

大数据背景下一门交叉学科的兴起*

——论数据计量学的构建

邱均平 余波 杨思洛

摘要 数据计量学是大数据背景下数据科学和计量学融合,而新兴的一门交叉学科。通过对异构、多源和海量数据进行有效的计量和统计分析,从而深入挖掘数据的知识价值和资源价值,丰富计量学的学科体系,促进数据科学的发展,这是数据计量学的研究目标所在。本文在阐述数据计量学的产生背景、研究现状、学科概念和内容体系的基础上,探讨数据计量学的发展趋势和有待突破的问题。认为数据计量学的内容体系由理论、方法和应用三个层面组成,呈现出学科融合、定量化、实用化、智能化的发展趋势,具有广阔的发展前景。参考文献 32。

关键词 大数据 数据科学 数量计量学 交叉学科 学科构建

分类号 G250.2

The Rise of Interdisciplinary Fields under the Background of Big Data: A Discussion of the Construction of Datametrics

QIU Junping, YU Bo & YANG Siluo

ABSTRACT

Datametrics is a new interdisciplinary subject. Effective measurement and analysis of heterogeneous, multi-source, and massive data, in-depth exploration of the knowledge value and resource value of data, enriching the discipline system of metrology, and promoting the vigorous development of the digital economy are the research goals of Datametrics. For different disciplines, the arrival of the big data era brings both opportunities and challenges. Big data will transform the research paradigm of traditional disciplines, and it will also bring a new research paradigm—the data paradigm to the development of various disciplines. At present, big data has become the focus of attention in various disciplines, especially in computer, statistics, management, and other disciplines. The fields of computer science and statistics have performed particularly prominently in big data research, and they even fought for the ownership of the discipline of data science spawned in the era of big data.

* 本文系国家自然科学基金重大项目“基于大数据的科教评价信息云平台构建和智能服务研究”(编号:19ZDA348)和国家自然科学基金一般项目“我国‘五计学’融合与图书情报学的方法创新研究”(编号:18BTQ080)的研究成果之一。(This article is an outcome of the major project “The Research on the Construction of A Cloud Platform for Science and Education Evaluation and Intelligent Service Based on Big Data” (No. 19ZDA348) and the project “The Research on the Integration of ‘Five Metrics’ and Method Innovation of Library and Information Science in China” (No. 18BTQ080) supported by National Social Science Foundation of China.)

通信作者:余波,Email:yubo@hdu.edu.cn,ORCID:0000-0001-8234-279X (Corresponding should be addressed to YU Bo,Email:yubo@hdu.edu.cn,ORCID:0000-0001-8234-279X)

From Bibliometrics to the development of Datametrics, Data Science, and Five Metrics, the relationship and course between them is a process of mutual connection and continuous development. With the development of big data, the trend of data research in the discipline of library and information science is highly integrated with other disciplines, and new discipline-data metrology has emerged.

Datametrics is an interdisciplinary field that has emerged in recent years. It mainly studies the measurement and analysis of data and the process of data measurement and analysis by extracting data from various structured or unstructured data. It is the development and extension of statistics, computer science, data analysis, machine learning, and data mining in the field of data analysis. At the same time, in the era of data intelligence, data has the advantage of quantitative analysis. Many qualitative and continuous objects are displayed in the form of data for easy analysis and processing.

On the basis of explaining the construction background, research status, subject concept, and content system of Datametrics, this paper profoundly discusses the future development trend and broad prospects of Datametrics, and proposes that Datametrics presents discipline integration quantification and practicality, and intelligence. It points out the problems and directions for the further development and breakthrough of Datametrics. It is a brand-new interdisciplinary and frontier subject emerging under the background of big data, and it has a good development prospect. 32 refs.

KEY WORDS

Big data. Data science. Datametrics. Interdisciplinary. Disciplinary construction.

在当今信息和数据时代,大数据技术蓬勃发展,人工智能应用快速渗透,开放科学方兴未艾,各行各业的数据呈海量增长,大数据的 Volume(大量)、Variety(多样)、Velocity(高速)、Value(价值)等特征日益凸显,给人类社会、经济、科技和文化等各个领域的发展都带来了巨大影响和深刻变革。人们的日常生活、工作学习及思维模式等正在不断被大数据改变^[1]。大数据也影响着传统的科学研究,对科研方式、科学交流、政策制定等产生了深远影响,形成了新的研究范式——第四范式:数据密集型科学发现^[2]。对不同的学科而言,大数据既带来了机遇,也带来了挑战;各学科的基础理论和技术方法如何适应这种革命性变化,是值得我们深入思考和亟需解决的重大课题。

目前,大数据已经成为众多领域关注的焦点,特别是在计算机科学、统计学、管理学等学科领域表现尤为突出,都在各自争取大数据时代催生出的数据科学的主动权。数据有定量分

析的天然优势,很多定性和连续的对象,也都通过数据的形式展示,便于计量分析和处理。文献计量学、科学计量学与信息计量学,与数据科学和大数据有着天然的紧密联系,在数据智能时代,“五计学”也在不断更新,在新时代、新环境、新背景下,催生了一门新兴的交叉学科——数据计量学。数据计量学从大数据的特点出发,探究新环境与技术产生的影响,将变革传统计量与评价活动,促进计量与评价活动的精准化、自动化和智能化^[3]。从文献计量学、科学计量学、信息计量学、知识计量学、网络计量学到数据计量学的发展,它们是一个相互联系又不断发展替代的过程。

基于此,本文对数据计量学的产生背景、研究现状、内容体系和发展趋势等进行较全面的探讨,分析大数据、数据科学和计量学的产生及其发展关系,在此基础上构建数据计量学的内容体系,探析数据计量学的发展趋势,以期数据科学和计量学领域的发展提供借鉴和参考。

1 数据计量学的产生与发展

1.1 数据计量学的产生背景

1.1.1 大数据背景

数据计量学的产生大致经历了从数据到大数据,从大数据到数据科学,以及与计量学深度融合发展的过程。

(1)从数据到大数据。数据通常与信息、知识联系在一起,人们获取数据是为了更好地获得信息和知识,并为自己所用。自从有了人类,数据就开始在人们的生产生活中发挥着重要的作用。例如,国外早期对人口数据的统计分析;国内几千年前,甲骨文符号的使用,也是一种记录客观事实的数据。一般而言,数据是用来对客观事件或事物的性质、状态及相互关系等信息进行记录的符号或符号组合,数据具体可以是数字、图像、文字、符号,也可以是声音、视频、实时记录或计算机代码^[4]。从某种意义上而言,数据不具有语义内涵,其内涵只有在特定背景和规则下才能体现。在图书情报学领域,数据是研究信息和知识的起点,美国乔治亚工业学院的斯拉麦卡教授认为“情报是有用的数据或被认为有用的数据”。

数据在最初是不分小数据和大数据的,但是在人类的持续进步和社会的不断发展中,数据在经济、政务、医疗和公共服务等方面一直发挥着重要的作用。进入21世纪后,计算硬件、网络 and 大规模并行处理框架的快速发展,以及计算基础设施成本的降低,促成大量数据的快速生成,网络环境下开始出现了海量的数据,“信息大爆炸”一词油然而生,对企业和个人驾驭数据的能力提出了新的挑战^[5]。从2012年开始,大数据(big data)的概念被普遍提及,人们用它来描述信息爆炸时代产生的海量数据。随后,大数据的内容出现在《纽约时报》《华尔街日报》的专栏封面,进入了美国白宫官网的新闻,世界主要国家纷纷制定大数据战略,大数据产业快速发展。最早提出大数据时代到来的是全球知

名咨询公司麦肯锡:“数据,已经渗透到当今每一个行业和业务职能领域,成为重要的生产因素;人们对于海量数据的挖掘和运用,预示着新一波生产率增长和消费者盈余浪潮的到来”。随着互联网的应用越来越广泛,新型社交网络平台和工具不断出现和快速发展,以智能手机、平板电脑为代表的新型移动设备改进升级,计算机网络技术更新换代,大数据受到人们越来越多的关注。2012年末出版的著作《大数据时代》强调,大数据正在改变人们的生活以及理解世界的方式,而更多的变革正蓄势待发。例如,美国总统奥巴马的成功竞选及连任的背后都有大数据挖掘技术及相关成果的支撑。

(2)从大数据到数据科学。随着大数据的发展,海量数据的搜集、处理、分析和管理工作成为人们高度关注的问题,催生了数据科学。数据科学大致历经了诞生、涵义演变、专业发展、广泛应用等四个阶段^[6]。数据科学是一门跨多个学科领域的交叉学科,包含了所有结构化和非结构化数据研究的相关内容,结合数学、计算机科学、统计学等学科知识,研究数据的获取、清理、分析和保存等内容。简而言之,数据科学主要是通过不同技术、算法和流程对数据进行分析 and 处理,从而获得有价值的信息和知识。具体可通过数据分析,使管理部门做出科学合理的决策。例如通过特定数据分析,推荐系统可以预测用户可能愿意购买的某个商品;数字广告比传统广告在展示方面显得更为逼真,为用户展现了更全面和具体的效果;网络搜索让用户能在最短的时间内找到与之相关的结果,全面掌握所需信息。

从时间维度看,数据科学是继“大数据”和“云计算”之后出现的另一个新概念,主要面向大数据时代的新挑战、新机会、新思维和新方法,以实现数据资源和能量之间的深层转换为研究目的,包括新的理论、方法、模型、技术、平台、工具、应用和最佳实践在内的一整套知识体系。从空间维度看,数据科学是统计学、机器学习、数据可视化以及某一学科领域的知识相互

融合后形成的一门新兴学科,其主要研究内容包括数据科学基础理论、数据加工、数据计算、数据管理、数据分析和数据产品开发,数据科学的兴起是云计算与大数据应用的必然结果^[7]。

1.1.2 理论背景

数据计量学基于数据科学理论和计量学理论,融合了数据与计量的两个范式。随着环境的变化和学科发展,数据计量学呈现出交叉融合和不断发展的趋势。

(1)数据科学理论。从计算机科学和技术领域角度,数据科学主要是以统计学、机器学习、数据可视化及领域知识为理论基础,包含数据科学基础理论、数据加工、数据计算、数据管理、数据分析和数据产品开发等^[8]。此外,还涉及大数据模因、数据挖掘、数据安全等与数据相关的诸多理论。数据科学的发展相当于统计和机器学习领域的超集。从某种意义上而言,数据科学理论的目的是为挖掘数据、计算数据、分析数据和管理数据等实践提供指导和支撑。因此,数据科学理论将影响整个科学领域数据分析实践的有效性。

(2)计量学理论。从统计书目学、文献计量学、科学计量学、信息计量学,再到知识计量学、网络计量学、替代计量学等,包括系列基本理论。其中文献信息流静态特性是指在一定时间内科学文献在空间的分布性质,包括布拉德福定律、齐普夫定律、洛特卡定律;动态特性是指科学文献随时间的延续而增长和老化的性质,包括文献信息增长规律、文献信息老化规律等。例如,布拉德福定律应用相对较广,应用在核心期刊的评定、辅助文献检索、考察专著分布、动态馆藏的维护、检索工具完整性的测定、学科幅度的比较、指导读者利用期刊、指导期刊订购工作和利用馆藏等方面^[9]。此外,还有加菲尔德在引文数据分析方面的规律性结论。传统计量学理论为数据计量学的产生和发展提供坚实的理论和实践基础。

数据科学理论和计量学理论已体现出交叉发展趋势,如数据分析、数据可视化、数据预测

等相关理论在数据科学和计量学中已广泛应用。进入21世纪后,特别是Web2.0、数据挖掘、深度学习、云计算等理念和技术的应用,为数据计量分析提供了更为广阔的发展空间,从而促进了数据计量学的形成和发展。

1.1.3 实践背景

数据分析与应用需要统计和计量。随着网络技术的不断发展,数字化、网络化不断加速,结构化与非结构化数据、超文本数据、多媒体数据等多样化数据大量出现,这些特点使得用户处在一个海量的数据环境中。如何通过定量的方法对数据进行科学的计量和管理,提高数据的使用效率,是一个亟待解决的理论问题和现实问题。大数据的特点决定了需要进行大量的数据处理、分析和挖掘才能获得其背后的信息和价值。计量指标是数据统计分析与应用实践的重要路径,越来越受到人们的高度关注。数据计量指标的构建涉及数据本身的质量,包括数据处理过程与数据效能的质量,需要对大数据背景下的数据质量进行全面的量化评估^[10]。另外,数据的影响渗透到人类社会各个方面,如何对数据进行管理和利用已成为国家和政府管理部门十分关注的重要课题。数据计量学就是在这样一个社会和技术背景下诞生。对数据进行定量描述和分析研究,为规范化和科学化管理和利用数据提供必要的定量依据,从而能够充分发挥数据的价值,而数据的定量化和管理工作将会促进数据计量学的全面发展。数字化、网络化为数据计量学的产生提供了必要的基础和条件,而且还将产生实践中的需求,进一步促进学科的形成和发展。

1.1.4 学科背景

数据计量学的产生是“五计学”在新时代发展的客观需要。数据计量学的兴起离不开计量学的发展,计量学与数据科学有着广泛的交叉和融合,计量学对数据科学的发展发挥着重要的作用。文献计量学的形成和发展大致经历了萌芽时期、奠定时期、发展时期和全面发展时期。信息计量学是在传统文献计量学和科学计

量学的基础上演化、扩展而成的。信息计量学是通过量的概念和定量分析方法对信息单元的数量变化规律和体系结构进行进一步的揭示,信息单位包含文献、数据、实物、消息和事件等^[11]。

大数据的到来推动了科学研究第四范式的出现,这种以数据为导向的科学研究方式彻底改变了科学研究的思维与方法,同时为科学交流带来了巨大变革;这些改变也推动着科学研究发生变化,迫切需要我们改变思维,完善政策革新理论,推出新的计量方法、流程和工具,以应对大数据的浪潮。大数据环境的影响是巨大的,传统计量与评价需要汲取大数据的特点来推动变革,比如计量对象从题录数据到全文本数据;大数据环境也呼吁更为科学的数据计量理论和方法为新的科学范式服务。因此,数据计量学从本质上说,是“五计学”在新的数据网络时代不断发展的结果。

1.2 数据计量学的研究现状

目前,国内外学者对数据科学的研究较多,涉及专门的数据科学领域研究和在各学科的应用研究^[12]。关于数据计量的研究主要集中在数据统计分析方面的探索,涉及计算机科学、信息科学、统计学及数学等领域。在国外数据科学研究领域,数据统计分析是关注的重点之一。在教育科学领域,有学者提出应研究课程分配的专门数据分析,认为其价值取决于对教学数据统计分析的直接或间接受益程度^[13]。在科学传播领域,科学期刊反映了数据计量的理论研究趋势,有学者利用在线数据库中不同学科的数据研究科学期刊,结合定量和定性方法识别期刊中的数据类型和范围^[14]。在医学领域,在新型冠状病毒研究中,有学者利用病例和死亡的时间序列横截面数据,发现相关数据统计结果在 COVID-19 的测量、推断和解释中发挥着重要作用^[15]。在医学领域,数据统计和机器学习正在彻底改变麻醉学的研究,通过数据计量可更准确地作出预测,通过数据关联分析,为揭

示未知的相关性知识提供了巨大的潜力^[16]。

近年来,以大数据为基础的计算社会科学、各种数据智能算法正在影响数据计量的进一步发展。同时,经济学、社会学和政治学也越来越重视数据的定量分析和数学模型方法,数据计量的发展呈现多学科交叉趋势。在图书情报研究领域,孟阳等认为数据计量需要根据自身特点发挥特有作用,需从其他计量学中吸取经验;数据计量是大科学时代的产物,从科学文献到科研数据,再到数据计量,反映了人类科学意识及科学素养的提高^[17]。顾立平认为数据级别计量是指在数据共享的背景下,主要与数据发布和数据引用相关的计量;“数据引用”不仅采用文献之间的参考文献,而且扩展到数据和文献、数据和数据集、数据和数据之间的多重关系,而且更进一步关注“数据”“学术记录”以及与学术记录紧密相关的“个人”^[18]。侯经川等认为,“数据引证”问题日益受到关注;数据引证实践在数据集标识系统建立等问题上取得了进展,统一规范化数据引证格式的趋势日益清晰;对数据引证现状的评估和研究进展的追踪,数据引证索引的编纂、指标体系研究、数据引证数据库的建立,以及基于数据引证行为、记录和索引的分析,应是未来重点突破的方向^[19]。另外,也有学者利用专利数据计量知识创造的方法,比较专利数量、专利被引率及被引加权专利数量,认为汽车生产供应链上下游企业可采用统一的计量方法表示知识创造^[20]。目前,国内外已有较多关于数据计量与统计分析的探索性研究,但还未发展成为系统的专门的学科。随着计算机网络技术的不断发展,人们对数据计量的研究越来越重视,数据计量将会影响各个学科,并促进各学科交叉融合、深入发展。

2 数据计量学的内容体系构建

2.1 数据计量学的概念和研究对象

作为数据科学与计量科学的交叉领域,数据计量学的主要研究对象是数据及数据资源。根

据数据科学与计量科学的原理以及交叉融合的事实,可以将数据计量学定义为:采用数学、统计学和计量分析方法,研究各类型数据的功能和特征、分布结构和变化规律,充分挖掘数据的知识价值和资源价值,一方面可以促进数据科学的量化发展,丰富数据科学和计量科学领域的学科体系,另一方面,有利于数据管理与决策的科学化和规范化,推动数字经济和社会的健康发展。数据计量学是在大数据背景下涌现的具有明显的创新性、交叉性、定量性的新兴前沿学科,具有广阔的发展前景。因此,数据计量学的产生和发展具有重要的科学理论价值和实际应用意义。

数据计量学聚焦于数据的统计分析和定量研究,揭示数据的内在结构特征和变化规律,从理论上提高数据分析的科学性和准确性,促进数据科学的发展,同时应用于各学科领域,为经济和社会发展提供服务。数据计量学的理论、方法是对各行各业数据管理和应用实践经验进行高度浓缩和提炼而形成的。其理论来源于数据资源的管理和实践,但高于实践,从而能指导数据科学的应用。数据计量学的基本问题包括:数据概念的定量描述,内容、范围及与相关学科的关系,数据的基本测度,数据计量模型,数据计量化方法的探讨,数据计量方法和工具的智能化实现,在不同学科领域的应用,等等。数据计量的研究对象主要涉及数据及数据资源。我们认为,对数据计量学的研究对象应从广义上理解。这里所讲的“数据”,不仅是指数字数据,也包括各种类型的文本、图像数据;既包括结构化的数据,也包括半结构数据和非结构化的数据;既包括数据单元的计量问题,也涉及数据单元间,如数据传递,单元间的相互关联等的计量问题。由此可见,数据计量学涉及范围很广,内容也非常丰富。

2.2 数据计量学的研究内容

数据计量学的内容体系可由理论、方法和应用三个层面组成。理论是基础、方法是手段、应用是目的,这三个部分共同作用、相互影响,

成为一个整体。

2.2.1 数据计量学的基本理论

基本理论是一个新兴学科的关键内容,数据计量学的基本理论主要包括数据计量学的概念、研究对象、内容体系、学科体系等问题。数据计量学与数据科学、计量学、计算机科学等学科之间有着密切的关系。首先,数据计量学的基本理论来源于数据科学,数据科学的内涵、发展历程、学科地位及知识体系等都为数据计量学提供基础。数据科学是一门以数据为中心和基于数据处理的科学^[21],可分为大数据及其运动规律的揭示、大数据的资产化利用、基于数据的行动范式、数据产品的研发、供给侧改革和数据生态系统的建设^[6]。其次,计量学领域的三大定律和两大规律是数据计量学重要的理论基础,从已有的文献分析发现,数据科学相关理论和“五计学”、信息管理学等学科的理论高度交叉和融合,为数据计量学奠定了丰富的理论基础。此外,数据计量学融合计算机科学和统计学的理论,如机器学习、数据可视化、数据挖掘等相关理论。因此,数据计量学吸收“五计学”、数据科学、计算机科学和统计学的基本理论,融合形成数据计量学独特的基本理论体系。

2.2.2 数据计量学的方法研究

如上文所述,数据计量学是理论、方法、应用三者结合的理论体系,方法作为理论和应用之间的桥梁,其重要程度不言而喻。理论通常被视为方法的基础,但是时常发生的方法突破带来科学理论突破的事例,也说明了方法对学科理论建立和发展的重要性。方法在计量学科应用中的重要性,还体现在离开方法数据计量学的研究很难进行。

一般来说,现代科学的方法论体系可分为哲学方法、一般科学方法、专门科学方法三个层次。哲学方法在数据计量研究中起到指引方向的作用,数据分析之前主题的选取、问题的提出,数据分析之后对结果的认识和总结,以及对未来发展的思考和展望,都需要哲学思辨方法。从一般科学方法来看,数据计量学关注数据定

量分析方面的问题,计量学和数据科学的研究方法在其中交叉和融合,构成了数据计量学研究方法体系的基础。从专门科学方法来看,数学、统计学、计算机科学等学科方法是其重要来源,具体包括数据计算、数据挖掘、数据处理、数据模型等方法,也涉及机器学习、计算智能、数据可视化等方法,另外计量学中的引文分析方法、内容分析方法、知识可视化方法等也融入其中,不断丰富数据计量学的方法体系。

2.2.3 数据计量学的应用研究

数据计量学为满足社会实践需要而产生,相关理论和技术可以指导社会实践,也将在实践中得到检验。大数据背景下数据科学和计量学不断发展,并在实践应用需求的推动下不断完善和成熟。数据存在于各学科和社会各领域,数据计量学也将广泛应用于各学科领域和社会经济生活中,如应用在管理学、经济学、计算机科学、传播学等学科,以及在社会管理、经济分析、计算机技术应用、网络分析、舆情监控等实践领域。在图书情报领域,主要应用于科学评价、信息资源管理、信息检索与服务、科技管理与预测等诸多方面。

数据计量学的应用,需要融合不同学科的理论和方法,通过不同层面的应用来检验该学科理论和方法的正确性,从而不断推进应用层面的发展和完善。数据计量学自身也可以通过优化现有学科体系,研究解决方案,改进技术方法,使得研究过程和实践应用适应学科发展的需求,并不断突破学科发展的局限。

2.2.4 数据和数据资源研究

数据是一种资源,也是一种经济要素。随着数字化、数据化时代的到来,数字经济的意义也日益凸显。通过数据计量分析,数据和数据资源可为推动数字经济发展发挥重要的作用。

数据和数据资源是数据计量学的主要研究对象,数据计量学也专门对二者进行过系统研究。其中,数据研究的内容包括数据的基本定义与概念、数据的发展与演化、数据的基本计量指标、数据内容的测度、数据结构的测度、数据

使用规律的统计等;数据资源研究的内容包括数据资源的基本内涵、数据资源配置计量问题、数据资源价值的定量评价、网络数据资源的定量分析、数据资源共享的计量、数据资源的成本效益定量评价等内容。因此,数据计量学能通过对经济数据的分析,挖掘数据资源的价值,为经济发展提供数据管理和决策服务^[22]。

3 数据计量学的发展趋势

对特定领域趋势的把握是所有学科普遍关注的关键性问题,通常可以通过咨询专家、参加前沿讲座、参加专业会议交流、阅读专业期刊等多种途径获取^[23]。大数据环境下诞生的数据计量学是一个变化发展着的新兴领域,处于萌芽和初步探索阶段,众多问题需要解决和突破;其相关研究进展也日新月异,研究内容纷繁复杂,涉及的学科和研究领域也十分广泛。基于定性与定量相结合的系统性思考,我们认为数据计量学的研究主要呈现学科融合、量化、实用化和智能化等发展趋势。

3.1 学科融合

学科融合是不同学科之间或同一学科门类不同子学科之间相互渗透、交叉到逐步形成新学科的动态发展过程。随着经济和科学技术的发展,学科交叉现象比较明显,数据计量学正迈向多学科融合的进程,具体包括图书情报学、数据科学、计量学、经济学、计算机科学等各学科的交叉融合^[24]。

大数据的涌现和计算机网络技术的发展,给数据计量学的研究模式带来了机遇和挑战。从数据计量学研究方法、技术工具、过程对象、实践应用等方面来看,数据计量学内部的融合、与数据科学和计量学科的融合、与所有学科领域的交叉融合,已成为一种趋势。例如,针对大数据环境下的安全情报研究,明确安全情报融合的基本内涵,分析安全情报与数据计量融合的意义及特点,构建并解析大数据环境下安全情报融合的核心分支体系^[25],是安全管理领域

与数据计量学共同关注的内容,体现了“数据计量情报”的发展趋势。学科整合中分支学科的形成和发展,不仅突破了计量学和数据科学的局限,大大拓宽了数据计量学的研究领域,还从多个侧面丰富了数据计量学的研究内容,完善了研究方法。各个领域的数据分析面临的问题、方法和工具基本是相通的,如共同面临数据的存储、数据的处理和分析、数据的计算等问题。可见,面对复杂性、综合性及交融性等重要问题,各学科不再通过单一的理论和方法解决各自领域的问题,而是进行全方位的综合探讨。因此,数据计量学带来的融合发展将成为解决人类发展重大问题的重要研究范式之一。

3.2 定量化

定量化是数据计量学显著的特征,也是该领域学术研究和实践应用的特色。在21世纪的信息爆炸时代,计算机技术和网络技术飞速发展,极大地提高了数据的采集和存储能力,为定量化研究奠定了基础。

进入大数据时代,数据来源多样化,有业务记录数据、传感器数据、网上爬取数据等,数据格式也呈现出结构化和非结构化、半结构化等多样性^[26],这些都为数据计量学的定量化研究和应用提供了良好的数据基础。近年来,国内外对数据的定量分析研究日益突出^[27],应用领域也不断扩展。国内的实践方面,图书情报领域的应用较早,主要应用在图书馆数字资源分析^[28]、用户行为分析以及学术成果评价研究等方面。另外,一个值得关注的现象是,数据计量研究在科学大数据领域引起广泛关注。在该领域,普赖斯是科学史定量研究的代表,他继承和发展了基于历史统计的科学史定量研究方法。近年来,在普赖斯研究的基础上,创新了科学数据计量研究的理念和方法,利用最新的统计分析技术,对科学大数据进行计量分析,取得了丰硕的成果^[29],同时也让我们看到了数据计量学广阔的发展前景。

3.3 实用化

实用化是大数据时代学科发展的基本趋势之一,数据计量学的实用性非常明显。相对于理论性学科和专门性的领域,数据计量学具有较好的普适性,更关注于研究结果的实际应用价值。由于计量的数据和研究成果比较真实可靠,可以为诸多部门的管理和决策提供依据。尤其是在国际经济迅速发展和激烈竞争的背景下,数据计量分析已逐渐进入到决策支持过程中,对国家、企业、个人发展都发挥着至关重要的作用^[30]。在图书情报领域,数据计量成果可为图书馆事业发展提供直观的数据参考;数据计量指标可用于科研成果质量、科研机构乃至整个国家的科技水平的评价;数据计量学研究结合知识可视化技术,结合实证分析的方法,广泛应用于信息资源管理、科学评价、信息检索等领域。随着数据计量理论和实践的成熟,它将会在更多的领域发挥作用。

3.4 智能化

数据与智能有着天然的紧密联系,在大数据与人工智能时代,更需要数据计量的智能化发展。

随着大数据、云计算、区块链、人工智能的迅速发展,数据计量的智能化趋势显得尤为突出。智能化是数据计量分析的高层次阶段,不仅满足于逻辑推理、定量计算或固定程序的需要,还具备灵活的分析判断能力,以及多路推理和模糊识别的能力。智能化的数据计量分析将综合利用人工智能、认知计算、数据挖掘、语义网等新技术,采用关联、序列、聚类、分类等方法从大量完整、彼此关系不明确的敏感性信息中找出隐含的、事先未知的有用信息,揭示数据内在的复杂性,进行深层次的分析,自动获得更多、更有价值的信息。在实践方面,近年来智能化技术与云计算技术相结合,基于云计算的智能化数据计量分析的优越性越发凸显,在实际工作中成效显著^[31]。

4 结语

数据计量学是一个新兴的跨学科领域,主要定量研究各类数据和数据资源,对结构化、半结构化或非结构化的数据进行计量和统计分析。数据计量学是统计学、计算机科学、数据科学、计量学等领域的交叉发展和延伸。随着信息技术和通信技术的快速发展,人们生活和社会生产的各类数据越来越容易

获取,为数据计量学的发展奠定了基础。数据计量学与信息科学、数据科学、计量学等在理论逻辑及技术方法上一脉相承,在数据—信息—知识—智慧链中,组成了数据—信息—知识—计算技术—数学方法—专业知识、人—技术—数据三个“三位一体”^[32]。目前,数据计量学还处于初步探索和创建的阶段,基本理论、技术方法、实践应用等都存在诸多问题需要解决,这些都是学科未来发展需要不断改进和完善的方面。

参考文献

- [1] Espinal Y. Big data: a revolution that will transform how we live, work, and think[J]. Research Technology Management, 2015,7:2727-2729.
- [2] Tony Hey, Stewart Tansley, Kristin Tolle, 等. 第四范式:数据密集型科学发现[M]. 科学出版社, 2012. (Hey T, Tansley S, Tolle K, et al. The fourth paradigm: data-intensive scientific discovery[M]. Science Press, 2012.)
- [3] 邱均平, 柴雯, 马力. 大数据环境对科学评价的影响研究[J]. 情报学报, 2017,36(9):871-877. (Qiu Junping, Chai Wen, Ma Li. Research on the influence of big data on scientific evaluation[J]. Journal of the China Society for Scientific and Technical Information, 2017,36(9):871-877.)
- [4] 马费成. 信息管理学基础[M]. 科学出版社, 2005. (Ma Feicheng. Fundamentals of information management [M]. Science Press, 2005.)
- [5] GIL D, Song I Y, Aldand J F, et al. Big data: new approaches of modelling and management[J]. Computer Standards & Interfaces, 2017, 54 (11):61-63.
- [6] 聂淑媛. 数据科学的发展与人才培养研究[J]. 统计与信息论坛, 2019, 34 (1): 117-122. (Nie Shuyuan. Study on development and talent training of data science[J]. Statistics & Information Forum, 2019, 34 (1):117-122.)
- [7] 朝乐门, 卢小宾. 数据科学及其对信息科学的影响[J]. 情报学报, 2017,36(8):761-771. (Chao Lemen, Lu Xiaobin. Data science and its implications on information science[J]. Journal of the China Society for Scientific and Technical Information, 2017,36(8):761-771.)
- [8] 朝乐门. 大力推进数据科学的理论研究[J]. 计算机科学, 2019,46(2):1. (Chao Lemen. Vigorously promote the theoretical research of data science[J]. Computer Science, 2019,46(2):1.)
- [9] 邱均平. 信息计量学[M]. 武汉大学出版社, 2007:88-110. (Qiu Junping. Informetrics[M]. Wuhan University Press, 2007:88-110.)

- [10] 刘金晶,王梅. 大数据下的数据质量评价指标构建实践[J]. 计算机技术与发展, 2019, 29(10): 46-50.
(Liu Jinjing, Wang Mei. Practice of data quality evaluating index construction under big data[J]. Computer Technology and Development, 2019, 29(10): 46-50.)
- [11] 宋艳辉,邱均平. 从“三计学”到“五计学”的演化发展[J]. 图书馆论坛, 2019, 39(4): 1-7. (Song Yanhui, Qiu Junping. The evolution from “Three-metrics” to “Five-metrics” [J]. Library Tribune, 2019, 39(4): 1-7.)
- [12] 朝乐门,邢春晓,张勇. 数据科学研究的现状与趋势[J]. 计算机科学, 2018, 45(1): 1-13. (Chao Lemen, Xing Chunxiao, Zhang Yong. Data science studies: state-of-the-art and trend [J]. Computer Science, 2018, 45(1): 1-13.)
- [13] Cleveland W S. Data science: an action plan for expanding the technical areas of the field of statistics[J]. Statistical Analysis and Data Mining: The ASA Data Science Journal, 2014, 7(6): 414-417.
- [14] Hong L, Moen W, Yu X, et al. The disciplinary research landscape of data science reflected in data science journals[J]. Information Discovery and Delivery, 2020.
- [15] Douglass R W. The data science of COVID-19 spread: some troubling current and future trends[J]. Peace Economics Peace Science and Public Policy, 2020, 26(3): 1-7.
- [16] Chae D. Data science and machine learning in anesthesiology[J]. Korean Journal of Anesthesiology, 2020, 73(4): 285-295.
- [17] 孟阳,屈宝强. 数据计量与文献计量之间的对比研究[J]. 情报理论与实践, 2017, 40(11): 139-144, 138.
(Meng Yang, Qu baoqiang. Contrastive study between data metrics and bibliometrics [J]. Information Studies: Theory & Application, 2017, 40(11): 139-144, 138.)
- [18] 顾立平. 数据级别计量——概念辨析与实践进展[J]. 中国图书馆学报, 2015, 41(2): 56-71. (Gu Liping. Data level metric: its concepts and progress [J]. Journal of Library Science in China, 2015, 41(2): 56-71.)
- [19] 侯经川,方静怡. 数据引证研究: 进展与展望[J]. 中国图书馆学报, 2013, 39(1): 112-118. (Hou Jingchuan, Fang Jingyi. Review on data citation in the context of big data [J]. Journal of Library Science in China, 2013, 39(1): 112-118.)
- [20] 林岩,陈剑. 以专利数据计量汽车零部件供应商的知识创造[J]. 科学学研究, 2008, 26(S1): 152-158.
(Lin Yan, Chen Jian. Using patent data measure components suppliers' knowledge creation in the auto manufacturing industry [J]. Studies In Science of Science, 2008, 26(S1): 152-158.)
- [21] Vasant D. Data science and prediction[J]. Communications of the Acm, 2013, 56(12): 64-73.
- [22] Online information seeking patterns and social inequality in a digital economy[J]. The International Review of Retail, Distribution and Consumer Research, 2021, 31(2): 211-228.
- [23] 杨思洛,王雨,祁凡. 系统视角下 Altmetrics 的发展趋势: 融合、开放、深化[J]. 情报理论与实践, 2020, 43(4): 32-39. (Yang Siluo, Wang Yu, Qi Fan. The development trend of Altmetrics from a systematic view: integration, openness and deepening [J]. Information Studies: Theory & Application, 2020, 43(4): 32-39.)

- [24] 孟小峰,杜治娟. 大数据融合研究:问题与挑战[J]. 计算机研究与发展,2016,53(2):231-246. (Meng Xiaofeng, Du Zhijuan. Research on the big data fusion: issue and challenges[J]. Journal of Computer Research and Development,2016,53(2):231-246.)
- [25] 黄玺,王秉,吴超. 大数据环境下安全情报融合体系构建[J]. 情报理论与实践,2020,43(10):39-44. (Huang Xi, Wang Bing, Wu chao. Building a system for safety & security-related intelligence fusion in big data environment[J]. Information Studies:Theory & Application;2020,43(10):39-44.)
- [26] 方匡南. 大数据时代统计学应拥抱数据科学[J]. 统计与信息论坛,2016,31(11):6. (Fang Kuangnan. Statistics should embrace data science in the era of big data[J]. Statistics & Information Forum,2016,31(11):6.)
- [27] 蒲显伟. 定量数据分析效应值研究综述[J]. 统计与信息论坛,2014,(11):18-22. (Pu Xianwei. A review of effect sizes in quantitative analysis[J]. Statistics & Information Forum,2014,(11):18-22.)
- [28] 谢紫悦,陈雅. 区域文化共享视域下我国双一流高校图书馆自建数据库现状定量分析研究[J]. 新世纪图书馆,2020(5):90-96. (Xie Ziyue, Chen Ya. Quantitative analysis and research on the status quo of self-built database in Chinese University libraries from the perspective of regional cultural sharing[J]. New Century Library, 2020(5):90-96.)
- [29] 董国豪,潜伟. 普赖斯与科学史定量研究[J]. 科学学研究,2017,35(5):667-675. (Dong Guohao, Qian Wei. Derek John de Solla Price and the quantitative study of the history of science[J]. Studies in Science of Science,2017,35(5):667-675.)
- [30] 骆云鹏,朱旋彤,毛慈伟,等. 一种基于连接关键词的实用化可搜索加密方案[J]. 计算机工程,2020,46(2):175-182. (Luo Yunpeng, Zhu Nitong, Mao Ciwei, et al. A practical searchable encryption scheme based on connection keywords[J]. Computer Engineering,2020,46(2):175-182.)
- [31] 苏宏庆. 基于云计算的自动化运维研究[J]. 大科技,2020(15):247-248. (Su Hongqing. Research on automated operation and maintenance based on cloud computing[J]. Super Science,2020(15):247-248.)
- [32] 叶鹰,马费成. 数据科学兴起及其与信息科学的关联[J]. 情报学报,2015,34(6):575-580. (Ye Ying, Ma Feicheng. Data science:its emergence and linking with information science[J]. Journal of the China Society for Scientific and Technical Information,2015,34(6):575-580.)

邱均平 杭州电子科技大学中国科教评价研究院资深教授、院长。浙江 杭州 310018。

余波 杭州电子科技大学中国科教评价研究院讲师。浙江 杭州 310018。

杨思洛 武汉大学信息管理学院教授,博士生导师。湖北 武汉 430072。

(收稿日期:2020-12-22;修回日期:2021-07-13)