

文献计量学驱动“双一流”农业研究生 科研知识体系的构建

何建强^{a*},和志豪^b,吴淑芳^a,冯浩^c,于强^b

(西北农林科技大学 a.水利与建筑工程学院 b.水土保持研究所 c.水土保持科学与工程学院,陕西 杨凌 712100)

摘 要 新时代农业高校的科技创新是促进我国“三农”事业健康高效发展的关键因素。开展农业科学研究工作,首要问题是要明晰本学科的研究方向、基本内容以及发展趋势,因此,培养农科研究生快速准确地了解本学科领域的时代变迁和未来发展趋势至关重要。文章旨在介绍一种基于编程和文献计量学的高效的科研知识体系,并希望以此指导相关研究生高效便捷地获取国内外文献,研究本学科领域的学术热点和变迁规律,继而在后续的科研工作中思路更清晰,创新性更强,成果应用和转化程度更高。研究有利于推动我国农业教育体系的建设,提升我国“双一流”农业高校教育的国际影响力。

关键词 农业教育 文献计量学 知识体系 “双一流”

中图分类号 G641 文献标识码 A 文章编号 1673-7164(2024)13-0101-05

2019年9月5日,习近平同志在给全国涉农高校的书记校长和专家代表的回信中说:“中国现代化离不开农业农村现代化,农业农村现代化关键在科技、在人才。”^[1]

新时代农业人才培养至关重要,以新型农业人才驱动新型农业科技发展是我国目前农业科技进步的重要路径之一。2022年中央一号文件《中共中央 国务院关于全面推进乡村振兴重点工作的意见》指出:“加强乡村振兴人才队伍建设。发现和培养使用农业领域战略科学家。启动‘神农英才’计划,加快培养科技领军人才、青年科技人才和高水平创新团队。”^[2]新型农业人才的培养目标是适应未来农业发展方向和掌握未来农业新型技术的创新型科技人才,因此,农业高校要不断深化农业科技创新型人才培养的

教学改革。

一、农业科研知识体系构建的原理

在我国农业高校的科学研究和教育教学工作中,科研教学人员主要以标准的科学研究理论和范式来教授和指导高校的研究生。首先,研究者需要基于唯物辩证法理论,通过自身的抽象思维能力和逻辑推理能力,思考和发现相关领域的科学问题。其次,需要对所选科学问题进行实践探索,传统的农业科学研究方式主要是进行科学试验或调研设计,并暂定解决该科学问题的可行性方案。最后,利用科学试验或社会调研等科学方式去验证其方案的客观现实性,若能解决该科学问题,则证明所设计的试验或调研具有科学意义上的准确性;反之则将该方案作为解决该科学问题

基金项目:西北农林科技大学2025年校级教育教学改革研究(一般)项目“从传统到智慧:中国农业水利思政元素挖掘与案例构建”(项目编号 JY250330) 西北农林科技大学2020年研究生教育教学改革研究(重点)项目“以多学科创新团队建设提高研究生科研水平和培养质量”(项目编号 JXGG2040)。

作者简介:何建强(1977—),男,博士,西北农林科技大学水利与建筑工程学院博士生导师,教授,研究方向为农业生态系统模拟;和志豪(1996—),男,博士,西北农林科技大学水土保持研究所,研究方向为农业生态系统模拟;吴淑芳(1977—),女,博士,西北农林科技大学水利与建筑工程学院博士生导师,研究员,研究方向为土壤侵蚀过程及水动力学;冯浩(1970—),男,博士,西北农林科技大学水土保持科学与工程学院书记,研究员,研究方向为水土资源过程模拟、调控与高效利用;于强(1962—),男,博士,西北农林科技大学水土保持研究所研究员,智慧农业、气候变化对农业的影响、生态过程与模拟。

的一次试错性措施,并基于之前的结果对试验或调研方案进行改进或重设,如此反复试错,直至获得最终解决该科学问题的标准理论或方法。从上述农业科学研究的基本范式可知,农业科学研究的核心在于发现本学科领域内的科学问题,并提出合理的解决方案,这通常是通过研读大量的科研文献资料来实现的,因此,及时、高效地搜集整理和消化吸收本学科领域科研文献的核心内容至关重要,是广大农业高校科教人员必须掌握的基本技能。

如今新的科学理论和技术不断影响着当代农业高校的科研和教学工作,例如人工智能技术和大数据应用的发展,使众多学科科学研究的方向产生了重大的变化和转型^[3]。传统的教学模式尽管能为高校人才提供经典的理论范式,但科学研究在新时代的提速发展,使得广大科研人员不得不及时主动获取最新的研究进展,并进行自我归纳和启发式探究,以便为日后的科学研究提供最前沿的支持。这也导致科研人员对本学科数据库的依赖性日益增加,而国内外的相关数据库基本上涵盖了各个学科,且数据量大、内容庞杂,科研人员在检索相关的科学问题时,往往会出现检索结果过多且与本研究不相关的额外信息夹杂其中的情况,这给广大科研人员带来极大的不便^[4]。因此,目前迫切需要一种廉价高效、操作简洁的工具,帮助科研人员从庞杂繁复的数据库中筛选出具有较高学术价值和启发性的文献。

二、文献计量学驱动农业教育的原理

文献计量学(Bibliometrics)主要是利用数学、统计学等学科的方法,定量分析科研人员所需的科研文献,以期获得科研文献的分布规律、研究演变、机构合作,以及个人或团体的影响力等信息^[5]。文献计量学主要基于几种经验统计规律,如布拉德福德定律(Bradford's law),该定律是由英国文献学家 B. C. Bradford 于 20 世纪 30 年代提出的一种定量描述文献分散规律的经验定律,且后期经过诸多研究者的修正和改进而发展成为一种著名的文献分布理论^[6]。20 世纪 20 年代美国学者 A. J. Lotka 提出了一种能够描述科研文献作者的分布规律,被称为洛特卡定律(Lot-

ka's law),又称为“倒数平方定律”^[7]。对于科研文献中关键词频率的分布,1948 年美国学者 G. K. Zipf 提出齐普夫定律(Zipf's law),旨在定量统计文献中关键词出现的频次,并按照频次的高低进行排序,最终获取关键词的频率等级^[8]。

如今,随着文献计量学的快速发展,其研究方向和应用领域已不再局限于图书文献学,其他众多学科领域也逐渐参与其中,其中就包括农业科学研究。通过文献计量学研究,相关领域的研究者能够快速获取本领域最新的研究趋势和发展方向,并归纳演绎出自己的科学问题。事实上,国内外已有诸多学者和机构开发了文献计量学的相关软件工具,以帮助研究者高效进行文献计量学研究,例如 CiteSpace、VOSviewer 和 BibExcle 等软件,均能够进行文献计量学分析,并能够将最终结果可视化呈现,但这些软件在操作难易程度、数据处理效率等方面存在不少问题,且普遍存在对文献分析不够深入,难以为研究者提供更直接信息的缺陷^[9]。

随着信息技术的发展,编程语言已经成为诸多科研人员必备的基础工具,其中 R 语言因为语言结构简洁友好,在数据科学方面具有强大优势,已成为多数农业科学研究者首选的编程语言^[10]。最近,Aria 和 Cuccurullo 开发了一个基于 R 语言的文献计量学工具包“bibliometrix”^[11],该工具包不仅严格遵循了文献计量学的逻辑工作流程,而且能够将基于数学和统计学方法处理所得的数据实现可视化,能够将文献计量分析结果更加直观地呈现出来。更为重要的是,R 语言作为一门面向对象的函数式编程语言且具有诸多统计学和可视化的程序包,意味着该平台能够将文献计量学的分析结果与 R 语言的其他功能模块进行有机结合,从而得出更为丰富的文献计量分析结果,帮助农业高校科研人员更为深入地理解自己的科学问题^[12]。

三、文献计量学构建农业知识体系的实践探索

(一) 文献计量学应用的基本范式

文献计量分析能够使研究者在国际视野下快速

获悉自身研究领域内影响力最高的学者、期刊和国家,并且通过一系列科学分析,获取该领域研究前沿热点的时空分布、交叉融合和历史变迁。因此,文献计量分析不仅能够帮助研究者高效获得本研究领域高质量的文献和期刊,更能够为研究者提供科学的研究思路,明晰自身学科的未来研究方向。但是在进行文献计量分析之前,首先需要厘清主要的工作流程,制订科学的文献计量学分析框架,从而获得高效精准的计量学分析结果。

本研究所述的文献计量分析的基本工作流程主要包括三个主要部分:1. 数据收集,即根据自身学科领域的特点进行文献计量学的研究规划和后续数据集的获取;2. 数据处理与展示,即对所获取的文献计量学数据进行处理与可视化呈现,主要涉及 R 语言中“bibliometrix”工具包以及数据可视化工具“ggplot2”的使用;3. 结果分析,即对文献计量学的数据进行分析解释,主要是基于 R 语言中其他强大数据分析工具与“bibliometrix”工具包之间的交互(见图 1)。

(二) 基于文献计量的知识体系构建流程

基于上述文献计量分析的基本工作流程,知识体系的具体构建工作又可分为以下 5 个主要步骤。

1. 研究规划。关于研究规划,最重要的是研究者首先应明确自己的科学问题,然后利用“bibliometrix”工具包设计自己的检索策略,定义自己研究的科学问题。在规划自身文献计量分析内容时,研究者主要是基于自己的研究主题和研究领域进行数据库构建。此

外,研究者还需选择特定的时间跨度,并基于时间段进行文献计量分析。

2. 数据收集。在获取文献计量所需的数据时,研究者应选择包含文献计量数据的数据库,并过滤核心文献集,之后从选定的数据库中导出数据,最终获得自己所需的文献计量分析数据集。随后,本研究以 SCI 科学引文索引 (Science Citation Index) 的 Web of Science 核心数据库的检索过程作为示例,介绍文献计量分析的数据收集和处理过程。

3. 数据处理。对获得的数据进行处理分析,主要是利用“bibliometrix”工具包和 R 语言中的相关函数来完成,研究者可以编写自己的 R 语言代码来满足自身的研究需求。对于文献计量数据处理分析,前人已经开发出诸多不同的技术方法(见表 1)。例如,文献共引分析是一种基于检索出文献所引用的参考文献来探究某个学科领域重大研究成果的分析技术。共同作者分析也是一种常见的文献计量学分析方法,该方法可获取作者及其隶属关系从而分析出作者的科研群体结构和合作网络。此外,另一种最常见的分析方法是引文分析,主要是利用所检索文章的被引次数来衡量文章、作者和期刊之间的相似性。引文分析又可分为文献耦合分析和共引分析,可以分析作者耦合、作者共同被引、期刊耦合和期刊共同被引等信息。需要注意的是,文献耦合的连接方式是基于相关文献的作者构建的,而共引分析的连接方式则是基于所引用文献的作者构建的。

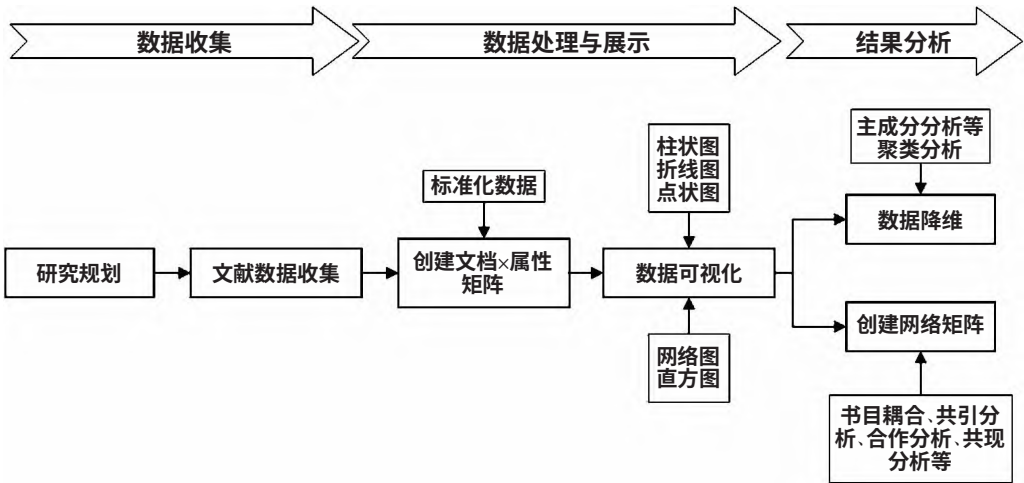


图 1 文献计量学的主要工作流程

表 1 基于文献计量的知识体系的构建内容

技术分类	分析对象	分析类别
文献计量耦合	作者	作者所著文章中常见的参考文献
	文章	文章中的关键词、参考文献等
	期刊	期刊中引用量较高的文献
共引分析	作者	共同被引用的作者
	参考文献	共同被引用的文献
	期刊	共同被引用的期刊
合作分析	作者	文章作者列表中同时出现的作者
	隶属国家	文章地址列表中同时出现的国家
	隶属机构	文章地址列表中同时出现的机构
共词分析	关键词或术语	文章中同时出现的关键词或术语

4. 数据可视化。在数据可视化中,研究者需要基于第 3 步中获取的不同数据进行不同的可视化呈现。文献计量学分析可从数据分析结果中获取可利用的分析内容,并通过可视化来直观地呈现分析结果,例如通过条形图、曲线图、树状图和网络图等。事实上,R 语言拥有非常丰富的数据可视化工具包,能够充分满

足研究者的绝大部分可视化需求。

5. 结果分析。在获得可视化的文献计量学分析结果后,研究者需要描述和分析所有的结果,并对该研究领域进行更深层次的总结和回顾。值得注意的是,文献计量学分析与综述文章(Review Paper)之间是存在差别的。一般地,综述类文章的作者往往在自己的

表 2 “bibliometrix”工具包的主要函数及其功能

主要函数	描述	输出
readFiles ¶ convert2df ¶	将一系列数据库的不同格式文件导入 R 语言中 创建文献计量数据框	数据框
biblioAnalysis ¶ summary ¶ and plot ¶ citations ¶ localCitations ¶ dominance ¶ Hindex ¶ lotka ¶ keywordGrowth ¶ keywordAssociation ¶	输出文献计量数据框中的所有对象 总结文献计量分析的主要结果 获取被引频次最多的文献或作者 获取本地引用频次最多的作者 获取作者的优势度排名 获取作者的 H-index 值并排序 获取作者的 Lotka 定律系数 获取热门关键词的年频次 关联作者和文献的关键词	结果列表
metaTagExtraction ¶ termExtraction ¶ cocMatrix ¶	提取其他字段的标签 从数据框中提取文本字段(摘要、标题、作者关键词等) 构建文档× 标签的矩阵	文档× 属性矩阵
normalizeSimilarity ¶ conceptualStructure ¶	计算文献计量网络中对象之间的关联强度 使用聚类分析等工具创建科学领域的概念结构图	相似性矩阵 重点词矩阵和聚类分析结果
biblioNetwork ¶ histNetwork ¶	获取常用的文献计量耦合、共引、合作和共现网络 从文献计量数据框中获取历史共被引网络	网络矩阵
networkPlot ¶ histPlot ¶ conceptualStructure ¶	利用 R 语言自带包绘制文献计量学网络图 获取历史直接引文网络 使用聚类分析等工具获取科学领域的概念结构图	用于其他计量学工具和 语义图的网络图

研究领域拥有较大学术影响力,他们在研究该领域大量科研进展之后,结合自身的研究进展,对该领域的研究进展和前景进行归纳整理,最终对该研究领域内的前沿科学问题提出一种前瞻性的指导或分析,属于探讨某一学科领域的内在结构。

(三) 基于编程的文献计量学的实现

如今越来越多的教师采用编程对研究生进行教学以及知识体系的构建,文献计量学亦能使用基于 R 语言平台和“bibliometrix”工具包进行农业科学的知识体系的构建^[11]。为构建研究生自身的知识体系,研究生可选择相应的科研数据库。

为使研究者迅速了解“bibliometrix”工具包的主要功能,帮助农科研究生快速地建立自己的知识体系,本节主要介绍该工具包中的主要功能。首先,研究者需要使用 R 语言代码“install.packages(“bibliometrix”, dependencies=TRUE)”和“library(bibliometrix)”来下载和加载该工具包。然后,将获取的文献检索数据进

行加载和转换,主要使用表 2 中的“readFiles ¶”和“convert2df ¶”这 2 个函数。将文献数据转化成“bibliometrix”工具包的可用数据后,就可进行简要的描述性文献计量分析,例如使用“biblioAnalysis ¶”函数可以获取文献计量分析中所有类型的数据对象;“citations ¶”和“localCitations ¶”函数可以分别获得全球和本地被引频次最多的文献或作者及其排名;而“dominance ¶”和“Hindex ¶”函数则可针对作者获得其在某一领域的优势度和 H-index 值及其排名;关键词的年频率则可以用“keywordGrowth ¶”函数来计算,而“keywordAssociation ¶”函数可以将作者和关键词进行关联,从而获取某领域具有影响力作者的研究热点和方向。之后,可利用相关的代码进行文献耦合和引文分析等功能网络的构建,这是进行知识体系构建中关键的过程。最后,研究生可根据自己已梳理的科研知识体系,利用 R 语言中相关的功能将知识体系中的一些关键节点进行可视化。

参考文献:

- [1] 习近平给全国涉农高校的书记校长和专家代表的回信 [EB/OL]. (2019-09-06) [2026-05-01]. http://www.xinhuanet.com/2019-09/06/c_1124967725.htm.
- [2] 中共中央 国务院关于做好 2022 年全面推进乡村振兴重点工作的意见 [EB/OL]. (2022-02-22) [2026-05-01]. http://www.gov.cn/zhengce/2022-02/22/content_5675035.htm.
- [3] 胡弼成,王祖霖. “大数据”对教育的作用、挑战及教育变革趋势:大数据时代教育变革的最新研究进展综述 [J]. 现代大学教育,2015 (4): 98-104.
- [4] 王婷,颜蕴,续玉红,等. 中国农业科学院科技论文产出及国际学术影响力分析:基于 Web of Science 数据库 [J]. 中国农业科技导报,2013,15 (2): 54-63.
- [5] 叶深溪,许为民. 文献计量学在科研评价中的应用进展 [J]. 图书馆论坛,2003 (4): 12-14.
- [6] Weinstock M. Bradford's Law [J]. Nature,1971,233 (5319): 434.
- [7] Kumar S, Sharma P, Garg K C. Lotka's law and institutional productivity [J]. Information processing & management, 1998,34 (6): 775-783.
- [8] LI W T. Random texts exhibit zipf-law-like word-frequency distribution [J]. IEEE transactions on information theory, 1992,38 (6): 1842-1845.
- [9] 周超峰. 文献计量常用软件比较研究 [D]. 武汉:华中师范大学,2017.
- [10] RCoreTeam. The R Project for Statistical Computing [EB/OL]. (2026-04-24) [2026-05-21]. <https://www.r-project.org/>.
- [11] Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis [J]. Journal of informetrics, 2017,11 (4): 959-975.
- [12] 蓝洋,何秀,朱诚勛,等. R 语言在生物科学研究绘图中的应用 [J]. 华东师范大学学报(自然科学版),2019 (1): 124-135.

(责任编辑 陈华康)