

# 基于 CiteSpace 知识图谱的数据要素研究： 热点、趋势与展望

王孟欣<sup>1</sup>，李静<sup>2\*</sup>，刘璐<sup>2</sup>

1. 广州大学 金融研究院，广东 广州 510405

2. 广州大学 经济与统计学院，广东 广州 510006

DOI:10.61369/ASDS.12186

**摘要：** 本文运用 CiteSpace 分析了 2016–2024 年数据要素研究的关键词共现、时序聚类以及突现情况。结果显示，国内外相关研究均呈增长态势但侧重点存在显著差异。国内研究聚焦理论体系与政策导向，以“数据要素”和“数字经济”为主线，丰富数字经济理论与实践；国外研究则强调跨领域应用，基于个人兴趣和市场需求进行多元化研究，促进了数据科学、医疗健康、环境科学等学科交叉发展，为数据流通与开放奠定理论基础。最后文章展望了数据要素未来的研究方向。

**关键词：** 数据要素；数字经济；文献计量；CiteSpace

## Research on Data Elements Based on CiteSpace Knowledge Graph: Hot Spots, Trends and Prospects

Wang Mengxin<sup>1</sup>, Li Jing<sup>2\*</sup>, Liu Lu<sup>2</sup>

1. Institute of Finance, Guangzhou University, Guangzhou, Guangdong 510405

2. College of Economics and Statistics, Guangzhou University, Guangzhou, Guangdong 510006

**Abstract:** This paper uses the CiteSpace to analyze the keyword co-occurrence, time-series clustering, and emergence of data element research from 2016–2024. The results show that both domestic and international studies show a growth trend, yet there exist significant differences in their focuses. Domestic researches focus on the theoretical system and policy orientation, and enrich the theory and practice of digital economy by focusing on "data elements" and "digital economy"; foreign researches emphasis cross-disciplinary application, conducting diversified research based on personal interests and market demands, which has promoted the development of multiple disciplines such as data science, healthcare, environmental science and other disciplines, laying a theoretical foundation for the circulation and opening of data. Finally, the article outlines the future research directions for data elements.

**Keywords:** data elements; digital economy; bibliometrics; CiteSpace

### 引言

数据作为新型生产要素，是数字化、网络化、智能化的基础，已快速融入生产、分配、流通、消费和社会服务管理等各环节，深刻改变着生产方式、生活方式和社会治理方式。2019 年美国发布《联邦数据战略与 2020 年行动计划》，明确将数据作为一种战略性资源进行开发。2020 年欧盟委员会公布了一系列围绕数据资源的发展规划，包括《欧洲数据战略白皮书》《人工智能白皮书》等多份文件，详细概述了欧盟未来五年实现数字经济所需的政策措施和投资策略，以及构建一个真正的欧洲数据统一市场的发展目标。2021 年国务院发布的《“十四五”数字经济发展规划》指出“数据要素是数字经济深化发展的核心引擎”。这些国内外政策的颁布和实施，显示了数据要素作为推动经济增长新动能的重要地位，也预示着全球范围内数据要素市场的竞争和合作将更加激烈，凸显数据要素在全球经济竞争中的核心地位。据《全球数字经济白皮书（2023 年）》显示，数字经济已成为全球经济增长的重要驱动力，数据要素作为其核心组

基金项目：2023 年度国家自然科学基金重大项目“数字贸易核算与全球产业链重构统计研究”（23&ZD127）；  
2023 年度全国统计科学研究重点项目“数据生产要素价值测度的理论与方法研究”（2023LZ030）。

作者简介：

王孟欣，广州大学金融研究院，教授，博士生导师，研究方向为数字经济与数字贸易；

刘璐，广州大学经济与统计学院，博士研究生，研究方向为宏观经济统计分析；

通讯作者：李静，广州大学经济与统计学院，硕士研究生，研究方向为宏观经济统计分析。

成部分，其价值和重要性无可估量。

为了全面、多维度地分析有关数据要素的研究状况，梳理数据要素的研究热点并分析其演变趋势，成为一项具有重要意义的研究工作。本研究基于 CiteSpace 可视化文献分析软件，以 CNKI 及 WOS 数据库中的相关中外核心文献为研究对象，从关键词的共现、时序聚类、突现等方面就有关数据要素的研究状况进行分析；并通过绘制研究知识图谱，结合国内数据要素相关政策和会议文件、国外数据要素相关立法和应用实践，把握数据要素的研究热点，并分析其发展趋势，为进一步推进相关研究提供重要的文献参考。

## 一、数据来源

### （一）文献来源

“生产要素”的概念在国外有着悠久的历史，其根源可追溯到 1662 年 William Petty 的税收理论，在 1848 年由 John Stuart Mill 提出为“生产的必要条件”（requisites of production），其后经 Alfred Marshall 的发展被正式表述为“生产要素”（factors of production）。然而，这一经典理论框架在数字时代的演变中逐渐迎来了新的诠释。1991 年 Tim Berners Lee 定义了超文本规范，标志着万维网的诞生。随后，数据的激增和复杂性催生了大数据的概念和技术。在这一背景下，数据不再仅仅是信息的载体，而是逐渐转变成为一种生产要素。1996 年 Don Tapscott 首次提出“数字经济”的特征即以数字方式呈现信息流的经济模式，在这一经济模式中数据成为连接生产和消费的桥梁。随着信息技术的飞速发展和互联网的广泛普及，数据的价值得到了深层次的挖掘与开发利用。人们开始意识到，通过收集、分析和利用数据，可以更加精准地把握市场需求，优化生产流程，提高产品质量和效率。这一趋势在 21 世纪初逐渐显现，并随着云计算、物联网、人工智能等技术的不断成熟而加速发展。2016 年 G20 杭州峰会正式提出了“数字经济”的概念，由此揭开了数据要素作为数字经济发展的核心生产要素的序幕。随着数字经济的蓬勃发展以及一系列数字经济政策的密集出台，经济社会中的数据要素积累日益丰富，从而引发数据要素与资本、劳动力、技术等传统生产要素的深层次交互，共同推进经济社会的全面转型升级<sup>[1]</sup>。这一年，欧盟不仅通过了《通用数据保护条例》（GDPR），还与美国签署了“隐私盾”协议，共同致力于保护跨大西洋数据流动中的个人隐私。与此同时，英国对“数字经济法”进行了全面修订，并颁布了《数字经济法（2017）》；加拿大政府也提出了数字宪章实施法案，旨在升级数字基础设施、保障消费者数据安全，以及推动数字经济的发展。在此背景下，国内外关于“数据要素”的文献研究呈现出显著的增长趋势。鉴于此，本研究聚焦于数据要素及其相关领域的演进，以 2016 年为研究起点，于 2024 年 12 月 31 日进行了系统性的文献检索。

本研究对于中文文献检索依托于 CNKI 数据库，采用高级检索策略，以“数据要素”为主题，并辅以“数据要素+数据生产要素+数据要素市场+数据要素资源+数据要素配置+数据要素产权”等关键词，进一步筛选出 CSSCI 来源的高质量学术期刊文献，最终选出了 2179 篇文献作为分析样本。英文文献来源于 WOS 核心合集，鉴于国内外学术用词表达差异，围绕“data factor(s)”与“data element(s)”这两大主题，以“data

factor(s)”与“data element(s)”为关键词并聚焦于“Article”，严格筛选“SCI”或“SSCI”索引文献，以确保学术研究的权威性和深度。经检索查找共得到 4871 篇文献，与中文文献相匹配共同构成本次研究的文献基础，并实施深入的文献分析。

### （二）研究文献的发文量分析

基于 2016–2024 年数据要素领域文献发文量的分析，本文绘制了国内与国外文献发文量的对比图（图 1）。图 1 显示国内外发文量及趋势存在明显差异。国内相关文献发文量在 2016 至 2019 年间经历了小幅波动，年发文量较少仅在 50 篇左右；随后 2020 年增长到了 94 篇，此后便呈现出显著的爆发式增长态势，到 2024 年增长到了 904 篇，并且在当年国内的发文量超过了国外发文量。国内发文量的急剧增加，可能归因于国家政策与资金对数据要素研究的重点扶持，以及学术界对该领域日益增长的兴趣与投入。在 2016 年“数字经济”提出之后，虽然社会对数字技术、数据、数据资产等的研究日益重视，但并没有将数据上升到生产要素层面。2019 年 10 月中共十九届四中全会首次明确将“数据”作为生产要素参与社会分配，这是中国首次在国家层面将数据作为一种新型生产要素，标志着数据要素市场化配置改革的正式启动，由此引发了学术界对数据及数据要素、数据生产要素等研究的热潮。2019 年学术界和教育界举办了如国际数据委员会学术大会、第四届大数据分析国际会议、第六届数据科学大会等研讨会，会议内容包括数据要素与数字经济的关系、数据要素流通、数据要素价值化等，由此进一步推动了关于数据及数据要素的研究，次年相关研究的发文量就达到了 94 篇。此后随着研究的深入，人们对数据要素日益重视，推动相关研究井喷式增长，直到 2024 年发文量达到了 904 篇。

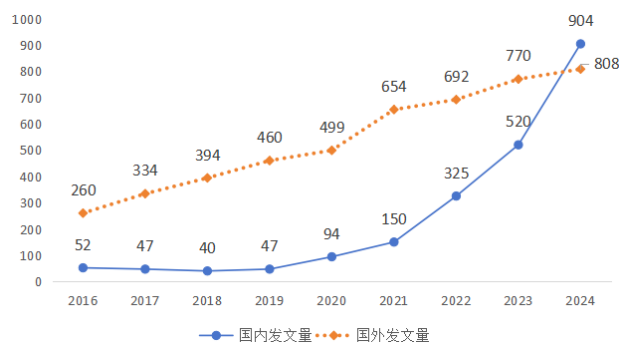


图 1 2016–2024 年有关数据要素的国内外研究发文量（篇）

对比同期，国外发文量则保持了较为平稳的增长轨迹，显示出研究活动的连续性和稳定性。国外在数据要素领域的研究最早可追溯到 20 世纪 90 年代公共部门信息市场化研究，并且伴随着大数据的步伐，国外研究更加聚焦于数据的使用价值、市场交易机

制以及与之相关的技术能力学习，如数据挖掘、机器学习等前沿技术，这种早期的学术积累为国外学者在数据要素的研究体系、方法和理论框架上提供了坚实的基础。因此在2016年之后，国外的研究理论并非从零开始，而是在原有丰富研究成果的基础上，进行了时代性的调整和创新。2016年有关数据要素的研究文献有260篇，约为国内研究的5倍多。2016年之前国外学术界对数据要素的研究平稳增长，基本呈现线性增长趋势，到2024年相关文献达到808篇。

国内外研究发文量态势差异显著，主要归因于研究关注点的驱动因素不同。国内学者对于研究领域的关注度往往紧密贴合国家政策的导向，这种政策驱动的模式在一定程度上导致了研究热点容易受国家政策的影响，在国家政策力推数据要素之后短期内使得相关研究呈现爆发式增长。相比之下，国外学者的研究动力则显得更为自主且富有一定的前瞻性。在大数据浪潮初现端倪之时，他们便洞察到“数据”的潜在价值，并基于个人兴趣、学术追求以及市场需求等多重因素，较早地展开了深入且持续的研究，这种前瞻性的研究视角使得他们在有关数据要素的研究成果上保持了一定的领先优势；但随着我国有关数据要素的研究成为热点并持续推进，相关研究文献大量涌现并很快缩小了与国外的研究差距，并且在发文数量上形成反超。

从整体趋势分析，无论是国内还是国外，关于“数据要素”的研究文章数量均呈现出逐年递增的趋势，这标志着该研究领域正不断拓展其边界，并吸引着全球范围内的广泛关注与探索。国内研究的快速增长，不仅体现了我国对数据要素战略价值的深刻认识，也预示着在国际竞争舞台上，中国将在数据要素领域发挥更加关键和引领性的作用。

## 二、“数据要素”研究领域关键词共现网络图谱分析

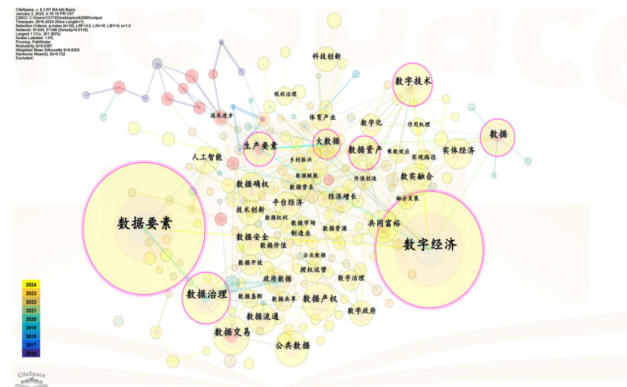
关键词共现网络是依据所有样本文献中“关键字（Keywords）”共同出现的次数所构建的网络，通过统计关键词的频次以及关键词之间的关联强度，来揭示某一领域内的研究热点和主题结构。关键词共现网络往往通过知识图谱的形式来反映，图谱中的每一个节点对应一个关键词，节点的数量反映关键词的个数，节点越多代表关键词的个数越多；节点的大小则直观展示了各关键词的出现频次，节点越大反映该关键词的频次越高；节点间连线表明关键词共现于同一论文，连线粗细则体现了关键词的共现强度。

在共现网络图谱分析中，节点中心性被视作评估一个关键词在所有样本文献中普及程度的关键指标，它通过计算所有经过该节点的节点对最短路径的数量，占这个节点对的所有可能最短路径数量的比例，来评估该节点在网络连接中的关键性或中心地位。这就意味着，如果一个关键词频繁地出现在其他两个关键词的最短连接路径上，那么就代表它具有高中心性，它为发现学科的研究主题演进趋势提供依据。一般认为，中心性值（Centrality）大于或者等于0.1是高中心性节点，即代表该关键词是所有样本文献中讨论最多的主题，并且是多个研究领域的交汇

点，对整个研究领域的发展有着重要的影响<sup>[2]</sup>。因此，分析关键词研究热点时需综合考量其频次与中心性。

### （一）国内文献关键词共现网络图谱分析

对国内文献数据要素领域的关键词进行共现网络分析，图2显示了文献关键词共现网络图谱，表1则显示了国内文献高频共现词TOP10情况。图2中245个节点与348条连接线清晰勾勒出关键词间的紧密联系，图中“数据要素”、“数字经济”、“数据治理”、“数字技术”及“数据交易”等核心关键词显著突出，表明这些关键词出现频次大且重要性较高。进一步地分析表1中的国内关键词列表发现，“数据要素”、“数字经济”、“数据治理”、“数字技术”、“数据”、“生产要素”以及“数据资产”的中心性超过0.1，表明这些概念是当前的研究热点，更揭示了它们作为连接多个研究领域的桥梁作用。



> 图2 有关数据要素的国内文献关键词共现网络图谱

表1 国内研究文献高频共现关键词 TOP10

序号	关键词	频次	中心性
1	数据要素	604	0.24
2	数字经济	473	0.35
3	数据治理	94	0.10
4	数字技术	67	0.16
5	数据交易	58	0.09
6	数据	57	0.12
7	数据产权	49	0.04
8	公共数据	49	0.04
9	生产要素	47	0.13
10	数据资产	46	0.10

在此基础上，可以从以下三个方面深入探讨数据要素的研究：（1）数据要素的基本理论研究至关重要。这包括辨析数据要素、数据资源与数据资产的概念，并探讨在数字技术迅猛发展的背景下，数据作为数字经济时代的第五生产要素所应具备的基本条件和特征属性。这些研究有助于我们更深入地理解数据的本质和价值，为数据的合理利用提供理论基础。（2）数据作为生产要素在市场交易流通中所面临的问题同样值得关注。例如，如何有效地进行数据治理以确保数据的质量和安全？如何明确数据的产权归属以促进数据的合规交易？以及如何将数据作为资产纳入会计核算体系？这些问题的解决对于推动数据要素市场的健康发展具有重要意义。（3）政策法规与标准体系建设在数据要素领域同样不可或缺。特别是针对公共数据这一数据要素的重要组成部分，如何通过政策法规的引导和规范来确保数据的合法获取、有效使用和合规交易，是当前亟需解决的关键问题。这些政策法规的制定和实施将为数据要素市场的健康发展提供有力的制度保

障。综上，前10个高频关键词共同描绘了我国数字经济蓝图：数据作为新兴的生产要素，在技术赋能和治理完善下，正向高价值数据资产转化；进一步促进交易市场繁荣，驱动实体经济转型升级，为共同富裕奠定坚实的经济基础。

## （二）国外文献关键词共现网络图谱分析

对国外文献数据要素领域的关键词共现图谱进行分析，各关键词的共现网络图谱如图3所示，表2显示了国外文献高频共现关键词 TOP10情况。图3由238个节点与835条连接线构成，不仅涵盖了基础性关键词，还揭示了与研究主题密切相关的核心词汇，如“model”“risk factors”“impact”“health”“prevalence”等。结合表1国外共现关键词列表，可以看到国外研究的高频关键词的中心性都比较低，并且第4个到第10个关键词的频次差异不大，表明这些关键词虽然各自在其特定领域内占据重要地位，但并未形成明显的核心和主导地位，而是各自独立发展、相互补充。

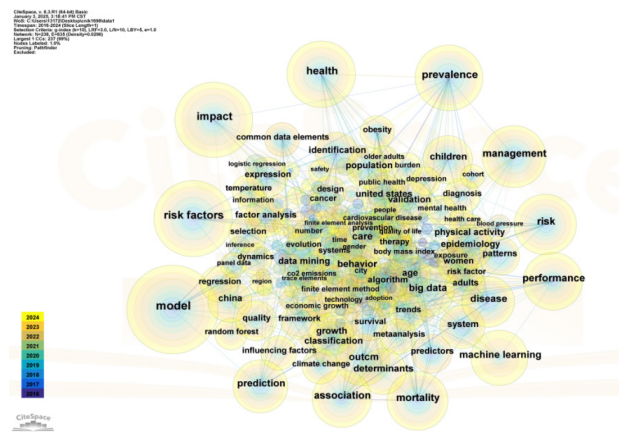


图3 有关数据要素的国外文献关键词共现图谱

表2 国外研究文献高频共现关键词 TOP10

序号	关键词	中文释义	频次	中心性
1	model	模型	318	0.05
2	risk factors	风险因素	239	0.02
3	impact	影响	233	0.04
4	health	健康	184	0.06
5	prevalence	流行率	182	0.07
6	management	管理	160	0.06
7	risk	风险	151	0.05
8	performance	绩效	144	0.09
9	mortality	死亡率	143	0.07
10	association	关联	135	0.07

尽管国外研究趋于多元化，但仍可以观察到一些主要的研究聚焦点：（1）模型构建与数据分析。以“model”为代表，旨在构建各类统计模型与机器学习模型，以处理和分析海量数据来理解和预测数据要素的行为和趋势。（2）风险因素与影响评估，以“risk factors”“impact”为代表，旨在识别和分析数据市场交易中的潜在风险因素，评估数据要素对经济社会的影响。（3）数据管理与绩效评估，以“management”“performance”为代表，旨在探讨数据质量管理、数据安全、数据治理以及数据系统能力方面的研究。（4）医疗健康领域，以“health”“prevalence”和“mortality”为代表，旨在利用数据要素来探讨疾病的流行率、健康影响因素以及死亡率等关键指标，进而更为准确地描绘疾病

的发病趋势和探寻其背后的影响因素。（5）跨学科融合研究，以“association”为代表，旨在使用数据关联技术，探索数据要素在多元领域中的创新应用，充分挖掘数据的潜在价值。

国内外研究在高频关键词上的差异，实质体现了两者研究线路的不同侧重点。国内学者研究侧重于相关理论的发展与体系的构建，注重从理论层面深入剖析数据要素的内涵、价值及作用机制，为数据要素的实践应用提供了坚实的理论基础；国外学者则是更加关注数据要素的使用价值，从应用角度出发，积极探索数据与其他学科领域的交叉应用，推动数据要素在各个领域中的广泛应用和市场化进程。

## 三、“数据要素”研究领域关键词时序聚类图谱分析

关键词聚类分析是在关键词共现网络图谱分析的基础上，通过引入时间维度，在图形中以时间轴展现群组内关键词的兴起、发展轨迹以及聚类情况，从而揭示研究主题的演变趋势和前沿动态；并且它还通过量化关键词间的相似性来形成聚类，一般情况下聚类序号越小，代表该聚类主题越热门<sup>[3]</sup>。与关键词共现网络图谱分析相比，它侧重于从时间和类别维度深入研究相关研究领域的进展情况。在一个时序聚类图谱中，节点标记了该关键词首次出现的年份，并用节点圆圈大小反映该关键词在考察时间段内出现的累积频次，这样能够避免不同年份频次的重复显示，节点圆圈越大反映该关键词出现的累积频次越高。在图谱中，从横向上看处于同一横线上的多个节点归属于同一个聚类类别，类别名称在最右侧一列显示；从列向上看，节点最上方小数字代表时间，表明位于同一列向上的多个节点同时出现。不同节点之间的弧线线条则勾勒出关键词间的共现关系，反映它们在所有样本文献中的联合出现模式，连接某一节点的弧线线条越多，说明该关键词具有越高的中心性。在关键词聚类图谱分析中，使用模块值（Modularity Q值）和平均轮廓值（Weighted Mean Silhouette S值）衡量评估关键词的聚类效果。其中模块值是衡量聚类结果内部连线密集程度的一个指标，是聚类内部连线的数量与所有可能连线的数量的比例；平均轮廓值（S值）是一种综合评估聚类内部紧密度和聚类间分离度的指标，通过计算每个数据点的轮廓系数，然后求其平均值来得到<sup>[4]</sup>。一般认为，当Q值 $\geq 0.3$ 时表明聚类结构显著；当S值 $\geq 0.7$ 时表示聚类结果高效<sup>[5]</sup>。

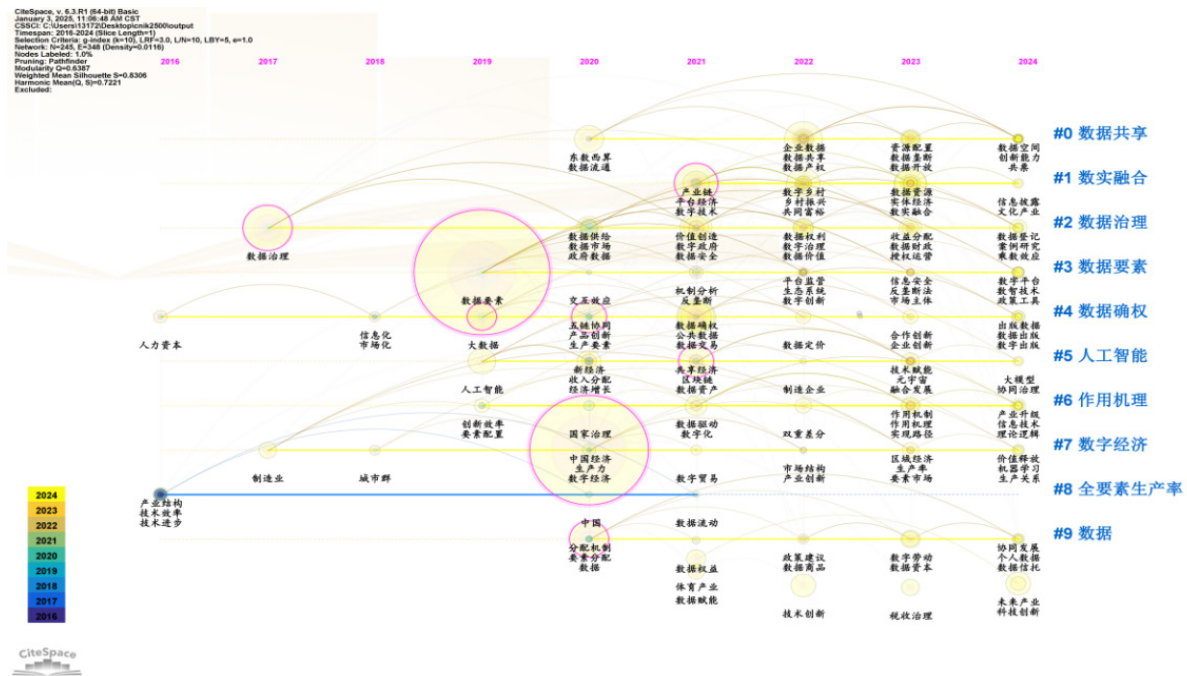
### （一）国内文献关键词时序聚类图谱分析

对国内文献数据要素领域的关键词进行时序聚类分析，图4显示了文献关键词时序聚类图谱，其上方显示了2016-2024年相关关键词出现的时间；右侧以序号标识出的词汇代表关键词的聚类和聚类名；图形当中的横线上的节点表示这些关键词归属于同一聚类。图4中Q值与S值分别为0.64和0.83，证明了聚类结果的可靠性和良好的聚类效果。为了保持图谱的清晰性，图中仅展示前10大聚类（#0~#9），以精炼地展现国内研究领域的核心热点和趋势。图4中显示国内研究关键词聚类依次为“#0数据共享、#1数实融合、#2数据治理、#3数据要素、#4数据确权、#5人工智能、#6作用机理、#7数字经济、#8全要素生产率、#9数

据”。表3则显示了文献热点主题与聚类表，包含了研究主题、聚类号、聚类名、聚类出现的时间以及主要节点。其中，研究主题的界定基于各聚类关键词进行归类，旨在精准提炼当前研究的核心热点。值得注意的是，图4与表3在呈现节点内容时存在差异，表3侧重于凸显每个聚类中的前5个最具影响力的关键词，而图4则根据关键词的频次来展现节点。因此，当一个聚类包含大量高频关键词时，图4会显示更多的节点，可能会导致部分节点未能在表3中体现；相反，若聚类中的关键词频次相对较低时，表3列出的关键词在图4反而会因为频次较低而不显示。

表3所示的归类结果表明，当前学者的研究主题可以分成三类：（1）数据要素的基础构建，包含聚类标签“#3、#5、

#9”，是数字经济大厦的基石，涵盖了数据要素的识别与整合、数字技术的融合与创新以及政策体系的规范与保障，为数字经济的繁荣奠定了坚实的基础<sup>[6-7]</sup>。（2）数据要素的流通交易，包含聚类标签“#0、#2、#4”，是数字经济活力的源泉，通过构建高效、透明的数据交易平台与生态系统，促进了数据资源的优化配置与价值释放，加速了数据要素向现实生产力的转化<sup>[8-9]</sup>。（3）数据要素的广泛应用，包含聚类标签“#1、#6、#7、#8”，是数字经济创新发展的动力所在，它渗透到实体经济的每一个角落，推动了数实融合与产业升级，催生新业态、新模式，为经济社会的高质量发展注入了强大的数字动能<sup>[10-11]</sup>。



> 图4 有关数据要素的国内文献关键词时序聚类图谱

表3 国内文献热点主题与聚类表

研究主题	聚类号	聚类名	时间	主要节点（LLR）
数据要素的基础构建	#3	数据要素	2022	数据要素；出版业；反垄断；数字创新；政策工具
	#5	人工智能	2021	人工智能；数据资产；经济增长；技术赋能；大模型
	#9	数据	2022	数据；数字劳动；数据信托；数据资本；马克思
数据要素的流通交易	#0	数据共享	2022	数据共享；数据流通；数据产权；企业数据；数字金融
	#2	数据治理	2021	数据治理；政府数据；数据安全；数字政府；数字经济
	#4	数据确权	2021	数据确权；数据交易；生产要素；数据定价；公共数据
数据要素的广泛应用	#1	数实融合	2022	数实融合；数字技术；实体经济；平台经济；数字鸿沟
	#6	作用机理	2022	作用机理；数字化；实现路径；实践路径；农业强国
	#7	数字经济	2021	数字经济；制造业；生产关系；要素市场；数字贸易
	#8	全要素生产率	2016	全要素生产率；技术效率；环境规制；技术进步；产业集聚

由图4并结合表3所示，自2016年起以“产业结构”“技术效率”“技术进步”等为关键词的研究开始出现，并在此基础上同年形成了“#8全要素生产率”聚类，该聚类的主要节点涵盖了“全要素生产率、技术效率、环境规制、技术进步、产业集聚”等，共同构成了推动全要素生产率增长的关键因素。特别地，数据要素作为新兴的生产要素，其在提升全要素生产率中的作用逐

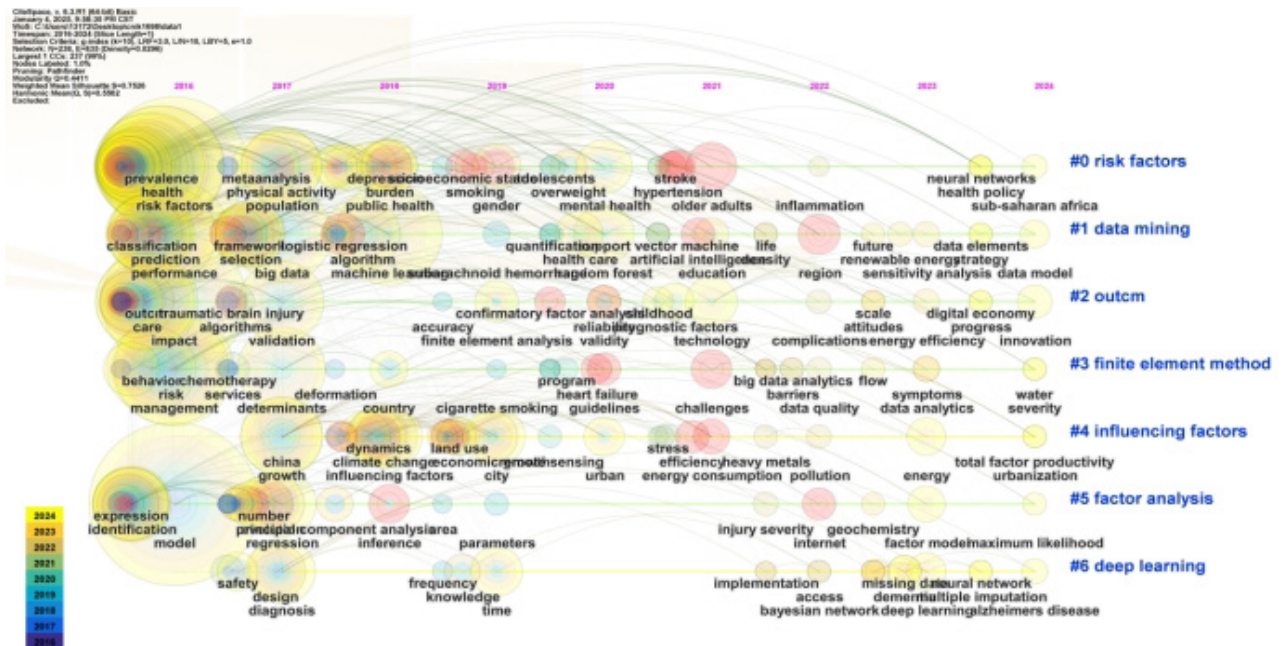
渐受到重视。通过数据的挖掘、分析、应用等，能更有效第配置资源，优化产业结构、提升技术效率和加速技术进步，并促进产业集聚效应的发挥，从而实现对全要素生产率的全面提升。同期以“人力资本”为关键词的研究也开始出现，但直到2021年才形成“#4数据确权”聚类，主要节点包括“数据确权、数据交易、生产要素、数据定价、公共数据”等，显然此方面的研究初期比

较分散，从初期的“人力资本”研究逐渐凝聚成“数据确权”的研究，指出了数据作为与劳动、资本等并肩的生产要素，使得人力资本的研究也渐渐涉及到数据产权、交易等问题，反映了数据要素研究从理论探讨转向实践应用。2017-2020年间相关研究热点比较分散，以“数据治理”“数据要素”“大数据”“数字经济”等为关键词的研究均有涉及，虽然没有形成明显的聚类，但也预示着数据在经济社会中的作用日益凸显，数据治理和数据要素的重要性也日益受到重视。2021年数据要素领域的研究迎来了新的突破，以“#2数据治理”“#4数据确权”“#5人工智能”以及“#7数字经济”等聚类为代表，研究开始深入到数据治理的具体实践、数据确权的政策体系、人工智能的应用前景以及数字经济的宏观影响等层面。这些聚类热词的出现，不仅丰富了数据要素领域的研究内容，还推动了相关理论的深化和拓展。2022年数

据要素领域的研究进一步呈现出多样化发展趋势，以“#0数据共享”“#1数实融合”“#3数据要素”“#6作用机理”“#9数据”等聚类作为标志，研究开始关注数据共享的机制与效果、数字经济与实体经济的融合发展、数据要素在经济社会发展中的作用机理以及数据的本质属性和价值挖掘等。

## (二) 国外文献关键词时序聚类图谱分析

对国外文献数据要素领域的关键词进行时序聚类分析，文献关键词时序聚类图谱如图5所示，表4则显示了文献关键词聚类情况。图5中Q值与S值分别为0.43和0.74，证明了聚类结果的稳定性和聚类效果的优异。国外研究关键词聚类依次为“#0risk factors、#1data mining、#2outcome、#3finite element method、#4influencing factors、#5factors analysis、#6deep learning”。



> 图5 有关数据要素的国外文献关键词时序聚类图谱

表4 国外文献热点主题与聚类表

研究领域	聚类号	聚类名	时间	主要节点 (LLR)	中文释义
医疗健康领域	#0	risk factors	2018	risk factors; hypertension; obesity; prevalence; public health	风险因素; 高血压; 肥胖; 患病率; 公共卫生
	#2	outcome	2018	outcome; therapy; confirmatory factor analysis; technology; prognostic factors	结果/成效; 治疗; 验证性因素分析; 技术; 预后因素
	#3	finite element method	2019	finite element method; determinants; myocardial infarction; behavior; management	有限元方法; 决定因素; 心肌梗死; 行为; 管理
数据科学领域	#1	data mining	2019	data mining; machine learning; big data; common data elements; random forest	数据挖掘; 机器学习; 大数据; 通用数据要素; 随机森林
	#5	factor analysis	2017	factor analysis; panel data; factor model; expression; identification	因子分析; 面板数据; 因子模型; 表达; 识别
	#6	deep learning	2021	deep learning; missing data; incomplete data; multiple imputation; predictive factors	深度学习; 缺失数据; 不完整数据; 多重插补; 预测因素
环境科学领域	#4	influencing factors	2019	influencing factors; carbon emissions; urbanization; climate change; economic growth	影响因素; 碳排放; 城市化; 气候变化; 经济增长

国外数据要素领域的相关研究起步比较早，伴随着数字技术发展的步伐，国外更加注重数据要素与数字技术融合，探索数据要素的创新应用。如图5所示，国外研究自2016年起便展现出显著的“集聚”特征，研究热词几乎同步涌现，如“management”“impact”“performance”“classification”“expression”

等，数字经济的发展带来海量数据爆发式增长，对数据分析、数据安全以及数据治理等提出了新的要求和挑战。在2017-2019年间，聚类标签的集中出现，如表4所示，“#0 risk factors”“#1 data mining”“#2 outcome”等表明该领域的在这一阶段的研究具有广泛的探索性和交叉性。此时，如何使用数字技术、构建合

理的数学模型去挖掘数据要素的潜在价值，并将其研究成果应用于环境科学、医学、统计学、管理学等领域成为了学者们的研究重点。到了2020年之后，数据要素与数字技术的融合趋势更为明显，进一步地拓宽了数据要素的应用领域。“#6 deep learning”作为最晚形成的聚类标签，不仅反映了数字技术发展的阶段性特征，也预示着数据要素与人工智能、大数据、机器学习等先进技术深度融合的未来趋势。整体而言，这些研究的发展轨迹逐步从“抽象概念”向“具象应用”过渡，突破了单一经济领域的界限，深入渗透到多个学科之中。

为了进一步厘清这些研究的多元化与跨学科特性，根据图5主要节点可将其重新归类（表4），表明国外研究的主要领域可以分为三类：（1）医疗健康领域，包含聚类标签“#0、#2、#3”，在该领域中数据要素是提升医疗服务质量、优化医疗资源配置、推动医疗创新的重要驱动力<sup>[13-14]</sup>。（2）数据科学领域，包含聚类标签“#1、#5、#6”，在该领域中数据要素是构建数据模型、进行统计推断、发现数据规律的核心研究对象<sup>[15]</sup>。（3）环境科学领域，包含聚类标签“#4”，在该领域中数据要素是理解地球系统、预测自然灾害、评估资源潜力等工作的重要基础<sup>[16-17]</sup>。随着计算机技术的飞速发展，特别是大数据处理、高性能计算、云计算、人工智能等技术的广泛应用，数据要素在数据科学的实践边界被极大地扩展，如医生通过病例数据把握疾病的发病机制、病变过程及患者个体差异，从而制定精准、个性化治疗方案；企业家利用微观市场数据分析消费者行为、市场趋势以及监控竞争对手

动态，从而调整投资策略、抢占市场先机；政府部门分析交通流量、环境检测等公共数据，提升城市管理的智能化水平。

#### 四、“数据要素”研究领域关键词突现图谱分析

关键词突现分析是揭示关键词出现频次激增的起始时间与强度的方法，反映了研究热点的兴起与衰落，直观地展示了该关键词在研究领域内的重要程度和受关注度的时间变化趋势。关键词从初现至结束的整个突现轨迹以红色横线标记（见图4和图5，若为黑白图，则以加粗为标记），突现轨迹的长度越长，说明该关键词在较长时间内保持高热度，其前沿性就越强。关键词的突现强度是通过计算该关键词在特定时间段内的出现频次与整个样本时间内该关键词的平均频次之差，再除以该关键词出现频次的标准差来衡量<sup>[12]</sup>。突现强度越高，意味着该关键词在该时间段内的影响力越大。

##### （一）国内文献关键词突现图谱分析

对国内文献数据要素领域的关键词进行突现分析，表5显示了国内文献前25个突现关键词的突现轨迹和强度。其中，国内关键词强度最高的是“生产要素（6.94）”，突现轨迹最长的是“产业结构（2016-2022）”，反映了国内对数据作为生产要素的重视程度不断提升，并且数据要素在产业结构调整和优化中发挥了重要作用。

表5 国内研究文献突现关键词 TOP25

关键词	强度	开始时间	结束时间	突现轨迹（2016 - 2024）
技术进步	5.86	2016	2019	██████████
技术效率	5.33	2016	2018	██████████
环境规制	3.67	2016	2020	██████████
要素禀赋	3.33	2016	2018	██████████
要素流动	3.31	2016	2021	██████████
影响因素	2.76	2016	2021	██████████
人力资本	2.55	2016	2018	██████████
产业集聚	2.45	2016	2020	██████████
产业结构	2.33	2016	2022	██████████
市场化	1.64	2018	2022	██████████
大数据	5.05	2019	2020	██████████
生产要素	6.94	2020	2021	██████████
数据	5.89	2020	2021	██████████
中介效应	3.41	2020	2022	██████████
国家治理	2.35	2020	2021	██████████
收入分配	2.18	2020	2022	██████████
异质性	1.7	2020	2022	██████████
数据市场	1.29	2020	2021	██████████
产权界定	1.28	2020	2022	██████████
经济增长	1.2	2020	2022	██████████
反垄断	3.3	2021	2022	██████████
数据交易	2.38	2021	2022	██████████
数据共享	1.9	2022	2024	██████████
扎根理论	1.59	2022	2024	██████████

为了进一步分析国内数据要素研究的变化趋势，本文搜集了自2016年以来国内有关“数据要素”的政策文件（见表6），结合表5对国内数据要素研究进行关键词突现分析，揭示了数字要素发展对数字经济的推动作用。2016年9月G20杭州峰会首次明确提出“数字经济”的概念，即以信息和知识的数字化为关键生产要素，以现代信息网络为重要载体、以有效利用信息通信技术为

提升效率和优化经济结构重要动力的广泛经济活动。此时“技术效率”“要素流动”“产业集聚”等关键词突现，说明国内研究聚焦于实体经济，探讨如何通过技术进步、要素优化以及产业集聚效应等来提升经济效率。这一时期，虽然数据的重要性已初露端倪，但研究重点尚未全面转向数据本身，而是更多地将其视为为推动实体经济转型升级的辅助工具。2017年12月中央政治局就“实





要素在不同领域的应用场景。在数据科学领域，关键词如“data mining”“logistic regression”与“inference”揭示了数据要素从海量数据中挖掘、逻辑分析至价值推断的完整流程。为促进数据流通与共享，这些国家推出一系列法案（详见表8），旨在完善数据要素市场机制，推动数据的开放利用。美国通过《开放政府数据法案》和《联邦数据战略2020年行动计划》等，确立了政府数据公开的基本原则和未来方向。欧盟通过《通用数据保护

条例》和《非个人数据自由流动条例》等，全面规范了个人数据和非个人数据的保护和管理；其余五个法律体系从不同领域对数据的发展和应用进行了具体的规定，构成了欧盟数字战略的法律基础体系。英国通过《数据保护法案》和《数据保护与数字化信息法案》等，更新和简化了数据保护框架，促进公共数据的开放和利用。这些立法和文件不仅为数据的流通与共享提供了法律保障，也为数字经济的发展奠定了坚实基础。

表8 国外有关“数据要素”的法律体系与概述

国家/组织	时间	法律体系	概述
美国	2018年	开放政府数据法案	确立了政府数据公开的基本原则，还详细规范了数据来源审查、数据清单更新、数据目录公开等方面的制度，形成了全面的信息立法体系。
	2018年	澄清域外合法使用数据法	由美国公司所拥有、监管或控制的数据无论是否存储在美国本土，美国均有权对其进行检查。
	2019年	联邦数据战略2020年行动计划	确立了未来10年联邦政府在数据开放和共享方面的方向。
	2024年	关于防止受关注国家访问美国人的大量敏感个人数据和美国政府相关数据的行政命令	限制特定国家对美国敏感数据的访问。
欧盟	2016年	通用数据保护条例	保护个人隐私数据。
	2019年	非个人数据自由流动条例	促进欧盟境内非个人数据自由流动，消除数据本地化限制。
	2020年	欧洲数据战略	为欧盟数据治理提供了战略指导。
	2020年	数字服务法案	明确数字服务提供者的责任，遏制大型科技的恶性竞争行为。
	2022年	数据治理法案	规范数据治理的基础设施
	2022年	数据市场法案	规范数字市场秩序
	2023年	数据法案	完善数据治理的法律基础
英国	2018年	数据保护法案	为个人数据保护奠定了基础性原则。
	2022年	数据保护与数字化信息法案	为公共数据开放奠定了坚实的法律基础。
	2023年	数据保护与数字信息（第2号）	更新和简化英国的数据保护框架，减轻数据跨境义务、消除国际贸易壁垒、鼓励自动化决策等。

2021年之后“artificial intelligence”的出现，标志着数据科学研究迈入智能化新纪元，利用AI技术挖掘数据价值，推动技术与社会各领域的深度融合。在医疗健康领域数据，关键词如“gene expression”“electronic health records”“evolution”等突现时间跨度基本覆盖了整个样本期，彰显了国外在医疗数据应用上的积极探索。美国国立卫生研究院推出“我们所有人”（“All of Us”）研究计划、欧盟委员会的“欧洲健康数据空间”（EHDS）以及英国生物库样本的基因组序列提供，均为医学研究提供了宝贵资源。随着AI技术的发展，美国、欧盟、英国均批准了基于AI的医疗设备和软件，如糖尿病管理AI、心脏病诊断AI算法等，推动了个性化医疗和精准治疗的进步。在环境科学领域，面对社会经济快速发展带来的空气污染与能源消耗等挑战，美国国家航空航天局NASA的地球数据搜索系统、欧洲环境署（EEA）的数据集以及英国自然环境研究理事会（NECR）的传感系统，均提供了大量环境数据支持相关研究。这些机构的研究方向聚焦于气候变化评估、空气质量改善及可持续发展路径等。特别是2024年1月英国通过“自然环境孪生能力”计划（TWINE）资助创新项目，旨在利用数字孪生技术推动环境科学变革。

## 五、研究结论与展望

本研究聚焦于“数据要素”这一前沿主题领域，通过系统搜集并深入分析国内外高质量核心期刊的相关文献，采用 CiteSpace

知识图谱分析，包括发文量统计、关键词共现分析、关键词时序聚类分析以及关键词突现分析等，展示了2016年-2014年间数据要素相关研究的发文量变化趋势、关键词的频次和中心性、关键词的出现时间和聚类关系以及关键词的突现轨迹和强度，并结合国内外数据要素领域内的相关政策体系及应用实践进行系统性阐述，旨在全面揭示并对比国内外在该领域的研究热点与趋势演进。

### （一）研究结论

研究结果显示，国内外在数据要素领域的研究均取得了显著进展，但各自的研究侧重点与特色有所差异。国内研究侧重于数据要素在推动数字经济实践方面的模式探索与应用创新；而国外研究则更加注重于数据要素跨领域应用的深度挖掘与理论拓展。其中，国内研究深入探讨了数据要素在数字经济中的核心地位，以及如何通过技术创新、政策引导和市场机制等手段，促进数据资源的有效整合与高效利用，进而驱动经济增长模式的转型与升级。国内研究则关注于数据要素在促进产业升级、优化资源配置、提升社会治理能力等方面的应用实践，为数字经济的可持续发展提供有力支撑。此外，从国内文献发文量趋势以及关键词频次激增等现象可以看出，国内研究受政策影响比较大。国内学者倾向于遵循国家政策的宏观导向，紧密围绕国家发展的实际需求，深入探究数据要素发展的潜在轨迹及其赋能数字经济的有效机制，旨在系统性地探讨与解析如何高效发挥数据要素的功能与效能，从而提升国家的综合竞争力。

相较之下,国外研究则展现出更为宽泛与深入的探索视角。国外学者倾向于采用跨学科的研究范式,将数据要素与数据科学、环境科学医疗健康等多个学科领域深度融合,揭示数据要素在不同应用场景下的独特功能机制与影响路径,形成了丰富多元的研究成果体系。这些研究不仅深化了对数据要素本质属性的理解,也为全球数字经济的理论构建与实践探索提供了宝贵的学术贡献与经验借鉴。此外,由国外文献高频关键词的低中心性可以看出,国外关于数据要素的研究主题相较广泛且分散,国外学者在探索数据要素的应用层面时,更多地是基于个人的研究兴趣以及市场需求的驱动,而非某一统一的政策导向或集中的研究议题。

## (二) 研究展望

国内外研究成果不仅极大地丰富了数据要素领域的学术内涵与外延,也为未来数字经济的全球发展指明了方向,提供了坚实的理论支撑与实践指导。基于此,进一步地展望该领域的未来发展趋势与研究方向,具体内容如下:

1. 数据要素的基础理论研究。包括数据要素的标准化、价值评估、要素流通与交易机制等基础理论问题需要进一步研究。在标准化方面,需构建完善的数据标准体系,涵盖数据基础设施、资源、技术、流通共享和隐私保护等多个维度,以促进数据要素的激活和应用。在价值评估上,需建立科学的评估体系,量化分析数据商品与服务经济效益,确保数据资产的合理定价,推动数据资源向数据资产转变。此外,要素流通与交易机制也是研究重点,需构建安全、可信、高效的数据流通交易市场,促进数据要素的市场化配置和价值释放。这些基础理论的深入研究将为数据要素的高效利用和数字经济的高质量发展提供坚实支撑。

2. 数据要素跨学科融合与综合研究。随着数据要素在各行各业的广泛应用,未来研究将更加倾向于跨学科、跨领域的深度融合。经济学、计算机科学、生物信息学、管理学、统计学等多个学科的交叉研究将成为常态,共同探索数据要素在复杂系统中的潜在价值。特别是随着人工智能技术的不断进步,借助大数据分析 with 机器学习的强大能力,从浩瀚的数据海洋中精准萃取核心信息,预判未来趋势,并据此优化决策流程,为各行各业带来一场前所未有的智能化转型与升级浪潮。数据要素与人工智能领域的融合,能够促进技术创新和产业发展,推动数字经济与实体经济的深度融合。

3. 数据要素的生态体系建设。研究如何构建和维护健康的、可持续发展的数据要素生态体系,要从数据确权、数据安全和隐私保护入手。首先要明确数据的产权归属和交易规则,为数据的合法流通和价值释放奠定坚实基础。通过法律法规和政策引导,清晰界定数据的所有权、使用权等,并建立数据资产登记制度,并建立数据资产登记制度,是数据成为可量化、可管理的资产。其次在数据安全与隐私保护方面,要注重数据全生命周期的安全管理,包括数据采集、存储、处理、传输、共享、销毁等各个环节的安全保障技术与方法。同时,数据伦理、隐私权利保护等法律与政策研究也将得到加强,以构建健康、可持续的数据生态系统。

4. 聚焦有关中国数据要素实际问题的研究。作为世界第二大经济体和数字经济大国,中国在数据要素的研究与应用方面拥有独特的优势和需求。一方面,需要深入剖析中国数字经济发展的现状、问题和趋势,探索数据要素在推动产业升级、促进经济高质量发展中的具体作用机制和路径。另一方面,也要推动数据要素在社会治理、公共服务、医疗健康等领域的创新应用,为不断提高人民生活幸福水平提供有力支撑。

## 参考文献

- [1]张可云和庄宗武.2024,《数字经济政策强度对制造业稳就业的影响研究》.吉林大学社会科学学报第4期:155-171+239.
- [2]邱均平和吕红.2013,《基于知识图谱的国内知识管理发展研究》.情报学报第5期:548-560.
- [3]盛亚和鲁晓玮.2021,《利益相关者管理理论研究的主导逻辑与议题框架:基于Web of Science核心数据库的研究》.商业经济与管理第4期:38-51.
- [4]董彦斌,潘海珠,王磊等.2024,《土壤盐渍化遥感研究的文献计量与可视化分析》.中国农业资源与区划第9期:236-248.
- [5]陈悦,陈超美,刘则渊等.2015,《CiteSpace知识图谱的方法论功能》.科学学研究第2期:242-253.
- [6]叶雅珍,刘国华,朱扬勇.2019,《数据资产相关概念综述》.计算机科学第11期:20-24.
- [7]占智勇,徐政,宁尚通.2024,《数据要素视角下新质生产力创新驱动的理论逻辑与实践路径》.新疆社会科学第3期:43-52.
- [8]任保平和贺海峰.2024,《按照高标准市场体系要求培育数据要素市场》.福建论坛(人文社会科学版)第3期:42-58.
- [9]张明,路先锋,吴雨桐.2024,《数据要素经济学:特征、确权、定价与交易》.经济学家第4期:35-44.
- [10]王德祥.2022,《数字经济背景下数据要素对制造业高质量发展的影响研究》.宏观经济研究第9期:51-63+105.
- [11]范德成和肖文雪.2024,《环境规制、数据要素配置水平与产业结构调整》.统计与决策第2期:105-110.
- [12]张恩祥,乔琴,郑玉萍等.2023,《生态产品价值实现研究进展——基于CiteSpace的文献计量分析》.中国国土资源经济第11期:26-35+74.
- [13]Bruland, P., McGilchrist, M., Zapletal, E. et al. 2016, "Common Data Elements for Secondary Use of Electronic Health Record Data for Clinical Trial Execution and Serious Adverse Event Reporting". BMC Med Res Methodol 16, 159.
- [14]Vest JR, Adler-Milstein J, Gottlieb LM, Bian J, et al. 2022, "Assessment of Structured Data Elements for Social Risk Factors". Am J Manag Care.28(1), pp.14-23.
- [15]Sengottaian S, Natesan S, Mathivanan S. 2017, "Weighted Delta Factor Cluster Ensemble Algorithm for Categorical Data Clustering in Data Mining". International Arab Journal of Information Technology 14, pp.275-284.
- [16]Emeline Lequy, Nicolas P.A. Saby, Ilia Ilyin, Aude Bourin, St é phane Sauvage, et al.. 2017, "Spatial Analysis of Trace Elements in a Moss Bio-monitoring Data over France by Accounting for Source, Protocol and Environmental Parameters". Science of the Total Environment 590-591, pp.602-610.
- [17]Tessari-Zampieri, M.C., Sarkis, J.E.S. Barbieri, C.B. 2022, "Metrological Aspects of Platinum Group Elements Atmospheric Deposition in Roadside Tree Leaves: Uncertainties and Environmental Data Interpretation". Water Air and Soil Pollut 233, 91.