大数据时代国内外个人信息保护研究热点和演化趋势
——基于科学知识图谱分析的文献计量方法分析

彭 飞[1] 肖荻昱[2]

(1.金陵图书馆 江苏南京 210019)

(2.江苏省社会科学院图书馆 江苏南京 210019)

摘 要:[**目的/意义**]对大数据时代国内外个人信息保护的研究热点和演化趋势进行了总结和回顾,旨在为相关领域的研究提供参考和启示。[**方法/过程**]运用文献计量法和科学知识图谱法,基于 CNKI 和 Web of Science 数据库,以 ITGInsight 为主体工具,再辅之 Gephi、Excel、SATI 等科学计量与知识网络分析软件,对大数据领域国内外个人信息保护研究领域的热点分布、主题演化以及研究内容进行分析。[**结果/结论**]大数据时代国内外个人信息保护相关研究主题分布广泛、演化规律较为复杂,呈现出显著的变化趋势,在未来的研究中,需要综合考虑技术、法律、政策等多个方面的因素,以构建更加全面、系统的个人信息保护体系。

关键词:大数据;个人信息保护;研究热点;科学知识图谱;文献计量方法

中图分类号:G353.1 **文献标识码**:A **doi**:10.3969/j.issn.1005-8095.2024.05.002

Research Hotspot and Evolution Trend of Personal Information Protection at Home and Abroad in the Era of Big Data: Bibliometric Method Analysis Based on Scientific Knowledge Graph Analysis

Peng Fei[1] Xiao Diyu[2]

(1. Jinling Library, Nanjing Jiangsu 210019)

(2. Library of Jiangsu Academy of Social Sciences, Nanjing Jiangsu 210019)

Abstract: [**Purpose/significance**] This paper summarizes and reviews the research hotspots and evolution trends of personal information protection at home and abroad in the era of big data, aiming to provide reference and inspiration for researches in related fields. [**Method/process**] This paper uses the bibliometric method and scientific knowledge map method, based on CNKI and Web of Science databases, with ITGInsight as the main tool, supplemented by Gephi, Excel, SATI and other scientometric and knowledge network analysis software, to analyze the distribution of hotspots, topic evolution, and research content in the field of personal information protection research in big data field at home and abroad. [**Result/conclusion**] In the era of big data, research topics related to personal information protection at home and abroad are widely distributed, and the evolution law is relatively complex, showing a significant trend of change. In future research, it is necessary to comprehensively consider technology, law, policy and other aspects to build a comprehensive and systematic personal information protection system.

Keywords: big data; personal information protection; research hotspot; scientific knowledge map; bibliometric method

0 引言

当前,互联网的普及使人们的生活进入了万物互联的时代,每个独立的个体都成为赛博空间中重要的一个节点,各节点在虚拟场景中从事各项业务的过程中不可避免地产生了大量结构化、半结构化和非结构化数据。这些海量数据由于超过了传统数据分析和处理工具的极限,进而推动了大数据分析技术的出现,大数据分析技术使用计算机算法分析能够对看似无关的海量原始数据转化为可操作的信息(即知识),以促进更好的决策[1]。麦肯锡全球研

收稿日期:2023-09-01

作者简介:彭飞(1980—),男,本科,副研究馆员,主要研究方向为文献资源建设、图书馆智库;肖荻昱(1985—),男,硕士,副研究馆员,主要研究方向为知识计量学与智库研究。

究院在医疗卫生、金融商业、银行信贷等领域明确了大数据的潜力,如医疗健康领域运用大量个人电子健康信息进行疾病预防、个性化医疗[2]。金融行业通过分析个人信用大数据来预测市场行为和投资表现[3]。谷歌、eBay、Twitter 等公司围绕大量关于个人行为、信息请求和偏好的信息创建了整个商业模式,实现商业智能化服务[4]。虽然基于个人信息的大数据分析提高了业务效率并为知识发现创造了重大机遇,但它无意中也增加了个人信息泄露的风险,给隐私信任带来了前所未有的挑战。如英国 Cambridge Analytica 公司通过非法收集 87 万 Facebook 用户的数据来为企业客户提供消费者研究、定向广告推送有关的服务[5];500 万法国患者的敏感信息从大约 30 个实验室被盗并在网上传播[6]。在我国,仅 2022 年就发生了诸如平安人寿泄露 4 万条公民信息、学习通 1.7 亿条用户数据疑被泄露[7]。此背景下,国内外有关个人信息保护相关研究进入了快速发展阶段,且取得一定的研究成果。目前,学界尚未有关于本主题系统、全面的综述研究,本研究则从国际国内视野出发,考察既有研究,试图分析国内外大数据背景下个人信息保护研究的最新进展与热点问题,以期拓宽国内学者学术视野、了解前沿学术动态,既为我国今后有关大数据背景下个人信息保护的实践提供有益参考,也为相关政策制定提供科学依据。

1　研究设计与数据来源

1.1　研究设计

本文采用计量学法和知识图谱法,以 ITGInsight 为主体工具,再辅之 Gephi、Excel、SATI 等科学计量与知识网络分析软件,对大数据领域国内外个人信息保护领域的研究热点、主题演化以及研究内容进行分析,并输出为可视化的关键词共现、关键词聚类、关键词演化图谱,进而从图谱中揭示出国内外关于该领域的研究主题变化。

1.2　数据来源

中文数据来源于中国知网数据库平台,外文数据来源于 Web of Science(WoS)核心数据库,中文检索式为“主题=大数据 AND 个人信息保护”or“主题=大数据 AND 隐私保护”,期刊来源类型限定为核心期刊、CSSCI、CSCD,学科类型不限。外文检索式为 TS=(“big data”or“Privacy information”or“individual information protection”or“Privacy information protection”),数据收集时间截至 2023 年 5 月 31 日,剔除与研究主题明显不符的文献,再使用 ITGInsight 进行数据清洗,最终获得国外样本文献 455 篇、国内 428 篇作为本文的研究对象。

2　大数据时代国内外个人信息保护的研究热点分析

2.1　大数据时代国外个人信息保护的研究热点分析

本文利用 ITGInsight 工具对大数据时代国外个人信息保护的数据进行统计分析,提取出排名前 10 的高频主题词进行词频统计,分别为 privacy protection(193)、big date(183)、sensitive information(61)、personal information(60)、personal date(54)、data protection(44)、data privacy(43)、large amount(41)、differential privacy(37)、general data protection regulation(35)。再利用 ITGInsight 工具绘制了高频主题词间的共现聚类(见图 1),图 1 反映出国外学者关于本主题的研究主要焦距在大数据时代个人信息基本理论、个人隐私(敏感)信息保护机制、个人信息安全风险等几个维度。

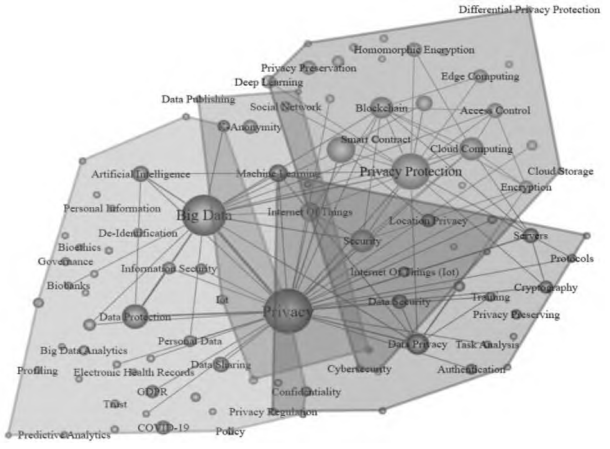

图 1　国外个人信息保护研究主题词聚类分布图

2.1.1　大数据时代个人信息保护基础理论问题研究

(1)个人信息保护的法理学分析。要探讨这一问题首先得认定个人信息的法律属性,有学者认为个人信息可以解释或定义为民法上的财产(或刑法上的财产),因此个体对于自己的个人信息拥有所有权[8]。也有学者认为由一组确定的个人信息组成的交易身份(transactional identity)具有法律性质,进而得出与个人相关联的个人信息亦是构成隐私权与身份权的基础[9]。个人信息的法律属性确定之后就要从法律的角度来对个人信息的使用作出规定。2018 年 5 月生效的《欧盟通用数据保护条例》(EU General Data Protection Regulation,GDPR)从法的角度对个人数据的收集、存储和处理做了规制[10]。首先,要求相关主体对个人数据的处理要遵循知情同意原则,只有在数据主体同意的情况下,才允许处理个人

数据。关于这一问题的研究主要集中在医学领域,如英国一般医学委员会规定医生如果在未获得患者完全知情同意的情况下处理患者病情信息将会面临《数据保护法》的起诉[11]。但也有学者指出,知情同意原则在大数据背景下面临着巨大挑战[12],譬如,为了给罕见病患者提供治愈可能性以及为了提升医疗的预防和诊断能力[13],不可避免地会收集大量个体的生物遗传数据,而且这些数据量越大越好[14],但严格遵守知情同意原则可能会威胁到许多其他有益研究的可行性,因为如果每项研究或每次使用参与者的遗传数据都需要同意的话会加剧研究成本[15]。此外,GDPR 等隐私法规规定个人数据收集必须要遵守最小必要原则(data minimization),它要求将个人数据处理过程限制为仅收集必要性信息内容,同时要删除垃圾数据(junk data)、限制数据囤积(data hoarding),还要定期评估收集个人信息的必要性。促成这一原则出现的主要原因是大数据环境下模糊了谁拥有数据、谁有权访问数据以及谁决定如何使用数据的界限,稍有不慎可能会造成数据丢失、被盗、泄露、伪造、变造或损坏[16]。在具体实践中实现最小必要原则可以通过两种不同的方式完成,一种是最小化个人数据量,另一种是最小化个人数据属性(类型)[17]。但也有学者认为从法的角度保护个人信息所规定的最小必要原则不利于数字社会的发展,同时小数据可能由于数据量的不足会增加决策的风险和成本[18]。

(2)个人信息保护的伦理学分析。个人信息保护的法理学分析主要从法律规范的视角对个人信息保护做出法律保障,并承认个人信息在法律上的有效地位或合法资格。而个人信息保护的伦理学分析则是从品格、理性和价值等道德规范角度出发判定相关主体对个人信息处理是否伤害到他人的合法权益[19]。当前我们身处大数据环境中,被称为数据资本(amount of data)的个人“数字足迹”(“digital footprints”)正被大量用作商业营销[20],加剧了信息隐私、信息安全、信息污染等伦理问题。从伦理学的视角研究信息安全主要涉及的问题是数据处理主体如何平衡个人信息保护和利用二者之间的关系,不可否认的是大数据通过分析个人信息对满足个性化信息需求、促进医疗进步以及推进智慧政务服务等有诸多好处,但也有可能侵犯到个人隐私。同时,目前还存在一种现象是个体很介意政府或企业处理个人信息,但他们却自愿在社交媒体上披露个人数据,如通过社交网站发布姓名、照片、出生日期、婚姻状况或在健康论坛上发布医疗数据,这种现象被称为“隐私悖论(privacy paradox)”[21]。此外,大数据分析造成的个人信息泄露还可能对一些特殊群体产生污名化现象,如艾滋病毒感染者以及精神病患者的个人信息泄露可能会被贴上负面标签,并受到社会排斥进而边缘化[22]。因此个人信息安全问题不仅仅是一个法律上的事实判断,还是一个具有社会学和哲学意义的价值判断。

(3)大数据时代个人信息保护的国际合作研究。大数据时代关于个人信息保护的“跨域研究”越来越普遍,主要集中在以下几个维度:

一是,大数据时代个人信息保护的国际或区域合作立法研究。目前通过远程网络利用个人信息进行电信诈骗、金融欺诈以及从事个人信息贩卖的国际犯罪案件时有发生[23],而仅仅使用本国或本地区的法律条款进行跨国司法互助或制裁显得愈发捉襟见肘,个人信息保护的国际立法重要性日趋受到重视。形成了像《欧盟通用数据保护条例》《保护和促进非洲个人隐私权指导原则》《美国-墨西哥-加拿大协定》《上海合作组织关于进一步加强互联网和信息安全及保护个人信息的声明》以及联合国中亚经济特别计划通过的《个人数据保护指导原则》等系列区域性的法律规范[24]。但以上除了《欧盟通用数据保护条例》,其他更像是一种软法(Soft Law),即在形式和效力上尚未完全法律化的规范[25],有学者将软法定义为国家之间缔结的不具约束力的协议[26],这种“不具约束力”就成为软法最致命的软肋[27],当行为主体不遵守时最多也只是受到声誉损害,所以国际或者地区性的软法在个人信息保护的实际操作上还面临诸多挑战。

二是,大数据时代不同国家或区域关于个人信息保护立法的比较研究。学界对国内外有关大数据背景下个人信息保护立法主要从是否为成文法与不成文法、是否为中央立法与地方立法以及是否为单一立法与复合立法等维度进行比较[28]。中国、欧盟和美国作为世界最大的经济体,在上述问题上有所差异。在中央立法与地方立法的比较上,欧盟和中国关于个人信息保护的相关法律规范由政府负责制定和实施,属于中央立法,也属于成文立法。而目前在美国尚未有类似于中国《中华人民共和国个人信息保护法》的统一联邦立法,相关规定散见于《隐私权法》《视频隐私保护法》《电子通信隐私法案》等法

律规范中[29],个人信息的保护更多是依靠行业的共识、自觉和习惯[30],所以应该属于不成文法(如习惯法)。同时也有学者认为美国相比欧盟在对个人信息的保护上比较宽松[31]。

三是,大数据时代个人信息的跨境流动相关问题研究。新兴的互联网技术不但加强了世界经济一体化格局,也进一步推动了跨境电子商务的快速发展,这给全球数字经济发展创造了新的发展红利和机遇[32]。目前,为了企业的国际化发展,跨境电子商务对个人信息的收集、存储、处理和应用已达到前所未有的水平[33],同时也伴随着个人信息泄露的风险,在国际领域个人信息泄露不仅关乎个人利益,更有可能威胁国家安全、破坏数据主权[34],从而降低甚至是削弱主权国家对个人信息的控制权限。为解决这一问题,构建国际性或区域间的双边数据安全流通机制、注重个人数据的本地化存储、强调出境数据的分类分级管理[35]、进行个人数据出境风险评估[36]以及采用大数据技术构建共用的个人信息传输渠道等方式受到学界普遍关注。

2.1.2 大数据时代个人信息保护实践应用技术研究

(1)数据收集阶段的反个人信息泄露技术。目前运用最为广泛的为匿名化技术(Anonymous technology),是指将个人信息和敏感数据从原始数据集中分离出来,从而保护个人隐私,具体包括删除个人敏感信息(如身份证号码、手机号码等),这种方法简单易行,但是可能会破坏数据的完整性[37],同时这种方法无法保护个人关联隐私,通过链接其他相关数据依然可能会推导出个人隐私。为解决这一问题,有学者引入了 k-anonymity("简称 k-匿名")方法,包括数据表 k-匿名模型和社交网络 k-匿名模型[38]。其主要思想是通过将数量不小于 k 的原始数据集中的敏感字段进行泛化[39],对于任意一条记录的攻击同时会关联到等价组中的其他 k-1 条记录,从而实现攻击者无法确定与特定用户相关的记录进而无法直接标识用户,实现对个人隐私的保护。但是 k-匿名依然存在数据质量损失以及可以通过反向查询进行攻击的缺陷[40]。为解决 k-匿名的不足,有学者提出了 L-diversity 模型,L-多样性的实现方法是对数据集进行预处理,使得每个敏感属性值(如姓名、身份证号码等)的出现次数不少于 L 次(其中 L 是一个预先定义的参数,但必须 L≥2),运用这种技术,攻击者就无法确定某个特定的人是否在数据集中,从而保护了个人隐私[41]。与之相似的还有(α,k)-匿名、(p,α,k)-匿名、t-逼近模型等。

(2)数据挖掘阶段的个人信息保护。为防止数据挖掘阶段对个人信息的泄露,具体技术有差分隐私法(Differential Privacy,DP)[42],DP 是一种用于保护个人隐私的统计技术,在数据挖掘中分析人员通过在原始数据中添加噪声来保护敏感数据,添加噪声的方法包括拉普拉斯密码(基于均值的机制)、指数机制(基于指数函数的机制)和哈密尔顿机制(基于哈密尔顿函数的机制)等。但上述技术要么需要个体将信息储存在一个收集个人原始数据的可信服务器中,要么受到计算和通信成本的影响在具体实施上有困难。为此,有学者介绍了一种本地差分隐私方法(Local differential privacy,LDP)[43],LDP 允许每个个体在自己的设备上对数据进行本地加密,然后将加密后的数据发送到分析人员,分析人员可以使用这些数据执行数据分析任务,而不知道任何特定个体的身份。在 LDP 基础上有学者提出了一种室内位置隐私的差分隐私方法(Local Differential Privacy for Indoor Location Privacy,LDP-ILP)[44],当前智能手机、平板电脑和智能可穿戴设备中的大多 APP 服务都需要获取用户位置信息,形成位置指纹(Location Fingerprinting,LF)[45]。LDP-ILP 为了保护个人位置信息的机密,在用户的个人位置信息发送到数据分析器之前,使用随机算法在本地对个人位置信息进行扰动(perturb)[46],但同时也确保扰动后人的位置信息仍然保留数据的统计特性[47]。这种技术在金融、医疗保健、科学研究等领域比较常见。

(3)数据传输阶段的个人信息保护。数据加密技术(Cryptographic Techniques)是在数据传输过程中常用的一种方式[48],主要有对称加密和非对称加密,此外,访问控制、审计日志、入侵检测系统、数据传输安全协议等方式也对数据传输阶段个人信息的保护具有一定作用。

2.2 大数据时代国内个人信息保护的研究热点分析

本文使用 Gephi 工具计算出国内文献中每个共现词的特征向量中心度,排名前十的分别为大数据(1)、个人信息(0.93)、隐私权(0.90)、个人信息保护(0.88)、隐私保护(0.66)、个人信息保护法(0.63)、大数据时代(0.62)、个人隐私(0.60)、个人数据(0.55)、人格权(0.53)。再使用 Gephi 工具对大数据时代国内个人信息保护关键词进行了网络聚类(见图 2),使用模块度(Modularity)算法计算出聚类值为 0.580,证明聚类良好。据此,可以将大数据时代

国内个人信息保护的研究热点概括为以下几个维度。

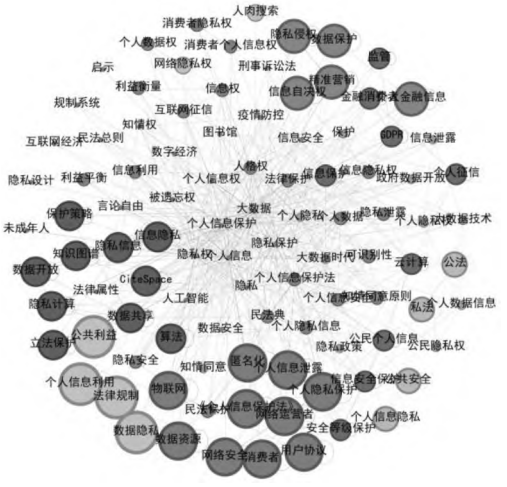

图 2　国内个人信息保护研究主题词聚类分布图

2.2.1　大数据时代个人信息保护基本理论问题研究

(1)大数据时代个人信息在具体领域的应用。一是,金融行业利用个人信用信息进行征信监管。能够表明个人信用的大量个人周边数据在信贷审批、贷后风险监测与预警等方面具有良好的应用效果[49]。二是,提升政府治理能力的现代化。如在抗击新冠肺炎疫情中,政府应用大数据技术,使用健康码、行程码、人脸识别等提高了应急管理的速度和力度[50]。还有在公安领域,基于大数据技术的智慧警务在应对突发公共安全事件风险、恐怖主义安全风险以及高科技犯罪风险中显现出强大的风险治理功能[51]。三是,在商业领域,运用个人信息在客户分析、风险管理、欺诈检测、产品研发、广告投放以及供应链管理都有良好的前景。四是,在教育领域通过使用大数据、人工智能等技术对受教育者的行为、表情等个人数据进行分析有助于进行学习干预、提升学习效果[52]。

(2)大数据时代个人信息存在的安全风险研究。一是,大数据技术的运用使得个人信息的收集边界日益模糊。主要体现在个人信息与非个人信息区分开始模糊[53],网络痕迹、购物习惯等间接与个人相关的信息在《个人信息保护法》中未做明确规定,个人信息收集边界的模糊加剧了个人信息泄露的风险。二是,大数据技术弱化了知情同意规则。“知情同意”的逻辑是基于程序性保护,即要经过“知情→同意→数据处理”,而大数据技术弱化了个体对个人信息的自由支配,换言之权利主体可能失去了对个人信息的控制,程序性规则被架空[54]。三是,个人信息共享阶段信息泄露问题严重。信息共享之前数据几乎处于分散状态,在数据共享之后零星分布的信息可以精确刻画出一个人的“数字形象”[55],第三方服务平台可能会用此信息分析一个人的特征、爱好,进而精准投放垃圾信息,甚至会监控个人行为。四是,个人信息安全素养缺乏。目前访问大多网站需要提交个人信息,对于网站和机构收集信息的目的个人一般不会主动去了解,这反映了个体对个人信息安全保护态度上的漠视[56]。此外,信息安全技术落后、风险防控系统缺位、内外监督制度薄弱等都是个人信息保护存在的安全风险[57]。

(3)大数据时代个人信息保护的法理学研究。本文使用 Gephi 工具对法学领域关于本主题的研究进行关键词网络共现,通过图 3 可以将大数据时代个人信息保护在法学学科的研究划分为以下几个维度,限于文章篇幅,仅做研究现状介绍,不做述评。

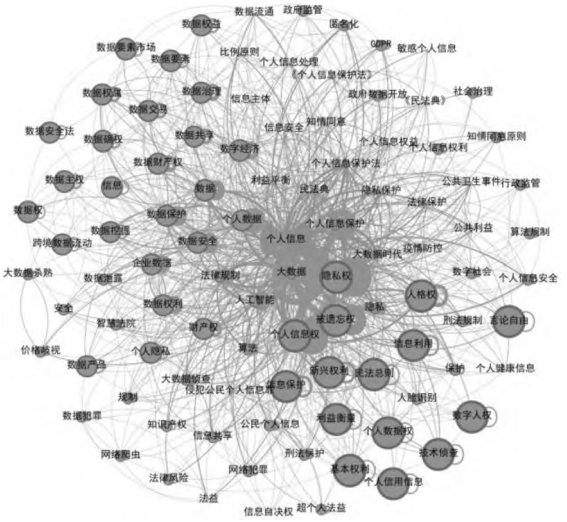

图 3　国内个人信息保护研究在法学学科主题词聚类分布图

一是,法理学将个人信息保护视为一种重要的权利,即个人信息权。从宪法学角度来看,个人信息保护是一种基本人权,个人信息的泄露、滥用和非法处理会对个人的自由、尊严和隐私造成严重侵犯[58];从民法的角度来看,个人信息是一种具有经济价值的财产,需要通过签订隐私协议、约定违约责任等方式,来维护自己的个人信息权利[59]。从刑法的角度来看,《个人信息保护法》中相关规定与刑事诉讼中限制公权与保障人权的制度追求相契合,因此,刑法可以规定对个人信息泄露和滥用的行为进行惩罚[60]。

二是,个人信息保护被视为法治和民主的必要条件。一方面,个人信息保护是维护法治社会的重要手段。个人信息保护可以防止个人信息的泄露和滥用,从而减少社会的不安全感和不信任感,促进社

会的和谐稳定,最终实现“以人民安全”为宗旨的总体国家安全观[61];另一方面,通过权力约束权力、权利对抗权力和诚信引导权力等措施对个人信息泄露提供法律救济[62],控制并排除个人信息泄露、误用以及贩卖等引起的侵权问题是实现社会公平正义的基础。

三是,新型技术环境中个人信息保护的法律障碍。在新型技术环境中,个人信息的跨境传输变得越来越普遍,但是源于不同国家之间个人信息保护法律规范上的差异,使得个人信息保护面临法律适用的困难[63]。同时,新型信息技术满足了用户隐私保护的需求,但个体身份的不易被识别,在客观上加大了对利用虚拟货币实施洗钱、毒品交易等违法犯罪活动的侦破难度[64]。此外,新型技术进步导致的立法滞后也成为个人信息保护的法律障碍之一。

2.2.2　大数据时代个人信息保护实践应用技术研究

有学者针对政府数据开放中个人隐私可能泄露的风险,提出了 ISM-MICMAC 技术模型[65]。对于云平台中的个人信息保护,有学者提出了采用位拆分与位合并的高性能数据隐私保护方法 BSBC(Bit Split Bit Combine,BSBC)[66]和基于单个准标识符组的最小化信息损失匿名方法(Anonymization with Minimum Information Loss,AMIL)[67]。此外,基于数据匿名的隐私保护技术包括 k-匿名、L-多样性和 t-接近等方法的研究备受国内学者关注。

3　大数据时代国内外个人信息保护的演化路径分析

关键词突显和关键词演化能够反映出一段时间内学界对该主题研究的主题演化趋势,本文使用 IT-GInsight 软件分别绘制了国外和国内对应可视化图谱,据此对大数据时代国内外个人信息保护的演化路径进行了分析。

3.1　大数据时代国外个人信息保护的演化路径

如图 4 所示,大数据时代国外个人信息保护相关研究主题分布广泛、演化规律较为复杂,呈现出显著的变化趋势。结合图 5 的主要关键词突现,可以将国外大数据时代个人信息保护研究分为三个阶段。

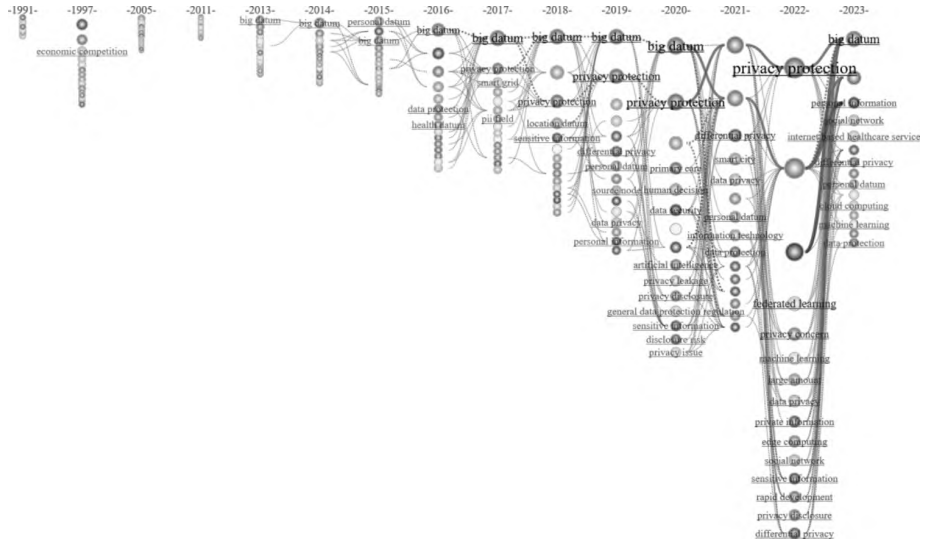

图 4　国外个人信息保护研究关键词演化图

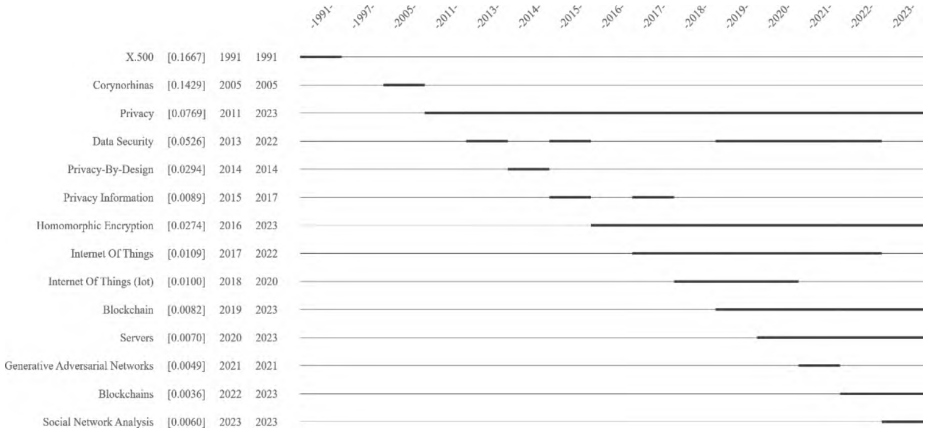

图 5　国外个人信息保护研究关键词突现图

第一阶段:基础理论探索(2012—2015 年左右)。这一阶段,学者们主要对大数据使用对个人信息带来的风险进行了探讨,认为数字技术使得个人信息可以被计算机轻易记录、存储和传播,从而给个人信息保护带来了极大威胁。为解决这些困境,学界提出了一些应对风险的措施,如采用数据加密、身份认证等技术防止个人信息泄露。

第二阶段:社交网络时期的个人信息泄露与保护(2015—2018 年左右)。社交网络的飞速发展使得个人信息被广泛采集和利用,进一步加剧了隐私泄露的风险。个人信息不仅局限于能够直接识别个体身份的显性个人信息,用户的搜索记录、浏览记录、购买记录等隐性信息也成为黑客收集的对象,例如,一些商业机构可能会通过隐性个人信息进行广告的精准推送,从而使用户感到不适和困扰。为解决这些问题,学界提出了差分隐私技术、数据最小化原则、区块链、数据备份以及信息加密等技术。

第三阶段:大数据与隐私保护的平衡(2019 年—至今)。大数据的使用能够促进社会治理更加智能化、智慧化和精准化,但也会导致个人信息泄露的风险,可见个人信息保护和信息流通之间存在一定的矛盾,因此如何平衡好二者之间的关系成为学界又一热点,“数据隐私边界”还需从技术哲学角度出发进行回应。

3.2　大数据时代国内个人信息保护的演化路径

如图 6 所示,可以将大数据时代国内个人信息保护研究概括为三个阶段。

第一阶段:个人信息保护的基础理论建构时期(2013—2015 年)。包括大数据时代的个人信息界定、个人信息泄露风险归纳、个人信息保护技术以及个人信息保护法律制度。大数据时代涉及的是个人信息可能不仅来自个人主动提供的信息,还可能包括由第三方通过数据挖掘、分析等手段从公开渠道或者第三方数据源获取的信息,导致个人信息的界定变得更加复杂。这一时期威胁个人信息保护的主要风险有:个人信息的不当收集或滥用、黑客攻击和网络钓鱼、移动设备安全漏洞、云计算平台中的数据泄露及供应链数据泄露等。对此,学者们从物理控制(锁住笔记本或硬盘)、法律制度(倡导制定个人信息保护法规、设立监管机构等)、技术手段(防火墙、杀毒软件)等维度研究了个人信息的保护问题。

第二阶段:个人信息保护的理论与技术创新时期(2016—2021 年)。在理论层面,个人信息权(人格权、隐私权)、隐私自我管理理论、数字身份理论、信息公正理论以及数字隐私规范理论为个人信息保护提供了理论依据。在技术层面,数据加密、数据脱敏、访问控制、分布式存储、智能合约、身份验证以及溯源追责为个人信息的保护提供了新的技术支持。

第三阶段:个人信息保护面临的新挑战与新问题(2022 年—至今)。元宇宙、ChatGPT 等技术使得人信息保护面临新的挑战。以元宇宙技术为例,用户在元宇宙中使用虚拟现实设备,如头戴式显示器、手套、耳机等,这些设备会收集用户的面部表情、眼球运动、手运动等个人数据,这些数据可能被用于广告、营销等目的,从而侵犯个人的隐私。

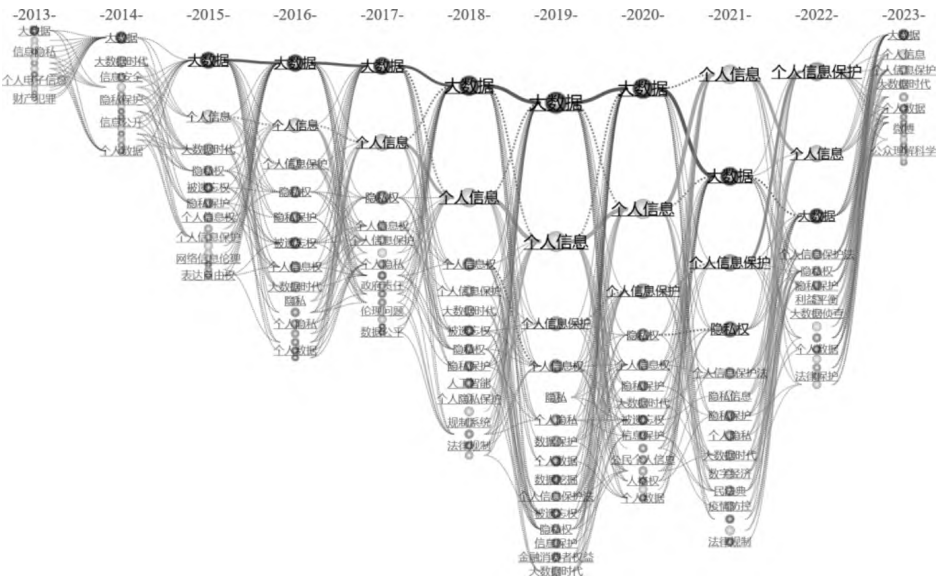

图 6　国内个人信息保护研究关键词突现图

4 结语

本文对大数据时代国内外个人信息保护的研究热点和演化趋势进行了总结和回顾,旨在为相关领域的研究提供参考和启示。在未来的理论研究中,需要综合考虑技术、法律、政策等多个方面的因素,以构建全面、系统的个人信息保护体系。同时,业界也需要不断探索新的隐私保护技术和方法,以应对日益增长的信息安全挑战。

参考文献

[1] CHEN J,CHEN Y,DU X,et al. Big data challenge: a data management perspective[J]. Frontiers of computer Science,2013(7): 157-164.

[2] PARK A,SONG J,LEE S B. Healthcare service analysis using big data[J]. Journal of the Korea Society of Computer and Information,2020,25(4): 149-156.

[3] WANG Y,ZHANG Y,LU Y,et al. A Comparative Assessment of Credit Risk Model Based on Machine Learning: a case study of bank loan data[J]. Procedia Computer Science,2020,174: 141-149.

[4] BURSTR M T,PARIDA V,LAHTI T,et al. AI-enabled business-model innovation and transformation in industrial ecosystems: A framework,model and outline for further research[J]. Journal of Business Research,2021,127: 85-95.

[5] HUANG L,ZHOU J,LIN J,et al. View analysis of personal information leakage and privacy protection in big data era—based on Q method[J]. Aslib Journal of Information Management,2022,74(5): 901-927.

[6] Libération. Les informations confidentielles de 500 000 patients français dérobées à des laboratoires et diffusées en ligne[EB/OL]. (2021-02-23)[2023-05-23]. https://www.liberation.fr/checknews/les-informations-confidentielles-de-500-000-patients-francais-derobees-a-des-laboratoires-medicaux-et-diffusees-en-ligne-20210223_VO6W6J6IUVATZD4VOVNDLTDZBU/.

[7] 赤峰市司法局. 2022 年国内十大信息泄露事件[EB/OL]. (2023-01-04)[2023-05-23]. http://sfj.chifeng.gov.cn/sfj_ztzl/wlaq/202301/t20230104_1939517.html.

[8] YUAN W. Fair data transactions across private databases[J]. IEEE Access,2020(8): 53720-53732.

[9] SULLIVAN C. Digital identity,privacy and the right to identity in the United States of America[J]. Computer Law & Security Review,2013,29(4): 348-358.

[10] GRUSCHKA N,MAVROEIDIS V,Vishi K,et al. Privacy issues and data protection in big data: a case study analysis under GDPR[C]// 2018 IEEE International Conference on Big Data (Big Data). Seattle: IEEE,2018: 5027-5033.

[11] PETO J,FLETCHER O,GILHAM C. Data protection,informed consent,and research[J]. BMJ,2004,328(7447): 1029-1030.

[12] RAU H,GEIDEL L,BIALKE M,et al. The generic Informed Consent Service gICS: implementation and benefits of a modular consent software tool to master the challenge of electronic consent management in research[J]. Journal of Translational Medicine,2020,18(1): 287-299.

[13] GREELY H T. The uneasy ethical and legal underpinnings of large-scale genomic biobanks.[J]. Annu Rev Genomics Hum Genet,2009,8(8): 343-364.

[14] VAYENA E,BLASIMME A. Biomedical Big Data: New Models of Control Over Access,Use and Governance[J]. Journal of Bioethical Inquiry,2017,14(4): 501-513.

[15] CONBOY C. Consent and Privacy in the Era of Precision Medicine and Biobanking Genomic Data[J]. American Journal of Law & Medicine,2020,46(2/3): 167-187.

[16] LI H,YU L,HE W. The impact of GDPR on global technology development[J]. Journal of Global Information Technology Management,2019,22(1): 1-6.

[17] PALLAS F,HARTMANN D,HEINRICH P,et al. Configurable per-query data minimization for privacy-compliant web APIs[C]//Web Engineering: 22nd International Conference. Cham: Springer International Publishing,2022: 325-340.

[18] HOSSNY A,NAHAVANDI S,CREIGHTON D. Minimizing impact of bounded uncertainty on McNaughton's scheduling algorithm via interval programming[C]// 2013 IEEE International Conference on Systems,Man,and Cybernetics. Washington: IEEE,2013: 970-976.

[19] SULA C A. Research ethics in an age of big data[J]. Bulletin of the Association for Information Science and Technology,2016,42(2): 17-21.

[20] GOLDER S A,MACY M W. Digital footprints: Opportunities and challenges for online social research[J]. Annual Review of Sociology,2014,40: 129-152.

[21] NORBERG P A,HORNE D R,HORNE D A. The privacy paradox: Personal information disclosure intentions versus behaviors[J]. Journal of consumer affairs,2007,41(1): 100-126.

[22] BABI D,BABI R,VASILJ I,et al. Stigmatization of mentally ill patients through media[J]. Psychiatria Danubina,2017,29(5): 885-889.

[23] ZIYI X. International Law Protection of Cross-Border Transmission of Personal Information Based on Cloud Computing and Big Data[J]. Mobile Information Systems,2022(12): 1-9.

[24] GREENLEAF G,COTTIER B. International and regional commitments in African data privacy laws: A comparative

analysis[J]. Computer Law & Security Review, 2022, 44: 105638.

[25] ABBOTT K W,SNIDAL D. Hard and soft law in international governance[J]. International organization, 2000, 54(3): 421-456.

[26] KLABBERS J. Constitutionalism and the making of international law[J]. No Foundations: Journal of Extreme Legal Positivism,2008,5: 84-112. .

[27] ELLIS J. Shades of grey: Soft law and the validity of public international law[J]. Leiden Journal of International Law,2012,25(2): 313-334.

[28] BOTHA J,GROBLER M M,HAHN J,et al. A high-level comparison between the South African protection of personal information act and international data protection laws[C]// ICML G. 5th International Conference on Management Leadership and Governance(ICMLG2017). Saint Petersburg:ICMLG,2017: 57-66.

[29] SEAN H. A Patchwork is Not Acceptable: Making the Case for a National Privacy Law[EB/OL]. (2019-07-29)[2023-05-23]. http://www.chnlawyer.net/law/subs/xingfa.html#347.

[30] RICHARDS N,HARTZOG W. Taking trust seriously in privacy law[J]. Stanford Technology Law Review,2015, 19: 431-472.

[31] MCGEVERAN W. Friending the privacy regulators[J]. Arizona Law Review,2016,58: 959.

[32] VORONKOVA V H,NIKITENKO V A,TESLENKO T V,et al. Impact of the worldwide trends on the development of the digital economy[J]. Amazonia investiga,2020,9(32): 81-90.

[33] LV Q. Supply chain decision-making of cross-border E-commerce platforms[J]. Advances in Industrial Engineering and Management,2018,7(1): 1-5.

[34] HUMMEL P,BRAUN M,TRETTER M,et al. Data sovereignty: A review[J]. Big Data & Society,2021,8(1):1-17.

[35] FERRACANE M. Restrictions on Cross-Border data flows: a taxonomy[J]. SSRN,2017(1):1-26.

[36] GRITZALIS D,ISEPPI G,MYLONAS A,et al. Exiting the risk assessment maze: A meta-survey[J]. ACM Computing Surveys (CSUR),2018,51(1): 1-30.

[37] TAMURA S,TANIGUCHI S. A scheme for collecting anonymous data[C]// 2013 IEEE International Conference on Industrial Technology (ICIT). South Africa: IEEE, 2013: 1210-1215.

[38] SWEENEY L. k-anonymity: A model for protecting privacy[J]. International journal of uncertainty, fuzziness and knowledge-based systems,2002,10(5): 557-570.

[39] El EMAM K,DANKAR F K. Protecting privacy using k-anonymity[J]. Journal of the American Medical Informatics Association,2008,15(5): 627-637.

[40] El EMAM K,DANKAR F K,ISSA R,et al. A globally optimal k-anonymity method for the de-identification of health data[J]. Journal of the American Medical Informatics Association,2009,16(5): 670-682.

[41] RAJENDRAN K,JAYABALAN M,RANA M E. A study on k-anonymity,l-diversity, and t-closeness techniques[J]. IJCSNS,2017,17(12): 172.

[42] DWORK C. Differential privacy[C]//International colloquium on automata, languages, and programming. Berlin, Heidelberg: Springer Berlin Heidelberg,2006: 1-12.

[43] CORMODE G,JHA S,KULKARNI T,et al. Privacy at scale: Local differential privacy in practice[C]// Proceedings of the 2018 International Conference on Management of Data. New York: Association for Computing Machinery,2018: 1655-1658.

[44] NAVIDAN H,MOGHTADAIEE V,NAZARAN N,et al. Hide me behind the noise: Local differential privacy for indoor location privacy[C]// 2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). Genoa: IEEE, 2022: 514-523.

[45] KAEMARUNGSI K,KRISHNAMURTHY P. Modeling of indoor positioning systems based on location fingerprinting[C]// Ieee Infocom 2004. Hong Kong: IEEE,2004,2: 1012-1022.

[46] SPANA S,DU L. Optimal information perturbation for traffic congestion mitigation: Gaussian process regression and optimization[J]. Transportation Research Part C: Emerging Technologies,2022,138: 103647. .

[47]ERROUNDA F Z,LIU Y. Continuous location statistics sharing algorithm with local differential privacy[C]// 2018 IEEE International Conference on Big Data (Big Data). Seattle: IEEE,2018: 5147-5152.

[48] PINKAS,BENNY. Cryptographic techniques for privacy-preserving data mining[J]. Acm Sigkdd Explorations Newsletter,2002,4(2):12-19.

[49] 李晓楠. 大数据技术下个人公平征信监管的数据治理维度[J]. 大连理工大学学报(社会科学版),2023(2): 65-74.

[50] 姜晓萍,杨舒雯. 速度与法度:疫情防控中智能应急管理的权力规范与权利保障[J]. 理论探讨,2022(4):83-92.

[51] 赫广平,孟昭阳. 人工智能时代警察执法行为创新:基于个人信息处理的视角[J]. 中国人民公安大学学报

(社会科学版),2022(3):129-139.

[52] 龚善要.教育场景中情感计算的应用风险及其法律规制[J].复旦教育论坛,2022(6):40-46.

[53] 刘绍宇.数字政府建设中个人信息保护的风险规制路径[J].财经法学,2023(2):51-67.

[54] 叶佩源,狄小华.数据正义与个人信息的程序性保护[J].中国特色社会主义研究,2023(2):54-61.

[55] 邢会强.政务数据共享与个人信息保护[J].行政法学研究,2023(2):68-81.

[56] 罗力.上海市民个人信息安全素养评价研究[J].重庆大学学报(社会科学版),2013(3):95-99.

[57] 胡尔西旦·卡哈尔,刘文轩.公共图书馆读者个人信息保护的原则与进路:参照《个人信息保护法》[J].新世纪图书馆,2023(2):36-41.

[58] 钱继磊.何以个人信息权为新兴(型)人权:人工智能与大数据新时代背景下的思考[J].北方法学,2023(2):5-14.

[59] 贾传昌,朱建明,高胜.隐私经济学研究进展[J].经济学动态,2022(3):139-157.

[60] 王仲羊.刑事诉讼中个人信息的权利保护[J].中国刑事法杂志,2022(3):155-176.

[61]相丽玲,李彦如."总体国家安全观"下的我国个人信息保护制度体系探析[J].情报理论与实践,2020(7):18-23.

[62] 王苑.数据权力视野下个人信息保护的趋向:以个人信息保护与隐私权的分立为中心[J].北京航空航天大学学报(社会科学版),2022(1):45-57.

[63] 沈俊翔.数字经济时代个人信息跨境保护的机制研究:兼论 CPTPP 视野下人民法院参与全球数据治理的新型路径[J].法律适用,2022(6):174-184.

[64] 齐爱民.区块链环境中个人信息保护的法律障碍与应对[J].现代法学,2022(5):180-193.

[65] 黄平平,刘文云,孙志腾.基于 ISM-MICMAC 模型的政府数据开放中个人隐私保护影响因素分析[J].情报理论与实践,2022(3):65-71.

[66] 孙辛未,张伟,徐涛.面向云存储的高性能数据隐私保护方法[J].计算机科学,2014(5):137-142.

[67] 杨海芳,王明征.基于最小化信息损失的用户隐私保护方法[J].系统工程理论与实践,2021(2):483-497.