doi:10.3772/j.issn.2095-915x.2024.01.004

中国个人信息保护应用与技术进展研究
——基于科学知识图谱视角

刘华玲 梁华璧 王希睿

上海对外经贸大学 统计与信息学院 上海 201620

开放科学（资源服务）标识码（OSID）

摘要：[目的 / 意义] 大数据时代，公众对个人信息保护的需求日渐多维化，基于科学知识图谱视角分析近十年中国个人信息保护演化趋势与前沿热点有其必要性。[方法 / 过程] 基于文献计量、突现检测、科学知识图谱等多种方式展开分析，以 CSSCI 和 CSCD 刊源分别表征应用与技术维度，梳理近十年国内个人信息保护研究的双维度发展趋势。[结果 / 结论] 我国个人信息保护研究文献数量呈现阶段性增长模式；个人、机构的合作方式以内部学术交流、地域性联系为主，多学科交叉格局有待形成；在应用维度上以图情类主导、公共化、数字化为研究趋势；在技术维度上以计算机类主导、精细化、智能化为研究趋势。全民监督与数据共治成为应对大数据时代个人信息保护新挑战的可行之策。

关键词：个人信息保护；科学知识图谱；文献计量；应用维度；技术维度；热点演化

中图分类号：G251；G35

Research on the Progress of Personal Information Protection Application and Technology in China——Based on the Perspective of Mapping Knowledge Domains

LIU Hualing LIANG Huabi WANG Xirui

Department of Statistics and Information, Shanghai University of International Business and Economics, Shanghai 201620, China

Abstract: [Purpose/Significance] In the era of big data, the public's demand for personal information protection is becoming increasingly multidimensional. It is necessary to analyze the evolution trend and cutting-edge hotspots of personal information protection in China in the past decade from the perspective of mapping knowledge domains. [Methods/Processes] Bibliometrics method, emergence detection and mapping knowledge domains are used to carry out analysis, while CSSCI and CSCD sources represent the application and technology dimensions respectively. [Results/Conclusions] The growth mode of the number of personal information protection research literature in China is phased stable and generally close to linear growth; The cooperation modes of individuals and institutions are mainly internal academic exchanges and regional connections, therefore

作者简介 刘华玲（1964-），博士，教授，主要研究方向为数据挖掘与欺诈识别；梁华璧（1997-），硕士研究生，主要研究方向为文献计量与风险管理，E-mail：m17801113522@163.com；王希睿（1999-），硕士研究生，主要研究方向为信息管理与知识图谱。

引用格式 刘华玲 , 梁华璧 , 王希睿 . 中国个人信息保护应用与技术进展研究——基于科学知识图谱视角 [J]. 情报工程 , 2024, 10(1): 42-58.

the interdisciplinary pattern needs to be formed; In the application dimension, it is dominated by library and information science, heading towards popularization and digitization; In the technical dimension, it is dominated by computer science, heading towards refinement and intelligence. National supervision and data co-governance have become feasible strategies to address the new challenges of personal information protection in the era of big data.

Keywords: Personal Information Protection; Mapping Knowledge Domains; Bibliometrics; Application Dimension; Technical Dimension; Hotspot Evolution

引言

随着“互联网 +”与大数据产业的蓬勃发展及其战略地位的上升，个人在充分享受着网络信息高度互联互通带来的服务便利的同时，也在无意识中将个人信息分享给了从事数据收集、分析与推断的调查者或互联网企业。个人作为以几何量级爆炸性增长的信息的贡献者，却在逐渐丧失对信息的掌控权，不仅难以成为信息多次集成再造新价值的受益者，还面临着个人隐私信息被大范围泄露和非法利用的风险。因此，在大数据时代建立完善的个人信息保护机制和体系已成为各界的共识。本文将采用文献计量方法与科学知识图谱可视化技术，对近十年我国相关个人信息保护文献进行系统梳理与总结，并深入分析与挖掘我国个人信息保护研究的前沿热点和动态发展趋势，为我国个人信息保护领域的政策实践与技术研发提供数据参考借鉴。

本文的贡献一方面在于当前从文献计量角度引入定量分析方法专门对个人信息保护研究领域发展概况进行科学阐述的文章比较少 [1-3]，另一方面则在于文献计量方面的研究大多是基于单一期刊来源或综合多个期刊来源开展关键词检索和计量分析 [4-5]，而本文在参考谭春辉和熊梦媛 [6] 在分析数据挖掘领域时将主题内容维度划分为理论类与应用类这一做法的基础上，考虑到 CSSCI 主要收录管理学、经济学等领域的人文社会科学学术期刊的应用性特点以及 CSCD 主要收录数学、工程技术等领域的自然科学学术期刊的技术性特点，分别基于 CSSCI 与 CSCD 刊源从应用和技术双维度对我国个人信息保护研究的近十年热点演变开展可视化分析，探究两个维度在发文体量、更迭模式、聚合方向等层面的差异与相同之处，以期促进应用与技术维度共建，进而推动个人信息保护研究的双向深入发展。

1 数据来源与研究方法

1.1 数据来源

在数字主导的信息膨胀时代，隐私与个人信息很难得到绝对的二元分割，看上去不存在私密性的个人信息在大数据技术的深度挖掘下极有可能会泄露个人隐私，对个人信息的保护最核心的指向是防范外界对隐私级别信息的侵犯，而维护潜在的隐私利益又能成为个人信息保护体系的建立基础 [7]。图 1 展示了个人信息保护与隐私保护的关系演变历程，可见个人信息保护的相关研究离不开隐私保护研究的成果支撑。

doi:10.3772/j.issn.2095-915x.2024.01.004

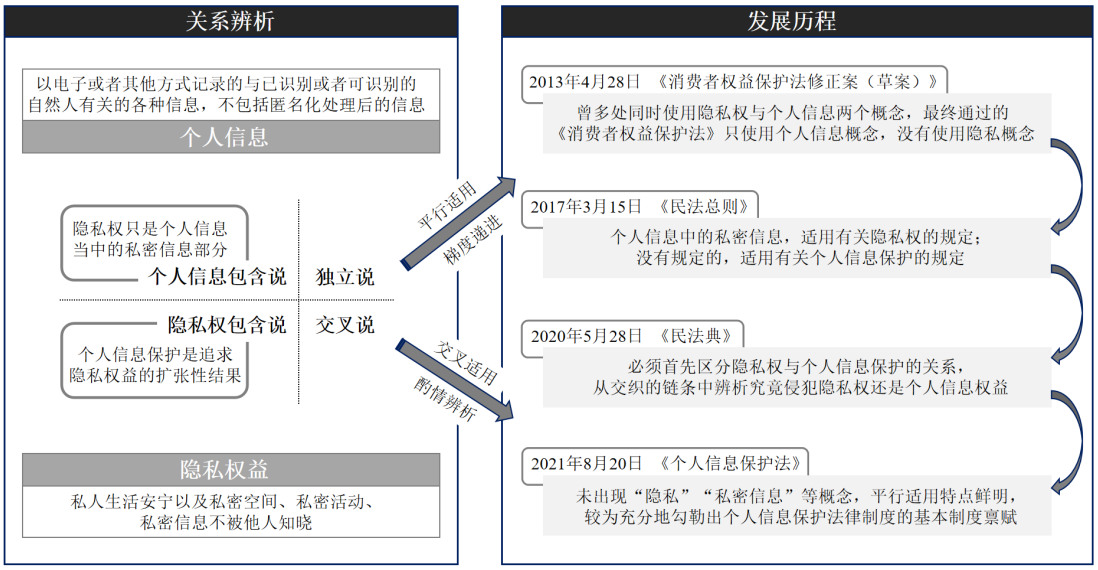

图 1 个人信息保护与隐私保护关系演变

为保证原始数据能够全面准确地反映我国当前的研究现状，并兼顾数据分析视角多样性与结果可解释性，本文拟以具有全面的学科覆盖面和完整的学科文献库的中国知网（CNKI）作为基础数据源，以“个人信息保护＋隐私保护”为主题、以“2011—2021 年”为时间节点、增加以 CSSCI、CSCD 来源期刊为检索来源的检索条件进行检索，经过筛选后共得到 3826 篇有效文献。

1.2 研究方法

首先，为了解国内个人信息保护研究的时间和空间结构分布特征，本文采用文献计量法分别按不同来源数据库、研究机构、作者统计发文量变化状况，并运用 ITGInsight 工具绘制研究机构与作者共现网络图来直观刻画其地域分布及合作关系 [8]。其次，为展示个人信息保护领域基于应用与技术双维度的演进态势和热点分布，利用词频探测技术获取高频关键词列表，并借助科学知识图谱软件 CiteSpace 分别绘制来源期刊分布图和细化关键词时间线分布图谱 [9]，实现研究焦点与未来趋势的可视化分析。

2 国内个人信息保护研究文献结构特征分析

2.1 时间分布特征

文献数量能够直观反映一个研究领域的受关注程度，本文数据集中包含 CSSCI 库论文 1790 篇，CSCD 库论文 2036 篇，总计 3826 篇文献，分别按 CSSCI 库发文量、CSCD 发文量和总发文量绘制年际变化趋势图，如图 2 所示。从总发文量来看，我国个人信息保护研究呈现出阶段性平稳、总体接近线性增长的特征，根据前后波动性的差异，可大体划分为启蒙探索阶段（2011—2012 年）、一次增长阶段（2013—2016 年）和二次增长阶段（2017—2021 年）。

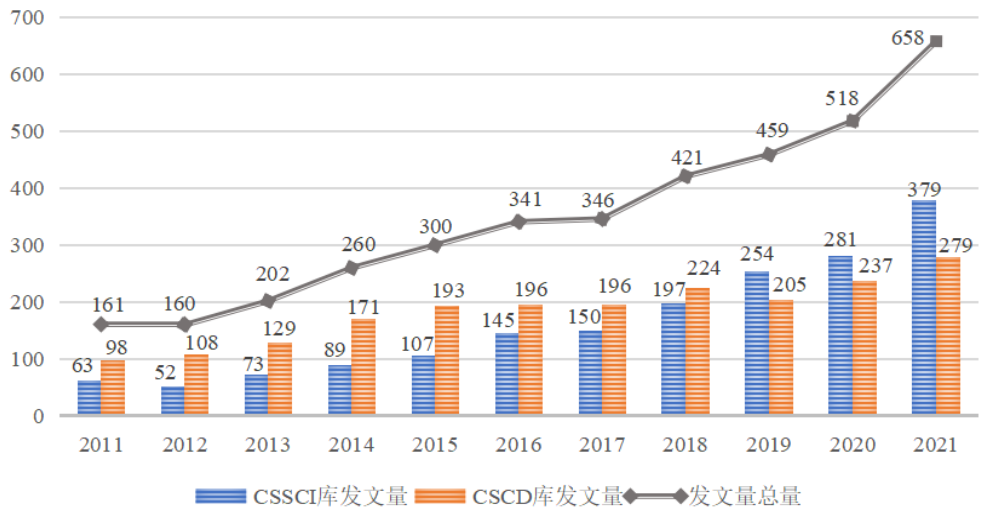

图 2　国内个人信息保护研究 CSSCI&CSCD 期刊发文量年际变化趋势

启蒙探索阶段的年均发文量约为 160 篇。在一次增长阶段，个人信息保护领域的研究迅速增多，共计发表论文 1103 篇，年均发文量约为 275 篇，该阶段发文量增幅为 68.81%，一系列法律法规的颁布表明了国家和政府对个人信息保护的重视，进而吸引了各界学者进入该领域从事深入研究。进入二次增长阶段后，共计发表论文 2402 篇，年均发文量约为 480 篇，发文量增幅高达 90.17%，2017 年个人隐私信息泄露事件频发、个人信息买卖形成非法产业链、企业间不正当数据竞争加剧等严峻现象，都推动着个人信息隐私保护研究的蓬勃发展。

从来源库视角看，CSCD 库发文量呈现逐年稳定增长趋势，即使在研究初期也近乎保持着 100 篇以上的年发文量，可见个人信息保护研究多从技术层面入手，且发文量随时间均衡分布。CSSCI 库发文量在近十年内存在指数增长态势，随着个人信息保护技术的更新与完善，学界开始转向研究如何将技术落实为经济社会领域的应用实践。2018 年以前，CSCD 库年发文量始终高于 CSSCI 库年发文量，随后三年 CSSCI 库发文量实现了逐年超越，成为二次增长阶段个人信息保护文献激增的主要推动力。

2.2　空间分布特征

2.2.1　研究机构

本文将样本文献进行规范化处理后形成的数据导入 ITGInsight，保留每个时间切片内排名前 30 的研究机构，绘制如图 3 所示的国内个人信息保护相关研究机构的共现图谱，图谱中每个节点表示一个研究机构，节点大小与该研究机构发文量呈正相关。节点间相连则表明两个机构间存在合作关系，连线的粗细反映了研究机构间合作关系的强弱。

从图 3 中可以看出，孤立节点大多为各大学的法学院，对于理论辨析、历史沿革等研究往往依赖于过往文献，多依靠机构内部力量推进；计算机科学、信息技术类学院形成的节点关联极为密切，针对算法模型、科学技术等研究更新迭代速度极快，与外部机构保持长期合作交流是把握前沿动态的必然之举。

doi:10.3772/j.issn.2095-915x.2024.01.004

图 3　国内个人信息保护研究机构共现图谱

由于大量研究机构以其二、三级下属机构署名发文，为了使文献计量更准确地反映一级研究机构的综合发文水平以及不同研究机构去除内部固有关联后的外部合作关系，本文采用人工统计的方法予以合并，表 1 展示了国内个人信息保护主题发文量排名前 10 的研究机构，共发表 987 篇论文，占研究文献总量的 25.79%，彰显了个人信息保护研究中的核心力量所在。其中，发文量最大的第一梯队是武汉大学（168 篇）、电子科技大学（163 篇）和中国科学院（154 篇），均是第二梯队研究机构发文量的近两倍。

表 1　国内个人信息保护主题发文量排名前 10 的研究机构

机构名称	发文量（篇）	机构名称	发文量（篇）
武汉大学	168	中国人民大学	78
电子科技大学	163	南京大学	68
中国科学院	154	南京邮电大学	64
西安电子科技大学	87	哈尔滨工程大学	63
清华大学	81	北京大学	61

研究机构分布呈现出三大特征：第一，研究成果在全国范围分布比较均匀，例如南京、北京、河南、哈尔滨、武汉、贵州、合肥等地都存在节点面积较大的研究机构，即国内各区域对个人信息保护的研究工作都予以足够的重视；第二，高校研究机构的参与程度最高，且以各高校中的法学院、信息学院与计算机学院为主体，这在一定程度上说明当前对于个人信息保护的研究主要还集中于理论层面，而在实际生活中与个人信息接触较多的如法院、图书馆、通讯社等应用型机构对个人信息保护的关

注相对低迷，有可能导致理论研究成果难以落地或者未达到预期效果的问题；第三，研究机构之间的合作具有较强的地域性特征，例如南京邮电大学、南京航空航天大学、安徽大学等长三角地区的高等院校间存在较强的合作关系，中国人民大学、中国科学院信息工程研究所等首都圈的研究机构也形成了较稳定的合作模式。

2.2.2 文献作者

以相同方式绘制我国个人信息保护研究相关作者的共现图谱，图 4 展现了发文量排名前 30 的文献作者信息，图谱中每个节点表示一个作者，节点大小与该作者发文量呈正相关。节点间的连线反映不同作者间的合作关系。

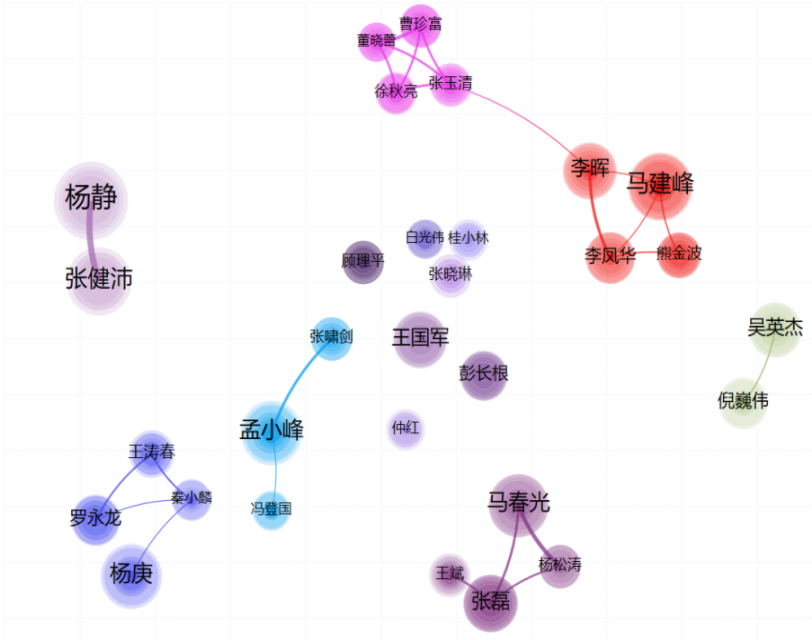

图 4 国内个人信息保护研究作者共现图谱

从合作的聚类效应来看，文献作者之间的关系存在内部性、包容性、联系性的特征。一方面可以看出大部分节点之间都存在连线，即高产作者之间普遍建立了不同程度的合作关系，孤立节点中不乏具备一定发文量的文献作者，但其研究联系可能存在于低产作者之间。其次，文献作者共现图中存在 4 个至少包含 4 个节点的研究群体，通常以发文量最大者为研究核心，这些研究群体中既包含来自同一研究机构的作者，也包含了所属研究机构在地理位置上跨度较大的作者。

来自西安电子科技大学的马建峰和李晖、来自福建师范大学的熊金波和来自中国科学院信息工程研究所的李凤华等人所组成的研究群体，主要涉及加密安全自毁方案 [10] 等创新类研究和云数据确定性删除 [11]、隐私计算 [12]、移动群智感知 [13] 等综述类研究；来自佳木斯大学的张磊、杨松涛和王斌与来自山东科技大学的马春光等人所组成的研究群体，研究内容高度集中在基于伪随机置换 [14]、随机网格 [15]、匿名区域层级扩张 [16] 等不同技术的位置隐私保护领域。这表明了当前个人信息保护领域的研究群

doi:10.3772/j.issn.2095-915x.2024.01.004

体在保持了同一研究机构内部紧密学术交流关系的同时，也注重开拓跨省份、跨地区的合作研究。

表 2 统计了国内个人信息保护主题发文量排名前 10 的作者，其近十年发表的关于个人信息保护的文章均在 20 篇以上，这些作者在共现图谱中均占据了较大的节点位置，是推动我国个人信息保护研究发展的中坚力量。由历年发文量来看，该主题研究的关注度一直处于较高的水平，但排名前 10 的作者发文量仅占 6.38%，表明个人信息保护领域作者基数庞大，高产作者与杰出作者比重有待提升。

表 2　国内个人信息保护主题发文量排名前 10 的作者分布

作者姓名	发文量（篇）	所属单位	作者姓名	发文量（篇）	所属单位
杨静	33	哈尔滨工程大学	马春光	24	山东科技大学
马建峰	27	西安电子科技大学	张磊	21	佳木斯大学
张健沛	27	哈尔滨工程大学	李晖	21	西安电子科技大学
孟小峰	25	中国人民大学	王国军	21	广州大学
杨庚	25	南京邮电大学	吴英杰	20	福州大学

杨静、张健沛和其他学者聚焦研究个性化轨迹隐私保护技术 [17] 与多敏感属性隐私保护方法 [18]；孟小峰、张啸剑等作者着重探讨大数据管理时代隐私风险管理问题 [19] 以及将差分隐私与自适应网格 [20]、矩阵分解 [21] 等技术融合开发隐私保护算法的可能性；杨庚、王涛春、秦小麟等作者联合提出传感器网络中的隐私保护方案 [22] 并综合分析联邦学习模式下隐私保护模型 [23]；王国军与其他学者合作设计移动社交网络中结合跨域代理重加密 [24]、混淆矩阵变换 [25] 等方法的朋友发现隐私保护机制。

3　国内个人信息保护研究应用维度演化分析

3.1　期刊构成

图 5 呈现了我国个人信息保护研究发文量排名前 25 的 CSSCI 来源期刊，其中情报类 7 种，共计发文 283 篇；图书馆类 7 种，共计发文 208 篇；法律类 5 种，共计发文 80 篇；新闻传播类 6 种，共计发文 77 篇。由此可见，CSSCI 刊源的个人信息保护文献主要集中发表在图情类期刊，而发文量占据前三的《现代情报》《情报理论与实践》和《情报杂志》均属于情报类期刊。

3.2　关键词突变分析

钟辉新 [26] 指出突现词检测相较于词频分析更易于锁定新兴主题的增长势头，特定领域在长期发展的过程中必然包含了多样化主题的相继突现，而单一突现主题往往伴随着批量含义相近关键词的诞生，共同阐释了领域内某一时间段而非时间点的研究态势。王梦婷 [27] 根据突变检测结果分时间段给出竞争情报领域的研究热点，卢新元等 [28] 借助关键词突变表观察企业知识转移领域在不同阶段的前沿趋势，因此，以核心关键词为导向并从阶段性视角分析个人信息保护领域的演进过程具备可操作性，也是应有之义。

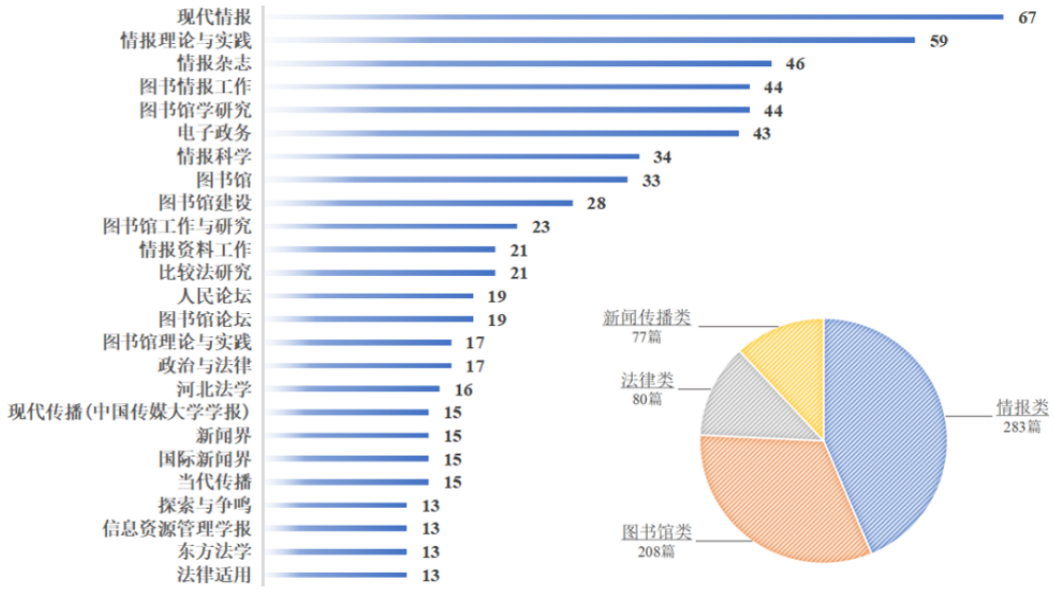

图 5　国内个人信息保护研究应用维度期刊分布

本文利用 CiteSpace 提供的词频探测技术，对个人信息保护领域的关键词进行突变分析，表 3 依次呈现了获取突现词的名称、突现强度、突现开始时间、突现结束时间以及突现持续时间段，所展示的 CSSCI 刊源在个人信息保护研究中最强突现的 26 个关键词基本属于社会科学领域概念，充分反映了 CSSCI 刊源的应用性特征，分别以具有较高的突现度的“图书馆”和“大数据”关键词为阶段核心，将近年来我国个人信息保护研究在应用层面的演进大致分为两个阶段：2011—2015 年和 2016—2021 年。

3.2.1　隐私与信息理论萌芽阶段（2011—2015 年）

该阶段的研究主要集中在如何完善法律体系中有关隐私权的界定，以及如何在图书馆等现实信息场所中实现对多样化信息的有序管理和高效传递。以“隐私”为主体的相关高频关键词多达五个，“立法”“法律保护”“伦理问题”关键词的存在说明了学者对于如何运用法律权威维护个人隐私权益投入了大量的研究资源，杜红原 [29] 阐明隐私权保护的重中之重是对隐私权概念的界定；安宝洋、翁建定 [30] 发现多元学科和部门的协同治理是解决大数据技术带来的网络信息伦理缺失问题的关键。

“图书馆”一词的突现度达 10.41，与“读者”“个性化服务”“个人信用信息”等关键词共同反映了当时图书馆作为线下信息流转的核心在信息服务中起到的重要作用；杨利军和高军 [31] 设计图书馆大数据可视化分析系统框架以期帮助图书馆员有效挖掘海量复杂大数据效能；陈臣和马晓亭 [32] 提出基于小数据的图书馆个性化服务推送模式更有助于响应读者的差异化需求。

3.2.2　大数据与法律权益融合阶段（2016—2021 年）

该阶段的研究主要集中探讨在大数据时代

doi:10.3772/j.issn.2095-915x.2024.01.004

表 3　国内个人信息保护研究应用维度突现词

突现词	突现强度	开始年	结束年	持续时间
保护	3.28	2011	2013	
读者	3.21	2011	2014	
实证研究	4.13	2011	2013	
隐私权保护	3.88	2011	2013	
个性化服务	3.75	2011	2014	
云计算	3.65	2011	2014	
隐私权	7.48	2011	2013	
个人信用信息	2.89	2011	2014	
图书馆	10.41	2011	2015	
隐私	4.49	2012	2015	
伦理问题	2.53	2012	2015	
电子商务	3.64	2013	2017	
立法	3.58	2013	2014	
人格尊严	2.67	2013	2015	
法律保护	5.22	2013	2015	
大数据时代	2.80	2014	2018	
大数据	10.16	2015	2016	
被遗忘权	4.76	2015	2018	
美国	4.08	2016	2017	
电子政务	2.76	2016	2019	
网络隐私	2.53	2017	2018	
政府数据	3.17	2017	2018	
高校图书馆	5.11	2017	2019	
数据开放	2.82	2017	2019	
影响因素	3.29	2017	2018	
知情同意	3.57	2019	2021	

如何维护衍生出的新型个人信息保护权利，以及如何在发达的信息网络及时获取有效的信息但又能保证个人隐私不被过度泄露。匡文波 [33] 指出当前的个性化推荐算法对多项个体权益都造成了不同程度的威胁，“被遗忘权”“网络隐私”“知情同意”等最新的高频关键词正是对这类问题的直接反映。蔡培如 [34] 通过欧盟法理实践历史探究以被遗忘权平衡言论自由与知情权利益的可能性；范海潮和顾理平 [35] 认为科学的知情同意构想能够缓解个人信息保护领域的知情主客体交流鸿沟、执行同意环节失灵等困境。

“大数据”一词的突现度达 10.16，与“数据开放”“电子政务”“网络隐私”等关键词共同描述了数据智能时代的信息对立现象，个人信息保护制度对传统隐私权规范的取代是维护政府数据安全的有益路径探索 [36]。值得注意的是，在高度开放的网络环境中人们能够便捷

搜寻到大量相关信息，但如何从驳杂的数据集中清洗出最有价值的信息具有较大的难度，反而会陷入一种相对信息匮乏的状态，甚至会主动以个人信息为代价去换取检索过程的简化。

3.3 热点演化趋势分析

从图 6 可以观察到每一主题在整个时间跨度上的演变进程以及由热点事件启发的研究方向。对于个人隐私保护、信息安全、图书馆管理的研究是贯穿整条时间线的，可见这些主题在个人信息保护研究中处于基础性地位。综合来看，应用维度的个人信息保护领域研究趋势主要呈现为“个人隐私转向公共数据”“人力调查转向数字治理”“信息自决转向政府责任”等宏观化特征，个人信息保护逐渐成为全民所需、全民参与、全民监督的一项公共事业。

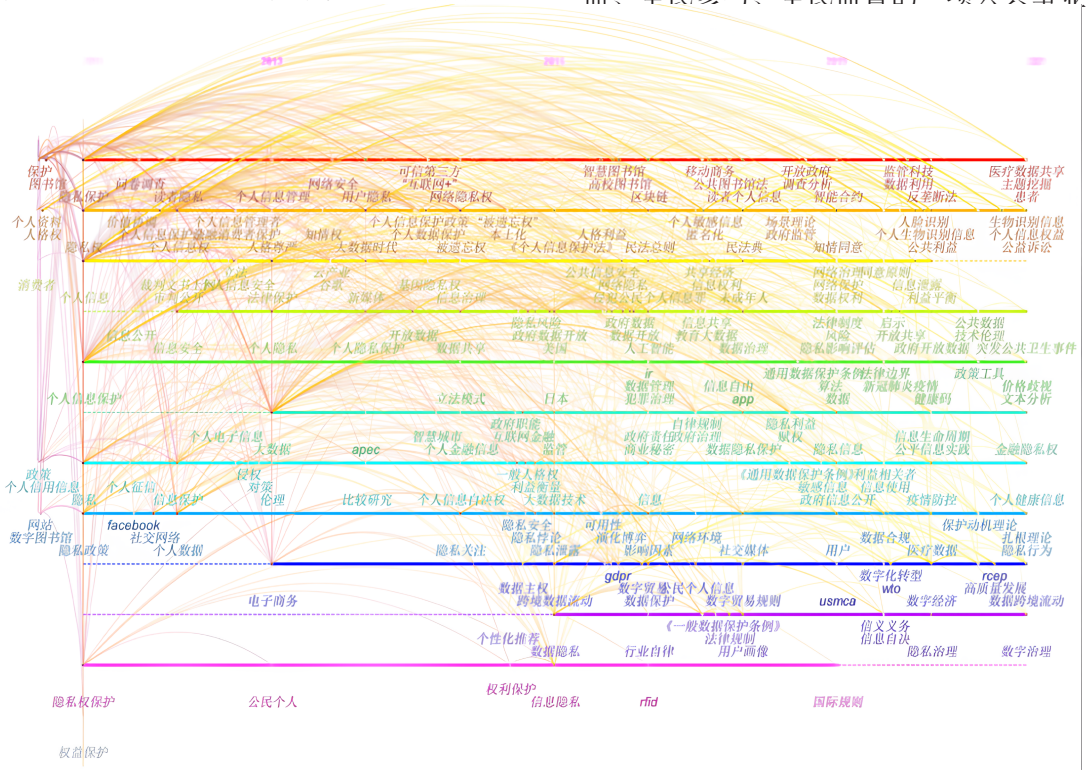

图 6　国内个人信息保护研究应用维度关键词时间线分布图谱

新冠肺炎疫情暴发以来，政府对于公民个人信息大范围、多频次的收集和使用引发了媒体和社会公众对个人信息安全的担忧，同时也掀起了化解由突发公共卫生事件造成的个人信息安全威胁的研究浪潮。苏今 [37] 通过涉疫信息的场景控制和权限设定分五阶段实现后疫情时代个人信息保护；占南 [38] 从三类个人信息保护关联主体视角出发构建面向隐私保护设计的全生命周期保护框架；沈伟伟 [39] 提出从紧急状态到日常状态的“恢复机制”理念并将其拓展到健康码记忆敏感信息的规制实践中。

医疗机构借助大数据技术以电子化的形式记录和存储了海量患者健康医疗信息，相关数据泄露将危害社会公共利益和国家安全，因此做好患者隐私和个人信息保护始终是研究焦点与难点。现行医疗隐私规则使得个人隐私保护机制生效范围有限，识别隐私利益的群体维度有助于获取统计意义上的健康信

doi:10.3772/j.issn.2095-915x.2024.01.004

息[40]；郭建[41]指出有必要建立一套伦理治理原则以规避健康医疗大数据应用中的伦理风险；相丽玲和陈琬珠[42]研究发现个人健康医疗信息保护逐步形成技术、管理、法律三位一体的跨领域保护模式。

4 国内个人信息保护研究技术维度演化分析

4.1 期刊构成

图 7 呈现了我国个人信息保护研究发文量排名前 25 的 CSCD 来源期刊，其中计算机类 12 种，共计发文 1026 篇；信息类 8 种，共计发文 431 篇；大学学报类 5 种，共计发文 95 篇。研究发现，25 种期刊仅占期刊总量的 4.96%，其 1552 篇的发文量占据文献总量的 40.56%，而 CSSCI 刊源文献仅占 16.94%。由此可见，国内学者更倾向于从技术维度研究个人信息保护问题，且高度集中于计算机领域，排名前五中的《计算机应用研究》《计算机科学》《计算机研究与发展》和《计算机应用》期刊发文量均超过 100 篇。

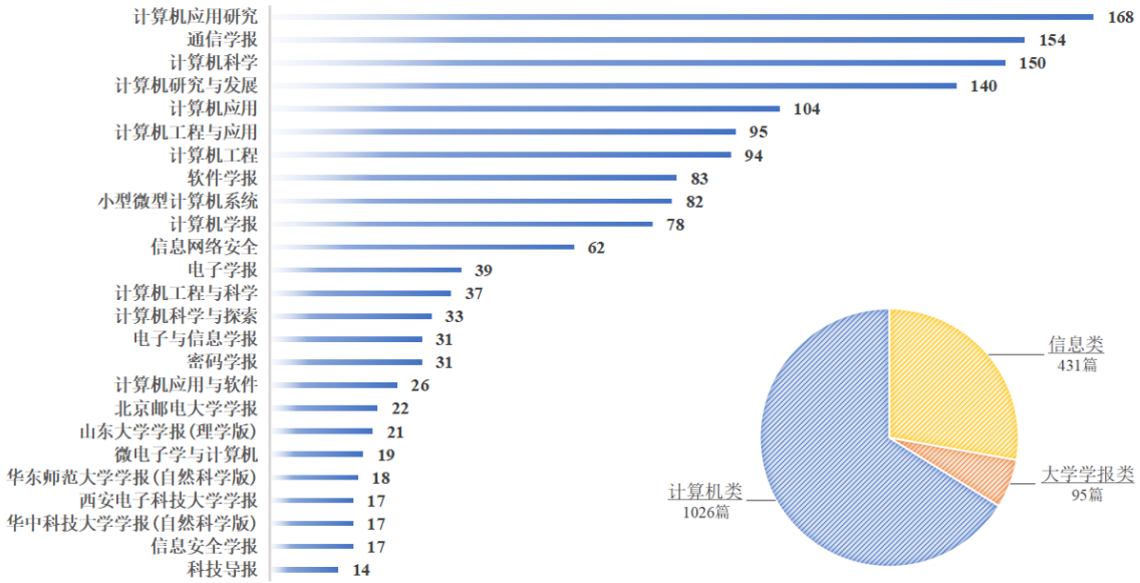

图 7 国内个人信息保护研究技术维度期刊分布

4.2 关键词突变分析

表 4 展示了 CSCD 期刊源在个人信息保护研究中最强突现的 37 个关键词，基本属于核心科技领域的概念，充分反映了 CSCD 刊源的技术特征，其中“数据发布”“无线传感器网络”“差分隐私”和“区块链”具有较高的突现度，分别以这些关键词为阶段核心，将近年来我国个人信息保护研究在技术层面的演进大致分为三个阶段：2011—2012 年、2013—2017 年和 2018—2021 年。此外，结合突现词数量容易发现，个人信息保护领域在技术维度的研究相较应用维度具有更快的更新速度、更多样化的研究对象以及更具波动性的未来趋势。

表 4　国内个人信息保护研究技术维度突现词

突现词	突现强度	开始年	结束年	持续时间
关联规则	3.2752	2011	2013	
l- 多样性	2.7388	2011	2013	
多敏感属性	3.4594	2011	2014	
泛化	2.7435	2011	2012	
聚类	4.1314	2011	2014	
有损连接	3.1074	2011	2013	
数据发布	7.6127	2011	2012	
数据挖掘	3.1659	2011	2013	
社会网络	4.2134	2011	2016	
数据融合	5.3095	2013	2014	
无线传感器网	8.1881	2013	2014	
数据聚合	2.4937	2013	2016	
数据聚集	2.7981	2013	2014	
完整性验证	2.8056	2013	2017	
大数据	2.7611	2014	2016	
匿名	2.8138	2014	2017	
轨迹数据发布	2.5775	2014	2015	
范围查询	3.5398	2014	2017	
移动互联网	2.5775	2014	2015	
位置隐私	2.8328	2015	2016	
移动社交网络	4.0756	2015	2017	
隐私	3.3081	2015	2016	
位置服务	4.2111	2015	2016	
k 匿名	3.0055	2015	2017	
隐私度量	2.8051	2015	2018	
属性基加密	3.5362	2017	2018	
轨迹数据	2.6797	2017	2018	
轨迹隐私	3.9837	2017	2018	
群智感知	2.5718	2018	2021	
移动群智感知	2.6686	2018	2021	
属性加密	2.8592	2018	2019	
智能电网	3.9256	2019	2021	
激励机制	3.4076	2019	2021	
区块链	30.902	2019	2021	
差分隐私	10.3813	2019	2021	
数据共享	3.858	2019	2021	
零知识证明	2.8032	2019	2021	

doi:10.3772/j.issn.2095-915x.2024.01.004

4.2.1 基础算法与隐私规则的探索优化阶段（2011—2012年）

该阶段的研究主要集中于对传统个人信息保护规则的优化更新和对基础个人信息保护算法的功能提升，“数据发布”作为该阶段突现度最高的关键词则表明，此时个人信息保护的研究热点依然围绕着数据本身，但一味追求数据的完全模糊化处理难以满足不同应用领域对匿名发布数据的质量需求，如何借助个性化隐私匿名技术平衡个性化服务质量和隐私保护效果受到广泛关注[43]。

刘彬等[44]在if-then算法设计中融入兴趣度、规则左件和逐步移项的思想，隐藏敏感信息的同时有效控制规则丢失率；刘峰等[45]借助安全多方计算和随机干扰矩阵缩减数据开销，高效解决半诚实模型下的隐私保护问题；王波和杨静[46]利用个性化扩展 l- 多样性逆聚类算法构建隐私匿名模型，以满足不同用户个性化的隐私保护需求；徐勇等[47]提出的考虑敏感属性权重的数据发布算法能够迎合不同应用领域的隐私保护意图和数据质量要求。

4.2.2 交互网络覆盖与隐私轨迹加密实践阶段（2013—2017年）

随着智能手机的普及和一系列自带定位功能的社交软件开发，位置信息的泄露风险急剧增加，世界高度互联互通使得行为个体往往会无意识被纳入多个社交关系网络，个人隐私数据更容易被窃取。因此，该阶段的研究更为注重如何紧贴数字时代的信息共享特性，研发能够迅速投入实践的个人信息保护通用技术和隐私攻击抵御机制，而非长期停留在实验精度和效度的数据模拟阶段。此时共现关键词数量出现激增，说明个人信息保护研究已经具有一定的热度。

以高频关键词“无线传感器网络”为核心，“移动互联网”“范围查询”“移动社交网络”等均体现了隐私数据的大范围散布，数据聚合对原始感知数据的篡改在一定程度上增加了隐私信息加密的难度。丁超等[48]基于隐私同态和聚合消息验证码技术提出的可恢复数据聚合方案能够兼顾数据隐私性与完整性；陈燕俐等[49]采用加法同态加密和同态消息认证码打造的轻量级安全数据融合保护方案便于处理多源异构数据场景；而这一时期诸如桶技术、前缀成员验证、保序加密等隐私保护范围查询处理技术尚未较好地实现隐私性、完整性、高效性和精确性四者之间的均衡[50]。

在隐私轨迹发布过程中如何既保证个体本身的敏感位置不泄露，并防止攻击者通过轨迹行为推导出其他敏感信息，也是当时专家学者热切关注的问题。杨静等[51]根据用户轨迹匿名的等级差异构造规模可变的个性化轨迹图模型以改善轨迹数据可用性低下的表现；张博闻等[52]通过构造隐私保护 Trie 树并根据（k,p）敏感轨迹进行剪枝重构来实现可信第三方隐私保护模块功能。

4.2.3 社群感知和差分隐私智能化应用阶段（2018—2021年）

该阶段中“区块链”的突现度高达 30.90，区块链技术所具备的难以篡改、智能合约、网状直接协作机制等优势令其实现了对中心式记账体系的颠覆，但性能及扩展性不足、数据隐私与访问控制机制不成熟和治理机制不完善等缺陷都限制着区块链产业的规范化发展，“同态加密”“属性加密”“零知识证明”等均反

映了当前研究者为解决区块链个人信息保护问题所做出的努力。田有亮等[53]基于属性加密和策略更新算法确保区块链场景下的交易隐私动态保护和溯源信息动态共享同步开展；许重建和李险峰[54]在蚂蚁区块链平台上验证其双重加密方案能以字段级别细粒度保护隐私数据并高效完成全链路操作；李莉等[55]综合群签名、隐私地址协议、零知识证明等多项技术制作了一类分布式可监管隐私保护方案。

“差分隐私”同样是该阶段突现度较高的关键词，通过向原始数据添加随机噪声以达到降低数据敏感度的目的，但在同等数据量的条件下，发布数据的有效比例会降低，故学界尝试构建不同隐私约束来规避数据恢复时的复杂计算，应用于位置、感知、电表等隐私数据进而追求高智能化个人信息保护水平。杨丽和陈思光[56]使用云雾协作的多级聚合模型节省智能电表通信开销，并借助同态加密技术实现隐私数据的轻量级认证；李卓等[57]考虑移动群智感知用户提交数据特点，分别从属性联合与属性独立视角出发设计本地差分隐私保护类算法。

4.3 热点演化趋势分析

图 8 表明通过主题聚类可以将技术维度的个人信息保护研究划分为八大主题。各主题间并非泾渭分明，而是相互渗透、相互演化、相互支撑，同时观察到研究早期各主题均有较多的分支，随着时间的推移，分布图谱的密度有所下降，细分研究方向主要集中于“差分隐私”“云计算”“区块链”和“基于位置的服务”四个主题。

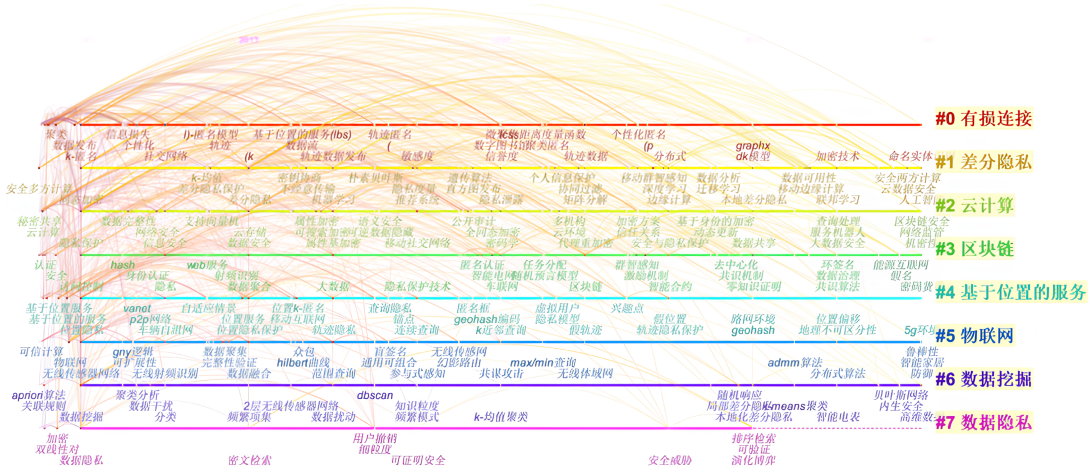

图 8　国内个人信息保护研究技术维度关键词时间线分布图谱

联邦学习是近两年来取得迅猛发展的研究热点之一，区别于传统隐私保护方法的原理，联邦学习不会出现数据和模型的传递，通过加密机制的参数交换方式保护用户数据信息的安全，常与区块链相结合用以抵御隐私攻击。方晨等[58]引入区块链的共识和激励机制促进设备参与联邦训练，辅以自适应差分方法防范边缘节点的隐私泄露可能；高胜等[59]以指数机制采样梯度贡献值高的隐私点，通过双因子调整机制实现全局模型更新，增强区块链体系下异步联邦学习方案的隐私保护效率。

随着第五代移动通讯技术（5G）的到来，数据传播速度和共享模式必将发生深刻变革，在方便用户长距离通讯和利用数据资源的同时，

doi:10.3772/j.issn.2095-915x.2024.01.004

个人数据的进一步透明化使个人信息保护将面临诸多未知的新挑战。冯登国等[60]将5G安全框架归纳为接入安全、网络安全、用户安全等七大层面，指出认证方案、密钥架构、切片隔离等是5G环境的关键安全技术。姜海洋等[61]在5G定位背景下通过降维处理、定位耦合和对称加密传输提升移动用户隐私保护能力。

5 结语与展望

本文借助CiteSpace和ITGInsight科学知识图谱软件，选取2011—2021年国内与个人信息保护研究主题相关的CSSCI和CSCD来源期刊文献，首先以文献计量手段和机构作者共现图谱识别个人信息保护研究的时空分布特征，其次分别从应用与技术维度可视化分析个人信息保护研究的热点演化趋势，得出结论如下：

第一，国内个人信息保护文献数量呈现出阶段性平稳、总体接近线性增长的时间分布特征，近年来应用类文章增长态势强于技术类文章。该领域发文在2013和2017年分别出现了显著的激增现象，随着《个人信息保护法》的颁布，我国个人信息保护研究必然出现更大的创新和突破。国内个人信息保护领域呈现出研究机构间的合作具有地域性和文献作者间的合作存在内部性的空间分布特征。总体来看，需要推动应用型研究机构积极参与到个人信息保护的研究当中，与理论型研究机构达成优势互补的合作，并注重增强不同研究群体跨省份、跨地区的科研联系。

第二，基于研究内容，可以将应用维度的研究区间分为隐私与信息理论萌芽阶段、大数据与法律权益融合阶段，将技术维度分为基础算法与隐私规则的探索优化阶段、交互网络覆盖与隐私轨迹加密实践阶段、社群感知和差分隐私智能化应用阶段。基于研究规模，技术维度排名前25的期刊发文体量是应用维度的两倍有余，在单一类型期刊上的集中趋势更强，突现词的平均持续时间更短，迭代也更为频繁。应用维度的热点演化随时间而进一步发散，技术维度的热点演化随时间逐渐聚焦于各自的核心要义，可见应用维度倾向于研究成果多点下探，技术维度倾向于研究成果单点泛化。

第三，中国个人信息保护具象于应用场景的研究呈现出公共化、数字化、宏观化的特征，研究主阵地坐落于情报、新闻传播与法律领域。信息爆炸性增长、信息主体纷繁交互、传播渠道日新月异致使个人信息泄露风险升级，进一步催生个人信息保护意识与隐私保护法律权益的诞生。然而随着个性化推荐算法等技术的不断普及，信息茧房与相对信息匮乏激化了情报价值与自我维权的对立，理论圈层的独立研究并不能从根本上平衡技术发展与信息保护的矛盾。

第四，中国个人信息保护具象于技术路线的研究呈现出专一化、精细化、智能化的特征，研究主阵地坐落于计算机和电子信息领域。早期匿名、脱敏和随机干扰等非密码技术兼顾了个性化信息保护需求与信息服务质量，同时以安全多方计算技术为代表的密码技术及其与其他密码技术的组合也被广泛运用于数据发布、共享和计算；随着位置轨迹与社交网络类信息的泛滥，学者通过改造图、树模型数据结构引入位置、社交网络外部信息以进行隐私攻击抵御机制与轨迹行为加密相关研究，多源异构的数据源提升了数据融合难度，主要使用全同态

加密技术解决数据不匹配、不均衡问题；当区块链的概念被提出，分布式个人信息保护方案与差分隐私算法更为适配，零知识证明具有数据最小化、关键信息隐私化的优势，在区块链方面也得到了广泛的应用，同时以联邦学习为代表的信息保护技术也在隐私 AI 领域被广泛研究，在保护个体核心数据资产的同时降低业务成本与风险 [62]。从基于数据本身的加密处理到同态加密、零知识证明等信息保护技术应用范围的拓展，再到感知不同的应用场景反哺优化技术主体，实际上是一个被动满足需求到主动感知拓展的过程。

个人信息保护事业的发展离不开应用与技术双维度的融合支撑，应用无法架空，技术终须落地，“事后惩罚”不如“事前监控”，以理论创新为更多具体的信息保护需求场景寻求更加合适和先进的技术解决方案，以应用场景的特殊性与真实性为技术发展引入人为知识的指导，从而推动个人信息保护研究领域内容不断丰富、完善。

参 考 文 献

[1] 孟小峰 , 王雷霞 , 刘俊旭 . 人工智能时代的数据隐私、垄断与公平 [J]. 大数据 , 2020, 6(1): 35-46.
[2] 熊金波 , 王敏燊 , 田有亮 , 等 . 面向云数据的隐私度量研究进展 [J]. 软件学报 , 2018, 29(7): 1963-1980.
[3] 方滨兴 , 贾焰 , 李爱平 , 等 . 大数据隐私保护技术综述 [J]. 大数据 , 2016, 2(1): 1-18.
[4] 沈光 . 文献计量视角下我国生态文明研究的热点与动态——基于 6240 篇 CSSCI 刊源文献信息的分析 [J]. 南京工业大学学报 (社会科学版), 2020, 19(5): 78-89, 116.
[5] 刘雅琦 , 贾利芳 . 知识图谱视角下的个人信息保护热点及前沿研究 [J]. 科技管理研究 , 2019, 39(18): 189-198.
[6] 谭春辉 , 熊梦媛 . 基于 LDA 模型的国内外数据挖掘研究热点主题演化对比分析 [J]. 情报科学 , 2021, 39(4): 174-185.
[7] 李忠夏 . 数字时代隐私权的宪法建构 [J]. 华东政法大学学报 , 2021, 24(3): 42-54.
[8] 陈超美 , 陈悦 , 侯剑华 , 等 . CiteSpace Ⅱ : 科学文献中新趋势与新动态的识别与可视化 [J]. 情报学报 , 2009, 28(3): 401-421.
[9] 刘玉琴 , 汪雪锋 , 雷孝平 . 科研关系构建与可视化系统设计与实现 [J]. 图书情报工作 , 2015, 59(8): 103-110, 125.
[10] 熊金波 , 姚志强 , 马建峰 , 等 . 面向网络内容隐私的基于身份加密的安全自毁方案 [J]. 计算机学报 , 2014, 37(1): 139-150.
[11] 熊金波 , 李凤华 , 王彦超 , 等 . 基于密码学的云数据确定性删除研究进展 [J]. 通信学报 , 2016, 37(8): 167-184.
[12] Li F, Li H, Niu B, et al. Privacy computing: concept, computing framework, and future development trends[J]. Engineering, 2019, 5(6): 1179-1192.
[13] 熊金波 , 毕仁万 , 田有亮 , 等 . 移动群智感知安全与隐私 : 模型、进展与趋势 [J]. 计算机学报 , 2021, 44(9): 1949-1966.
[14] 周长利 , 田晖 , 马春光 , 等 . 路网环境下基于伪随机置换的 LBS 隐私保护方法研究 [J]. 通信学报 , 2017, 38(6): 19-29.
[15] 杨松涛 , 王慧强 , 马春光 . 基于随机网格的位置隐私保护方法 [J]. 系统工程与电子技术 , 2018, 40(2): 422-426.
[16] 张磊 , 马春光 , 印桂生 . 匿名区域层级扩张的位置隐私保护方法 [J]. 系统工程与电子技术 , 2021, 43(2): 561-566.
[17] 杨静 , 张冰 , 张健沛 , 等 . 基于图划分的个性化轨迹隐私保护方法 [J]. 通信学报 , 2015, 36(3): 5-15.
[18] 谢静 , 张健沛 , 杨静 , 等 . 基于属性相关性划分的多敏感属性隐私保护方法 [J]. 电子学报 , 2014, 42(9): 1718-1723.
[19] 孟小峰 , 张啸剑 . 大数据隐私管理 [J]. 计算机研究与发展 , 2015, 52(2): 265-281.
[20] 张啸剑 , 金凯忠 , 孟小峰 . 基于自适应网格的隐私空间分割方法 [J]. 计算机研究与发展 , 2018, 55(6): 1143-1156.
[21] 张啸剑 , 付聪聪 , 孟小峰 . 结合矩阵分解与差分隐私的人脸图像发布 [J]. 中国图象图形学报 , 2020, 25(4): 655-668.
[22] 王涛春 , 秦小麟 , 张吉 , 等 . 传感器网络中基于路线的隐私保护数据聚集算法 [J]. 电子学报 , 2017, 45(6): 1334-1341.
[23] 杨庚 , 王周生 . 联邦学习中的隐私保护研究进展 [J]. 南京邮电大学学报 (自然科学版), 2020, 40(5): 204-214.
[24] 罗恩韬 , 王国军 , 陈淑红 , 等 . 移动社交网络中跨域代理重加密朋友发现隐私保护方案研究 [J]. 通

doi:10.3772/j.issn.2095-915x.2024.01.004

信学报 , 2017, 38(10): 81-93.
[25] 罗恩韬 , 王国军 , 刘琴 , 等 . 移动社交网络中矩阵混淆加密交友隐私保护策略 [J]. 软件学报 , 2019, 30(12): 3798-3814.
[26] 钟辉新 . 新兴趋势探测研究综述 [J]. 现代情报 , 2017, 37(12): 162-167.
[27] 王梦婷 . 基于突变检测的主题突变分析研究 [J]. 情报科学 , 2016, 34(12): 36-39.
[28] 卢新元 , 张恒 , 王馨悦 , 等 . 基于科学计量学的国内企业知识转移研究热点和前沿分析 [J]. 情报科学 , 2019, 37(3): 169-176.
[29] 杜红原 . 论隐私权概念的界定 [J]. 内蒙古社会科学 (汉文版), 2014, 35(6): 105-109.
[30] 安宝洋 , 翁建定 . 大数据时代网络信息的伦理缺失及应对策略 [J]. 自然辩证法研究 , 2015, 31(12): 42-46.
[31] 杨利军 , 高军 . 图书馆个性化服务中的大数据可视化分析与应用研究 [J]. 现代情报 , 2015, 35(7): 68-72.
[32] 陈臣 , 马晓亭 . 基于小数据的图书馆个性化推送服务与服务质量保证研究 [J]. 情报理论与实践 , 2015, 38(10): 95-99.
[33] 匡文波 . 对个性化算法推荐技术的伦理反思 [J]. 上海师范大学学报 (哲学社会科学版), 2021, 50(5): 14-23.
[34] 蔡培如 . 被遗忘权制度的反思与再建构 [J]. 清华法学 , 2019, 13(5): 168-185.
[35] 范海潮 , 顾理平 . 探寻平衡之道 : 隐私保护中知情同意原则的实践困境与修正 [J]. 新闻与传播研究 , 2021, 28(2): 70-85, 127-128.
[36] 商希雪 , 韩海庭 . 政府数据开放中个人信息保护路径研究 [J]. 电子政务 , 2021(6): 113-124.
[37] 苏今 . 后疫情时代个人涉疫信息的控制特点及其路径修正——以隐私场景理论为视角 [J]. 情报杂志 , 2021, 40(9): 124-132, 123.
[38] 占南 . 重大疫情防控中的个人信息保护研究——基于隐私保护设计理论 [J]. 现代情报 , 2021, 41(1): 101-110.
[39] 沈伟伟 . 论数字紧急状态的恢复机制——以新冠疫情防控为例 [J]. 清华法学 , 2021, 15(2): 121-142.
[40] 刘士国 , 熊静文 . 健康医疗大数据中隐私利益的群体维度 [J]. 法学论坛 , 2019, 34(3): 125-135.
[41] 郭建 . 健康医疗大数据应用中的伦理问题及其治理思考 [J]. 自然辩证法研究 , 2020, 36(3): 85-90.
[42] 相丽玲 , 陈琬珠 . 个人健康医疗信息保护的研究进展与未来趋势 [J]. 情报科学 , 2020, 38(6): 170-177.
[43] 王波 , 杨静 . 数据发布中的个性化隐私匿名技术研究 [J]. 计算机科学 , 2012, 39(4): 168-171, 200.
[44] 刘彬 , 孟凡荣 , 汪楚娇 . 基于兴趣度的隐私保护关联规则挖掘算法 [J]. 计算机工程与设计 , 2011, 32(6): 2124-2128.
[45] 刘峰 , 薛安荣 , 王伟 . 一种隐私保护关联规则挖掘的混合算法 [J]. 计算机应用研究 , 2012, 29(3): 1107-1110.
[46] 王波 , 杨静 . 一种基于逆聚类的个性化隐私匿名方法 [J]. 电子学报 , 2012, 40(5): 883-890.
[47] 徐勇 , 秦小麟 , 杨一涛 , 等 . 一种考虑属性权重的隐私保护数据发布方法 [J]. 计算机研究与发展 , 2012, 49(5): 913-924.
[48] 丁超 , 杨立君 , 吴蒙 . 一种同时保障隐私性与完整性的无线传感器网络可恢复数据聚合方案 [J]. 电子与信息学报 , 2015, 37(12): 2808-2814.
[49] 陈燕俐 , 张乾 , 许建 , 等 . 无线传感器网络多应用场景下的安全数据融合方案 [J]. 计算机科学 , 2017, 44(9): 162-167.
[50] 张晓莹 , 董蕾 , 陈红 . 无线传感器网络隐私保护范围查询处理技术 [J]. 华东师范大学学报 (自然科学版), 2015(5): 1-13.
[51] 杨静 , 张冰 , 张健沛 , 等 . 基于图划分的个性化轨迹隐私保护方法 [J]. 通信学报 , 2015, 36(3): 5-15.
[52] 张博闻 , 陈晶 , 杜瑞颖 . 一种个性化移动社交网络轨迹隐私保护方案 [J]. 计算机应用研究 , 2017, 34(3): 871-874.
[53] 田有亮 , 杨科迪 , 王缵 , 等 . 基于属性加密的区块链数据溯源算法 [J]. 通信学报 , 2019, 40(11): 101-111.
[54] 许重建 , 李险峰 . 区块链交易数据隐私保护方法 [J]. 计算机科学 , 2020, 47(3): 281-286.
[55] 李莉 , 杜慧娜 , 李涛 . 基于群签名与属性加密的区块链可监管隐私保护方案 [J]. 计算机工程 , 2022, 48(6): 132-138.
[56] 杨丽 , 陈思光 . 雾辅助的轻量级隐私保护数据多级聚合研究 [J]. 小型微型计算机系统 , 2020, 41(6): 1224-1230.
[57] 李卓 , 宋子晖 , 沈鑫 , 等 . 边缘计算支持下的移动群智感知本地差分隐私保护机制 [J]. 计算机应用 , 2021, 41(9): 2678-2686.
[58] 方晨 , 郭渊博 , 王一丰 , 等 . 基于区块链和联邦学习的边缘计算隐私保护方法 [J]. 通信学报 , 2021, 42(11): 28-40.
[59] 高胜 , 袁丽萍 , 朱建明 , 等 . 一种基于区块链的隐私保护异步联邦学习 [J]. 中国科学 : 信息科学 , 2021, 51(10): 1755-1774.
[60] 冯登国 , 徐静 , 兰晓 . 5G 移动通信网络安全研究 [J]. 软件学报 , 2018, 29(6): 1813-1825.
[61] 姜海洋 , 曾剑秋 , 韩可 , 等 . 5G 环境下移动用户位置隐私保护方法研究 [J]. 北京理工大学学报 , 2021, 41(1): 84-92.
[62] 霍炜 , 郁昱 , 杨糠 , 等 . 隐私保护计算密码技术研究进展与应用 [J]. 中国科学 : 信息科学 , 2023, 53(9): 1688-1733.