

数据素养研究热点与主题变迁

——基于 CiteSpace 知识图谱可视化分析

田 华^a, 陈 杰^b

(信阳师范学院 a. 生命科学学院; b. 马克思主义学院, 河南 信阳 464000)

摘 要:文章基于 CiteSpace 知识图谱可视化分析,以 CNKI 数据库 206 篇数据素养文献为研究对象,采用关键词共现网络及聚类分析,生成数据素养领域的研究作者、研究机构、研究热点和研究时间态势的知识图谱,重点分析关键词共现网络的信息素养、教育大数据、科学数据素养、数据素养教育、科学数据等 5 个自动聚类,宏观揭示数据素养领域研究文献的演进路径,为进一步深化数据素养研究提供参考。研究表明:数据素养研究受国家政策驱动明显,整体研究水平还有待于提升;图书馆是数据素养教育的主阵地;数据素养研究内容以国外教育案例介绍为主,研究方法以理论研究为主,实证应用研究不足。

关键词:数据素养; CiteSpace; 关键词共现; 可视化分析

OSID:



中图分类号:G434 **文献标志码:**A **文章编号:**1003-0964(2020)04-0052-08

随着大数据时代的到来,数据驱动创新和决策的新模式对行业与个人都产生重要影响^[1]。数据在《现代汉语词典》中指“进行各种统计、计算、科学研究或技术设计等所依据的数值”,网络中的数据是指收集、测度并保存起来可管理可利用的信息。总体来看,大数据是财富,是资源,是科学研究的战略高地,因此要把大数据提升到国家战略层面,要真正把大数据当作国家重要资源来谋划发展战略^[2]。要发挥国家大数据战略功能,科学数据的采集、存储、处理、传播、管理以及共享等技能素养,就显得极其重要^[3],这种技能素养是当前大数据领域研究的热点,是科研人员提升科研创新能力的必备因素和核心竞争力,也称数据素养^[4,5]。

对于数据素养的确切定义,目前学术界尚未形成一致的认识。较早明确提出“数据素养”这一概念的是 2004 年发表的《信息素养、统计素养和数据素养》一文^[6]。Sarah 等认为,数据素养是指能够将数据研究转化为科学决策与政策实施的数据使用能力,这种能力是所有研究者的必备素养^[7]。Carlson 等认为,数据素养的生成过程就是研究者正确认知、理解、运用和转化数据的实施过程^[8]。在国内,张静波等认为数据素养指“研究者在科学数据的采集、组

织和管理、处理和分析、共享与协同创新利用等方面的能力,以及研究者在数据的生产、管理和发布过程中的道德与行为规范”^[9]。孟祥保认为数据素养是指“科学研究中收集、加工、管理、评价和利用数据的知识与能力”^[10]。郝媛玲把数据素养划分为数据态度、数据意识、数据知识、数据技能和数据伦理 5 个方面^[4]。尽管数据素养目前还没有一个明确一致的定义,但是综合多方面文献,本研究认为数据素养强调个体具备较高的数据意识和伦理道德,能够创设最优化的数据获取环境和数据收集工具,进而规范获取数据、批判性地选择和评估数据、规范化地管理和处理数据、合理地应用和共享数据。本文立足 CNKI 数据库,应用软件 CiteSpace V 生成数据素养研究文献的知识图谱,进而开展可视化分析,以实现学科知识的前沿热点和演变趋势的可视化呈现,宏观揭示数据素养领域文献的演进路径,为研究者深入开展数据素养研究提供参考。

一、数据素养:信息素养的拓展与延伸

信息素养在 1974 年首次被 Paul Zurkowski 提出,是当前发展比较成熟的概念。1979 年, Taylor 将该词汇应用到图书馆研究。1989 年,美国图书馆协会(ALA)指出信息素养的本质是发现、检索、分析

收稿日期:2020-01-11;**收修日期:**2020-03-20

基金项目:河南省教师教育课程改革与研究重点项目(2019-JSJJYD-017);河南省高等学校哲学社会科学应用研究重大项目(2018-YYZD-14);信阳师范学院马克思主义学院 2019 年教学改革与研究 A 类项目

作者简介:田 华(1979—),女,河南南阳人,博士,副教授,硕士生导师,研究方向为教育信息化。

和利用信息。2011年,英国国立和大学图书馆协会指出信息素养包括数字素养、视觉素养、媒介素养、学术素养、信息技术、信息处理、数据管理等7个要素。2015年,美国大学与研究图书馆协会(ACRL)定义信息素养“是包含反映发现信息、理解信息生产和价值、使用信息创造新知识和参与社群学习的综合能力的集合”^[11]。简单来说,信息素养是强调人们利用信息解决问题的能力。在大数据时代,信息存在方式发生变化,信息的形式多样化,信息素养的内涵和外延在不断扩大,对信息素养提出了挑战,尤其是大数据时代的到来更是如此。2015年,中华人民共和国国务院审议通过了《促进大数据发展行动纲要》,标志着我国在顶层设计上对大数据的实践与实施做出了总体部署,数据已成为国家基础性战略资源^[12]。获取数据、挖掘数据、应用数据已不仅仅是数据科学家和技术专家必备的技能,也是新时代我们研究者所需的综合素质,这使得数据素养日渐成为一个活跃的研究领域。

比较信息素养和数据素养,我们发现二者都是适应时代发展提出来的,数据素养可以理解为信息素养的拓展和延伸,但专家的意见也不尽一致。张静波认为数据素养是信息素养的子集^[9];黄如花认为数据素养比信息素养的外延更大^[11];杨晓菲认为数据素养是信息素养的最佳延伸与表达^[13]。在信息的获取与挖掘、分析与应用、安全与伦理方面,信息素养与数据素养具有很大的相似性,二者都强调个体需要具备发现、查找、确认、评价、利用和解决问题的能力,但数据素养尤为强调个体的数据分析应用能力、批判性思考能力和数据伦理与道德。数据素养所涉及的数据类型多是非结构性数据,处理起来更复杂,难度更大。在大数据时代,运用云计算、人工智能、深度学习等大数据技术将数据连接、贯通,彰显了数据的战略性地位和时代价值。这是大数据分析的魅力所在,也是大数据时代信息素养所欠缺的薄弱环节。数据素养是大数据时代信息素养的延伸和发展,是一种能够辩证、科学、正确认识和挖掘数据的价值,在符合道德伦理的基础上对数据加以操作和管理,能够通过分析、存储、转化等操作或长久保存信息,以备再次应用以实现其蕴含的巨大价值的能力。

二、文献结构特征分析

(一)数据来源与研究方法

由于近10年来数据素养领域研究文献众多,本文从CNKI数据库中选取CSSCI期刊论文,检索时间为2019年8月30日,以“主题=数据素养”进行

检索,共检索出214篇记录。基于检索结果,根据题目和摘要,将非学术论文、与主题不相关的论文进行剔除,人工筛除不相关文献8篇,最终确定学术论文共计206篇作为分析样本,采用美国Drexel大学陈超美博士开发的CiteSpaceV可视化分析软件,对样本文献进行关键词共现分析、聚类分析和时间序列分析等可视化呈现,能够有效反应文献之间的互动、交叉、联系和衍生等关系。

(二)年度分布

数据素养领域的CSSCI论文最早出现于2011年,2015年发文开始呈规模增加,发展趋势曲线呈现出迅速上扬态势,这表明对于数据素养的理论研究已经进入快速发展阶段。由图1所示,2011—2014年数据素养领域论文发文总量9篇,而仅仅2015年即达22篇,2016—2018年分别为48篇、45篇、55篇,这与国家的政策驱动是分不开的。2015年,国家发布《促进大数据发展行动纲要》,首次提出发展“国家大数据战略”。大数据被提升到国家战略层面,已经成为国家的基础性战略资源,是重要的生产力。国家的重视与扶持为大数据产业的发展提供了一个良好的契机。

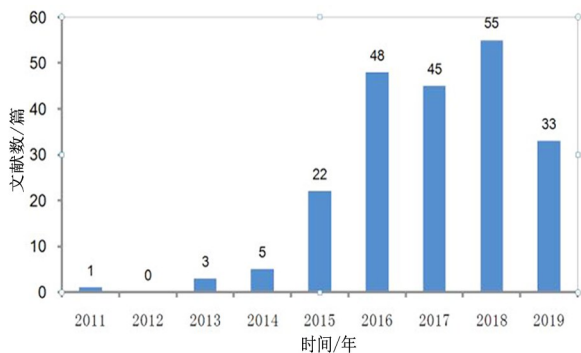


图1 数据素养领域高质量论文的年度分布

(三)核心作者和核心机构

图2是数据素养领域作者合作网络图谱,节点34个,连线18条,展示的是数据素养领域作者之间的合作关系,节点表示作者,节点越大表示该作者的文章被引频次越多,节点间的连线表示合作关系,连线越粗表示合作关系越强,连线的颜色表示不同的年份。图3是数据素养领域研究机构知识图谱,节点23个,连线6条。节点表示机构,节点越大说明该机构的文章被引频次越多,节点间的连线表示合作关系,连线越粗表示合作关系越强,连线的颜色表示了合作的时间段。

由图2可以看出,共被引频次排名前10的作者是胡卉、邓李君、吴鸣、张群、沈婷婷、黄如花、杨现民、杨文建、李新、郝媛玲。由图3可以看出,共被引

频次较高的研究机构有中国科学院文献情报中心、武汉大学信息管理学院、四川外国语大学图书馆、中国科学院大学、上海大学图书馆等。总体来看,各研究机构间比较独立,没有较高的合作度。综合图 2 和图 3,数据素养研究领域共被引频次最高的作者胡卉,供职于中国科学院文献情报中心。其次是邓李君、吴鸣和张群,分别供职于四川外国语大学图书馆、中国科学院文献情报中心和江南大学图书馆。

大学图书馆是海量数据资源的集散地和学术信息的交流平台,并且在服务科研一线的过程中积累了丰富的数据管理经验,成为提供数据管理服务和开展数据素养教育的主要阵地^[14]。围绕数据素养的教育和研究,图书馆的服务面临新的挑战 and 机遇,积极开展数据素养教育已经成为当前一项紧迫和重要的任务。

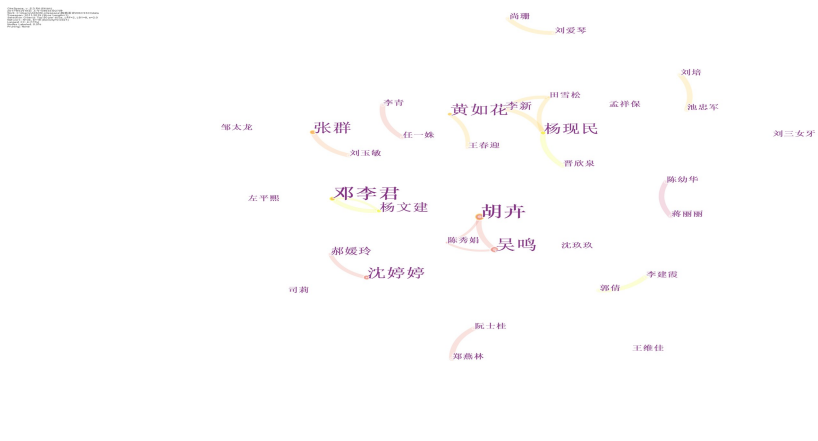


图 2 数据素养领域作者合作网络图谱

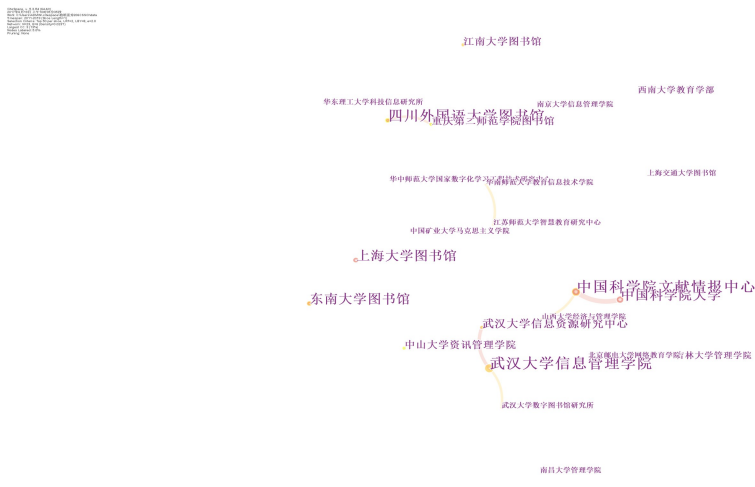


图 3 数据素养领域研究机构知识图谱

(四)期刊分布

从学术论文的期刊分布来看,一共有 19 种期刊就“数据素养”主题刊登了超过 4 篇以上的论文(见图 4)。从载文数量来看,以《图书与情报》《图书馆理论与实践》《图书情报工作》3 种期刊最为突出,其中 3 种杂志 9 年间共发表“数据素养”主题论文分别为 20 篇、20 篇、16 篇。从期刊的主题分布来看,其主题分布比较集中,主要涉及图书情报与数字图书馆、教育理论与教育管理。数据素养的跨学科研究也发展迅猛,已深入到图书馆、情报与档案管理以及教育学等多个学科,并衍生出多个交叉学科主题。

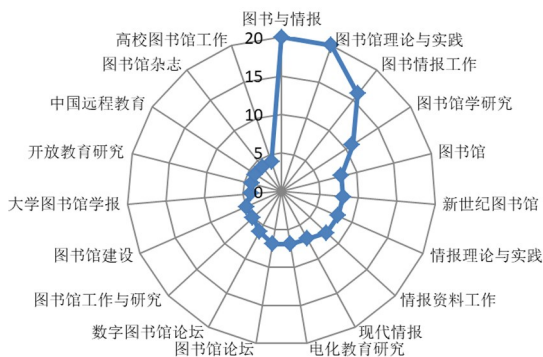


图 4 数据素养领域载文 4 篇以上的期刊分布

三、研究热点分析

(一)关键词共现热点分析

CiteSpace V 通过对高频关键词的分析来确定该领域的研究热点。关键词是对文章核心内容浓缩和提炼出的词语,能够有效将文章的研究目的、研究

对象、研究方法和研究结果进行概括总结,关键词的知识图谱见图 5。从图 5 和表 1 可以看出,“数据素养”“大数据”是该知识图谱中频次最高的两个节点。此外,“数据素养教育”“高校图书馆”“信息素养”等关键词的频次也位于前列。



图 5 数据素养领域关键词共现知识图谱

表 1 数据素养频次 TOP20 的关键词

序号	频次	关键词	序号	频次	关键词
1	101	数据素养	11	5	科学数据素养
2	61	大数据	12	4	美国
3	37	高校图书馆	13	4	教育
4	35	数据素养教育	14	4	信息素养教育
5	25	信息素养	15	4	科学数据管理
6	12	图书馆	16	4	教师专业发展
7	8	教育大数据	17	4	MOOC
8	7	科学数据	18	3	数据新闻
9	7	大学图书馆	19	3	科研数据管理
10	6	大数据时代	20	3	数据生命周期

其次,中介中心性代表着节点在网络中的媒介能力。关键词的中介中心度越高,说明其占据节点的位置越多、控制的关键词之间的信息流越多,可以认为该节点居于重要地位。通常在一个共词网络中,中介中心度较高的节点很大可能预示了未来的研究趋势。分析数据素养中介中心度排列前 20 名的关键词(简称 TOP20)发现,关键词中介中心性位于前 5 的分别是“数据素养”“数据素养教育”“大数据”“高校图书馆”“信息素养”等(见表 2)。综合表 1 和表 2,关键词词频与中心中介性指标呈现部分一致趋势,有 15 个关键词同时位列 TOP20 高频关键词和 TOP20 高中心中介性关键词,分别是:“数据素养”“大数据”“高校图书馆”“数据素养教育”“信息素养”“图书馆”“教育大数据”“科学数据”“大学图书馆”“大数据时代”“科学数据素养”“美国”“教育”“科学数据管理”“科研数据管理”。可见,大数据是数据素养研究的重中之重,数据素养教育、信息素养、教

育大数据等也是该领域重要的研究议题,图书馆、大学图书馆是与数据素养紧密相关的机构。

表 2 数据素养中心中介性 TOP20 的关键词

序号	中介中心度	关键词	序号	中介中心度	关键词
1	0.66	数据素养	11	0.02	教育
2	0.42	数据素养教育	12	0.01	图书馆
3	0.27	大数据	13	0.01	科学数据素养
4	0.25	高校图书馆	14	0.01	科学数据管理
5	0.13	信息素养	15	0.01	科研数据管理
6	0.08	人才培养	16	0.01	数据素养能力
7	0.03	科学数据	17	0.01	数字人文
8	0.03	大学图书馆	18	0.01	指标体系
9	0.03	美国	19	0.01	科研数据服务
10	0.02	教育大数据	20	0.01	大数据时代

(二)关键词聚类热点分析

为了进一步探索该领域的研究热点,本研究通过 CiteSpace V 软件对关键词共现知识图谱进行自动聚类,生成自动聚类标签视图(见图 6),分别是“信息素养”“教育大数据”“科学数据素养”“数据素养教育”“科学数据”。Modularity 为 0.3714, Mean Silhouette 为 0.5534。

1. 聚类 0: 信息素养

信息素养的研究主要集中在信息素养的内涵、教育模式和评估三个方面。信息素养的内涵前文已经分析,在此不再赘述。第一,信息素养教育模式的多元化。信息素养教育模式比较成熟的主要有通识(学分)课程模式、嵌入整合教育模式、基于 MOOC 的混合教学模式三种。国外学者侧重于 MOOC 环

境下信息素养技能、信息素养教育能力标准以及信息素养教育模式等方面的研究,国内学者侧重于国内外信息素养 MOOC 建设现状的调研、基于 MOOC 的信息素养平台设计、信息素养教育内容框架与教育模式以及翻转课堂在信息素养教育中的应用等方面的研究^[15]。第二,信息素养评估。信息素养评估主要包括学生信息的评估、教师信息的评估和国际健康信息的评估。学生信息评估的对象包括大学生和中小學生,主要针对大学生,针对中小學生

的信息素养实证研究很少,评估的维度通常为信息素养思维观念和实践能力两个层面,助力《教育信息化 2.0 行动计划》所制定的学生信息素养评价指标体系^[16]。健康信息素养评估自 1997 年开始日益受到各国学者的关注,美国在健康信息素养评估研究中遥遥领先于其他国家,我国在该方面的能力标准和标准化评估工具的研究设计方面还处于探索阶段^[17]。

图6展示了数据素养领域关键词共现网络自动聚类标签视图。图中包含多个聚类标签，如#0 信息素养、#1 教育大数据、#2 科学数据素养、#3 数据素养教育、#4 科学数据、#5 数据素养、#6 数据素养教育、#7 数据素养、#8 数据素养、#9 数据素养、#10 数据素养、#11 数据素养、#12 数据素养、#13 数据素养、#14 数据素养、#15 数据素养、#16 数据素养、#17 数据素养、#18 数据素养、#19 数据素养、#20 数据素养、#21 数据素养、#22 数据素养、#23 数据素养、#24 数据素养、#25 数据素养、#26 数据素养、#27 数据素养、#28 数据素养、#29 数据素养、#30 数据素养、#31 数据素养、#32 数据素养、#33 数据素养、#34 数据素养、#35 数据素养、#36 数据素养、#37 数据素养、#38 数据素养、#39 数据素养、#40 数据素养、#41 数据素养、#42 数据素养、#43 数据素养、#44 数据素养、#45 数据素养、#46 数据素养、#47 数据素养、#48 数据素养、#49 数据素养、#50 数据素养、#51 数据素养、#52 数据素养、#53 数据素养、#54 数据素养、#55 数据素养、#56 数据素养、#57 数据素养、#58 数据素养、#59 数据素养、#60 数据素养、#61 数据素养、#62 数据素养、#63 数据素养、#64 数据素养、#65 数据素养、#66 数据素养、#67 数据素养、#68 数据素养、#69 数据素养、#70 数据素养、#71 数据素养、#72 数据素养、#73 数据素养、#74 数据素养、#75 数据素养、#76 数据素养、#77 数据素养、#78 数据素养、#79 数据素养、#80 数据素养、#81 数据素养、#82 数据素养、#83 数据素养、#84 数据素养、#85 数据素养、#86 数据素养、#87 数据素养、#88 数据素养、#89 数据素养、#90 数据素养、#91 数据素养、#92 数据素养、#93 数据素养、#94 数据素养、#95 数据素养、#96 数据素养、#97 数据素养、#98 数据素养、#99 数据素养。

图6 数据素养领域关键词共现网络自动聚类标签视图

2. 聚类 1: 教育大数据

2018年4月,中华人民共和国教育部正式发布《教育信息化 2.0 行动计划》,全面开启以大数据为重要载体的教育信息化发展升级新征程^[18]。教育大数据是大数据的一个子集,特指教育领域的数据集。教育大数据是教育活动过程中所产生的以及根据教育需要采集到的一切可用于描述教育教学活动的数据集,包含教学过程采集和捕捉的教学大数据^[19]。正因为大数据的嵌入,使得教育被技术赋权,拥有了表征教育过程的大数据量的有形价值,更揭示出了教育与数据之间产生的关联性、系统性、趋向性的无形价值。这是传统教育模式面临的新挑战。依照不同层级的教育主体和教学活动的各项内容,教育大数据可以分为4个层次和6大类型。4个层次包括个体、学校、区域和国家;6大类型包括基础数据、教学数据、科研数据、管理数据、服务数据和舆情数据^[20]。教育大数据关注的焦点集中于3个方面^[21-23],一是教育大数据的基础理论,如教育大数据的概念、内涵、特征以及价值;二是教育大数据支持下的学习方式、学习分析的探索,如运用教育大数据改善学习、评价学习;三是教育大数据在教育领域的应用,重点关注重塑教学模式、转型研究范式、

创新教育决策、变革教育管理等方面,包含高等教育、中等教育、基础教育、职业教育及思想政治教育等。

3. 聚类 2 和聚类 4: 科学数据与科学数据素养

科学数据素养是数据素养的一个子集,将“数据素养”的概念和应用范畴限定于科学研究领域,而且对数据的范围给予更加明确的界定,特指“科学数据”。传统的科学数据是指人类在认识世界、改造世界的一切科技活动中所产生的原始性、基础性数据,以及按照不同需求系统加工的数据产品和相关信息^[24]。随着信息技术的进步,科学研究的信息化日益成熟,科学计算促使人类描述社会复杂事物的能力不断提升,科学数据被赋予了新的内涵。2019年2月颁布的《中国科学院科学数据管理与开放共享办法(试行)》中指出,“科学数据主要包括在自然科学、工程技术等领域,通过基础研究、应用研究、试验开发等产生的数据,以及通过观测监测、考察调查、检验检测等方式取得并可用于科学研究活动的原始数据及其衍生数据”^[25]。在大数据时代,科学数据已经不再是孤立的静态数据,而是处于一种整体框架、系统逻辑之内的一种动态数据,进而形成一种相互关联、相互影响的数据网络。因此,数据越来越成

为包括教育在内的各行各业创新发展不可或缺的基础资源,如何对其进行长时段、全方位的采集,系统化、科学化的分析就显得至关重要了。

从现有研究来看,科学数据素养主要指从事科学研究的科研工作者在数据的收集、处理、管理、评价、利用、再生产等科学研究过程中所具备的一般的思维、知识和技能,以及在数据管理中形成的一种伦理道德与行为规范。科学数据素养重在强化对数据的理解、管理和利用的能力,目的在于实现数据的再生产和再利用。纵观当前科学数据素养研究路径,主要体现在三个方面:一是数据理解,不同的理解将会带来不同的数据利用,这是数据利用和管理的前提;二是数据利用,通过收集和分析用户数据行为,进而实现数据的再利用和再生产;三是数据管理,主要是在数据生命周期内有效地对数据进行收集、加工、存储、管理、评估和使用等。三个研究目标相互影响、相互作用,其根本目的在于培养学习者科学数据素养能力。

4. 聚类 3: 数据素养教育

数据素养教育是提升数据素养能力的必要路径。目前,国内外数据素养教育研究主要集中于以下几个方面:一是数据素养教育主体。主要指政府与相关机关,如国家统计局、科学基金会、教育部、高等院校和图书馆行业等^[28,29]。二是数据素养教育的对象。主要包括图书馆员、科研人员、高校教师及大学生和研究生、中小学教师及中小学生、编辑等。三是数据素养案例研究。主要包括对美国、加拿大、德国、英国等高校图书馆数据素养教育实践经验的总结。四是数据素养课程研究。主要集中在调查国内外高等院校、图书馆等数据素养课程设置情况^[30]。五是数据素养教育模式以及评价研究。主要涉及嵌入内容的教育模式、分层次的阶段教育模式、跨部门协同发展模式、开放式在线教育模式等^[31]。但是,目前国内关于数据素养教育理论研究

仍然极度薄弱,尤其在数据素养教育的总体框架、实践操作、评价标准等方面还存在诸多空白区域。这与社会各界对于数据素养教育的广泛关注的实际情况之间存在较大差距,尤其数据素养教育实证研究方面亟须人们对其进一步深化和突破。

四、文献主题变迁分析

为了更好地理解 2011—2019 年数据素养文献的主题变迁,本文通过 CiteSpace V 软件绘制出了该领域的关键词时间线视图(见图 7),大致可以分为 2011—2013 年、2014—2015 年、2016—2018 年 3 个阶段。第一阶段,2011—2013 年。这个阶段是数据素养的萌芽期,相关文献还较少,研究方向主要集中在数据素养重要性的论证。第二阶段,2014—2015 年。2014 年,大数据首次被写入我国《政府工作报告》,得到了国家层面的支持,大数据迎来了重要的发展机遇。2015 年,在党的十八届五中全会中首次提出实施大数据战略,紧接着国务院印发《促进大数据发展行动纲要》,数据已成为国家基础性战略资源,是“十三五”规划的重点。基于国家大数据战略,如何加速大数据应用的落地生根,方便快捷的大数据分析工具必不可少,具有较高数据素养的人才尤其重要。基于此背景,从 2014 年开始,关于数据素养研究的文献逐渐增多,研究领域主要集中在如何开展数据素养教育、图书馆在数据素养教育中的地位和策略。第三阶段,2016—2019 年。从 2016 年我国发布《大数据产业发展规划(2016—2020 年)》开始,数据素养领域呈现出研究内容丰富、发表论文数量增多的现象,数据素养的研究由宏观研究转向微观研究,研究重心由数据素养的重要性和国外案例介绍转向大数据时代数据素养教育的具体实施策略和模式研究,呈现出一大批数据素养教育的模式、评价模型构建、培养机制等理论研究,但系统性的实证应用研究还有待加强。

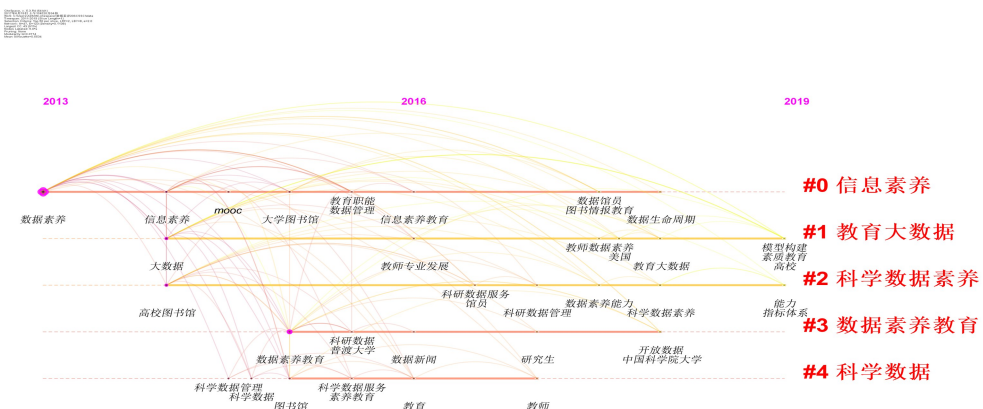


图 7 数据素养领域关键词时间线视图

五、结论

纵观2011—2019年我国数据素养领域的研究成果,从2011年开始,数据素养日渐成为大数据时代的“新宠儿”,信息素养越来越广泛地被拓展到各个研究领域,并在各个领域实现对现有数据的再整合、再挖掘和再管理,进而在新的社会实践中彰显出更大更新的价值。通过以上分析,数据素养的研究领域研究呈现如下特点,尤其是一些关键问题必须引起我们的重视并积极为之改进。

(一)数据素养研究受国家政策驱动明显,整体研究还有待提升

虽然国内学者对数据素养进行了大量研究,但是该领域仍然处于起步和探索阶段。从2011—2019年的研究热度和研究主题来看,我国数据素养领域的研究与国家政策出台的步伐密切相关,尤其是国家大数据战略与推动教育变革的教育大数据。2015年国家首次提出发展大数据战略。数据素养CSSCI期刊论文数由2015年的22篇猛增到2016年的48篇。胡卉、邓李君、吴鸣、张群、沈婷婷、黄如花、杨现民等在数据素养研究方面都是先行者。特别是2015年以来他们把美国、英国、加拿大等国家开展数据素养教育的先进经验陆续介绍到国内,研究对象从国家层面、地方政府到高校,从教师、学生到编辑,在数据素养的概念、内涵、必要性、评价模型构建、培养机制、教育模式等方面开展了大量的理论探讨,但我国数据素养的整体研究水平还非常薄弱,亟待进一步提升。

(二)图书馆在数据素养教育中起着极其重要的作用

2015年12月31日中华人民共和国教育部印发的《普通高等学校图书馆规程》指出,图书馆应重视开展信息素质教育,采用现代教育技术,加强信息素质课程体系建设,完善和创新新生培训、专题讲座的形式和内容。数据素养教育是大数据环境下国家和社会对图书馆赋予的新责任和高要求。图书馆作为数据素养教育教学和科研的基地,不仅需要整合资源和服务,而且需要在实践中提升数据素养教育从业馆员(数据专家)的数据管理经验、数据操作技能和培训能力,从受教育对象的不同需求和特征出发,开展数据素养的合作教育模式,合作的对象可以是专业教师、学生和数据库运营商等,同时扩展数据素养教育的覆盖范围和教育方式,密切关注数据素养教育的发展变化,保障数据素养教育的实践性、持久性与连续性。

(三)数据素养研究方法以理论探讨为主,深度系统的实证研究不足

国内数据素养研究还处于探索阶段,研究内容更多的是停留在概念的理论探讨层面。在实践层面虽然开展了MOOC、翻转课堂、嵌入专业课等教学模式,但缺乏运用大规模数据的实证研究,并且既有的实践运行模式还存在诸多不足。如数据素养教育内容尚未形成体系,与专业领域相结合的数据素养教育十分薄弱;课程设计、课程资源配置不完善,教育形式仍是单一的讲座和培训,内容创新不足,缺乏结合社交网络、移动网络等在线学习平台开展多渠道的数据素养教育案例;缺乏微观尺度的个案深入刻画和宏观层面的社会文化解读;高校图书馆在数据素养教育体系中的定位及培养策略方面的研究还不充分。

总之,当前国内数据素养教育虽取得了一定的成绩,但是如何结合大数据时代背景和当前人工智能、机器学习和预测分析等新技术,有针对性、系统性地提升数据素养,尤其是提升大学生数据素养,目前还缺乏国家层面的政策引领和实践指导。数据素养教育要把数据意识和数据伦理作为导向,数据知识和能力作为核心,数据分析应用作为关键,数据的合理使用、加工与传播作为基础。从实践上看,数据素养教育对象不仅包括专业人员,而且应包括普通大众;数据素养能力的提升,不仅需要学习者自身的不断学习探讨与领悟,而且需要政府、高校、企业等多部门、多渠道、多模式、多层次的有效引导。

参考文献:

- [1] 胡元元,朱慧敏.我国科研数据管理服务:内涵、主题演变、研究框架及其启示[J].图书馆学研究,2019(4):17-25.
- [2] 郭华东.科学大数据——国家大数据战略的基石[J].中国科学院院刊,2018,33(8):768-773.
- [3] 马腾,孙玲.信息生态视域下高校大学生数据素养评价研究[J].情报科学,2019,37(8):120-126.
- [4] 郝媛玲,沈婷婷.数据素养及其培养机制的构建与策略思考[J].情报理论与实践,2016,39(1):58-63.
- [5] 林秀清,杨现民,李新.中小学教师数据素养的发展路径与培养策略[J].现代教育技术,2020,30(1):59-65.
- [6] SHIELDS M. Information literacy, statistical literacy, data literacy[J]. IASSIST Quarterly,2004,28(2):7-14.
- [7] SARAH J, ALEXANDER B, CUNA E. The Data Audit Framework: A First Step in the Data Management Challenge[J]. International Journal of Digital Curation, 2008, 3(2):112-120.

- [8] CARLSON J, FOSMIRE M, MILLER C C, et al. Determining data information literacy needs: a study of students and research faculty[J]. *Portal-Libraries and the Academy*, 2011, 11(2): 629-657.
- [9] 张静波. 大数据时代的数据素养教育[J]. *科学*, 2013, 65(4): 29-32+4.
- [10] 孟祥保, 常娥, 叶兰. 数据素养研究: 源起、现状与展望[J]. *中国图书馆学报*, 2016, 42(2): 109-126.
- [11] 黄如花, 李白杨. 数据素养教育: 大数据时代信息素养教育的拓展[J]. *图书情报知识*, 2016(1): 21-29.
- [12] 国务院关于印发促进大数据发展行动纲要的通知(国发[2015]50号)[EB/OL]. (2015-09-09)[2020-02-20]. <http://www.scio.gov.cn/xwfbh/xwfbh/wqfbh/33978/34896/xgzc34902/Document/1485116/1485116.htm>.
- [13] 杨晓菲. 从信息素养到数据素养: 内涵、关系及培养途径[J]. *河北科技图苑*, 2015, 28(6): 30-32.
- [14] 胡卉, 吴鸣. 面向不同参与主体的国内外数据素养教育现状研究[J]. *数字图书馆论坛*, 2016(9): 61-67.
- [15] 蒋丽丽, 陈幼华. 国内外高校信息素养 MOOC 关键成功因素研究[J]. *图书情报工作*, 2015, 59(15): 62-67.
- [16] 马晓玲, 张心如, 阮凌志, 等. 亚太地区基础教育阶段学生信息素养评估比较研究[J]. *中国电化教育*, 2018(8): 60-66.
- [17] 李信, 李旭晖. 基于文献计量的国际健康信息素养评估研究[J]. *现代情报*, 2017, 37(2): 134-140.
- [18] 教育部关于印发《教育信息化 2.0 行动计划》的通知[EB/OL]. (2018-04-18)[2020-02-20]. http://www.moe.gov.cn/srcsite/A16/s3342/201804/t20180425_334188.html.
- [19] 涂涛, 胡柯铭. 一极两仪: 教育大数据与大数据关系辨析[J]. *中国电化教育*, 2019(8): 18-22.
- [20] 孙洪涛, 郑勤华. 教育大数据的核心技术、应用现状与发展趋势[J]. *远程教育杂志*, 2016, 34(5): 41-49.
- [21] 王学男. 不同教师群体对教育大数据的认知及影响因素——基于全国 5434 名教师的调查[J]. *开放教育研究*, 2019, 25(3): 81-91.
- [22] 蒋鑫, 洪明. 国际教育大数据研究的热点、前沿和趋势——基于 WOS 数据库的量化分析[J]. *中国远程教育*, 2019(2): 26-38.
- [23] 杨现民, 周宝, 郭利明, 等. 教育信息化 2.0 时代教育数据开放的战略价值与实施路径[J]. *现代远程教育研究*, 2018(5): 10-21.
- [24] 秦小燕, 初景利. 科学数据素养内涵结构研究[J]. *图书情报工作*, 2019, 63(18): 30-39.
- [25] 中国科学院关于印发《中国科学院科学数据管理与开放共享办法(试行)》的通知[EB/OL]. (2019-09-17)[2020-02-20]. http://www.ecas.cas.cn/gzwwj/ygzdzd/201909/t20190917_4552523.html.
- [26] 张群, 彭奇志, 张路路, 等. 高校图书馆用户科学数据素养需求调查分析[J]. *图书馆学研究*, 2020(1): 67-75.
- [27] 张群, 张慧, 张路路, 等. 认知视角下用户科学数据素养研究架构设计及其干预策略[J]. *新世纪图书馆*, 2019(12): 11-15.
- [28] 邓李君. 国内外数据素养教育发展现状研究[J]. *图书馆理论与实践*, 2017(7): 30-33.
- [29] 杨晓琼. 大数据时代高校数据素养教育的合作路径[J]. *情报资料工作*, 2015(3): 98-102.
- [30] 张璇, 孟祥保. 面向数字人文的高校数据素养教育案例研究[J]. *大学图书馆学报*, 2019, 37(5): 87-94.
- [31] 郭倩, 李建霞. 基于多元主体的高校数据素养教育生态模式构建研究[J]. *图书馆理论与实践*, 2019(5): 49-57.

Visualization Analysis of Knowledge Map of the Research Hotspots and Thematic Changes of Data Literacy based on CiteSpace

TIAN Hua^a, CHEN Jie^b

(a. School of Life Science; b. School of Marxism, Xinyang Normal University, Xinyang 464000, China)

Abstract: Based on the visual analysis of CiteSpace knowledge map, this paper presents the research hotspots and thematic changes of data literacy, reveals the evolutionary path of research literature in the field of data literacy, and provides reference for researchers to carry out theoretical research and practical exploration. 206 data literacy documents from CNKI database were analyzed using key word co-occurrence network and cluster analysis to generate knowledge maps of research authors, research institutions, research hotspots and research time trends in the field of data literacy. This paper focused on five automatic clustering of key word co-occurrence networks, such as information literacy, education big data, scientific data literacy, data literacy education, scientific data. The results showed that data literacy research was driven by national policies, and the overall research level needed to be improved. The library was the main position of data literacy education, and the content of data literacy research was mainly foreign education cases, and the research methods were mainly theoretical research, while the empirical research was insufficient.

Key words: data literacy; CiteSpace; key word co-occurrence; visualization analysis

(责任编辑: 金云波)