

关于国内新型冠状病毒相关文献的分析研究

——基于中国知网抗击新冠疫情知识服务平台

刘 海, 姚兴兴

武汉工程大学光电信息与能源工程学院 数理学院, 湖北 武汉

收稿日期: 2022年12月2日; 录用日期: 2023年1月2日; 发布日期: 2023年1月10日

摘 要

目的: 应用文献计量学对国内关于新型冠状病毒的相关文献进行分析和研究, 挖掘国内新冠疫情下不同时期的研究热点和发展态势, 完善有关的数据材料, 为该主题的有关研究提供参考。方法: 选取中国知网抗击新冠肺炎疫情知识服务平台的文献, 利用VOSviewer、BiComb等文本分析工具, 配合Python语言编程, 对获得的8038篇中文文献进行分析并可视化结果。结果: 高频关键词为作用机理、中医药等, 影响力最大的研究领域为医药卫生, 发文量最多的机构为华中科技大学, 研究热点集中在呼吸系统疾病, 刊文量最多的刊物是《中华流行病学杂志》, 有科研基金支持的文章有6089项。结论: 疫情对各行业的影响逐步扩大, 国民对疫情的关注点随时间不断变化, 不同领域的研究在国家及社会的支持下陆续开展, 高水平论文的发表也屡见不鲜。

关键词

新冠疫情, 文献计量, 数据分析, 共现分析

Analysis on Domestic Literature Related to Novel Coronavirus

—Knowledge Service Platform for Combating COVID-19 Epidemic Based on CNKI

Hai Liu, Xingxing Yao

School of Optical Information and Energy Engineering, School of Mathematics and Physics, Wuhan Institute of Technology, Wuhan Hubei

Received: Dec. 2nd, 2022; accepted: Jan. 2nd, 2023; published: Jan. 10th, 2023

Abstract

Objective: To apply bibliometrics to analyze and study the related literature on novel coronavirus in China, explore the research hotspots and development trends in different periods of the COVID-19 in China, improve the relevant data materials, and provide reference for the related research on this topic. **Methods:** We select the literature from the knowledge service platform of CNKI to fight against COVID-19, use text analysis tools such as VOSviewer and BiComb, and cooperate with Python language programming to analyze the 8038 Chinese literature obtained and visualize the results. **Results:** The high-frequency keywords were mechanism of action, traditional Chinese medicine, etc. The research field with the greatest influence was medicine and health, the institution with the largest number of papers was Huazhong University of Science and Technology, the research focus was on respiratory diseases, the journal with the largest number of papers was the Chinese Journal of Epidemiology, and there were 6089 articles supported by scientific research funds. **Conclusion:** The impact of the epidemic on various industries is gradually expanding, and people's concerns about the epidemic have changed over time. Research in different fields has been carried out successively with the support of the state and society, and the publication of high-level papers is not uncommon.

Keywords

COVID-19, Bibliometrics, Data Analysis, Co-Occurrence Analysis

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

2019 年底爆发的新型冠状病毒(以下简称“新冠”)疫情,对中国人民甚至全球人民来说都是一次重大的考验,疫情不仅影响着人们的学习、工作和生活,还制约着全球经济的发展和社会的进步。由于疫情对人类生命安全和健康带来了重大威胁,所以大众对新冠知识的需求和关注度水涨船高,国内外不同领域的科研人员也针对新冠疫情开展了许多研究,产出了大量科研论文[1]。

文献计量学是通过对已发表的文献进行定量分析和描述,评价研究状态,预测发展趋势,从宏观角度把握相关领域研究现状的方法。针对新冠领域,国内也有部分专家和学者做了相关的研究,例如张荣娜等人基于 Web of Science 核心合集对新冠相关文献进行了文献计量学分析,陈锐锋等人在大数据视域下开展了相关研究,郭婷等人针对 2019 年 12 月~2020 年 3 月的相关文献数据做了研究,陈玉恒等人以 CNKI 和 PubMed 数据库的数据展开探索[2]-[9]等,这些文章都从一定程度上丰富了国内新冠领域的研究。但考虑到上述文章存在研究时间较久远或研究范围较宽泛的问题,没有针对近两年国内的新冠形势进行论述,进而导致国内的研究缺乏相关数据和资料,故基于中国知网抗击新冠疫情知识服务平台(以下简称“新冠知识平台”),对该平台的中文文献开展文献计量学分析,统计分析相关文献的各项数据,并以可视化形式展示各时间段新冠研究的热点分布,为国内各行业工作人员及普通民众提供参考。

2. 研究设计

2.1. 研究对象

中国知网凭借优质的资源内容、领先的技术方法和专业人性的服务,在学术科研界享有极高的声誉,其学术性、科研性和权威性已得到各行各业的学者认同。为保证文献的研究内容都是新冠领域的且文献具有一定的影响力,本研究选取来自中国知网新冠知识平台的中文文献数据。该平台数据涵盖了科学引文索引(SCI)、中文社会科学引文索引(CSSCI)、中国科学引文数据库(CSCD)、工程索引(EI)、北大核心等多种类型的期刊,通过限定文献发表的日期为2020年01月01日至2022年09月30日,文献类型为期刊对文献数据进行纳入,并对数据进行清理。

2.2. 研究方法

文献计量学是一门用数学和统计学的方法,定量地分析一切知识载体的交叉科学,属于图书情报学分支科学。通过文献计量学分析,能将原本独立的文献汇集成一个大批量的文献集,深度挖掘出文献间隐藏的信息,从而定量的分析文献规律,清晰的把握计量对象特征。本研究从新冠疫情相关文献发文量、关键词共现、学科主题及类别、刊物数量及发文机构、检索类型、资助基金等多方面进行科学合理的分析,并将结果以可视化的形式展现。为保证对数据进行文献计量分析时的科学性、合理性和客观性,采用定性和定量研究相结合的方法,使用CiteSpace、VOSviewer和BiComb软件作为文献数据统计软件,辅以Python语言编程,得到相关结果并给出结论。CiteSpace是一款着眼于分析科学文献中蕴含的潜在知识,在科学计量学、数据可视化背景下逐渐发展起来的一款引文可视化分析软件[10]。VOSviewer是荷兰莱顿大学科技研究中心的Van Eck和Waltman开发的一款软件,主要面向文献数据,侧重科学知识的可视化[11]。BiComb是一款文本挖掘的基础工具软件,可对国际上诸多数据库的文献记录进行读取分析,并允许用户对系统功能进行修改、增加等拓展[12]。将中国知网新冠知识平台中检索到的文献数据导入Excel电子表格中,编写程序对数据进行清洗并规范数据格式,最终纳入有效文献8038篇,其中2020年共5551篇,2021年共1525篇,2022年截至9月共962篇。

3. 结果分析

3.1. 发文量分析

发文量分析能获得月发文量数据,月发文量大小反映出当月国民对新冠疫情的关注程度。记录每篇文献的发表时间,统计出月发文量数据如图1所示,可以看到在2020年中,1月的发文量很低,2月急剧增长,2、3、4月是发文量最多的三个月,其中3月更是高达1399篇,5、6月发文量在450篇左右,7月开始月发文量逐渐下降并趋于稳定,最终保持在200篇左右;在2021年中,月发文量稳定在150篇左右,偶有发文量较低的月份如11月,只有71篇文献,总的来看,整年的发文量数据无大幅波动;2022年截至9月的月发文量稳定在100篇左右,无大幅波动。纵观整个时间轴,月发文量最多的时间段集中在2020年的2、3、4三个月份,后续月份月发文量比较稳定,结合2020年初国内严峻的新冠疫情态势,加之国民急需快速了解有关新冠肺炎疫情防控的知识,与发文量变化曲线的各时间点吻合,从侧面佐证了所选数据和所用分析方法的合理性和科学性。

3.2. 关键词共现分析

关键词共现分析能得到出现频次和关联强度较高的关键词,对把控新冠相关文献的主旨内容和研究方向有着指导作用[13]。某一关键词的词频指该关键词出现的总次数,关联强度指文献中出现其它关键词

同时出现该关键词的次数总和。假设共有 M 篇文献, 其中 N ($N \leq M$) 篇文献包含预设关键词, 记这 N 篇文献中的第 i 篇文献有 K_i 个关键词, 则该预设关键词的频次和关联强度可表示为:

$$\text{频次} = N \quad \text{关联强度} = \sum_{i=1}^N (K_i - 1)$$

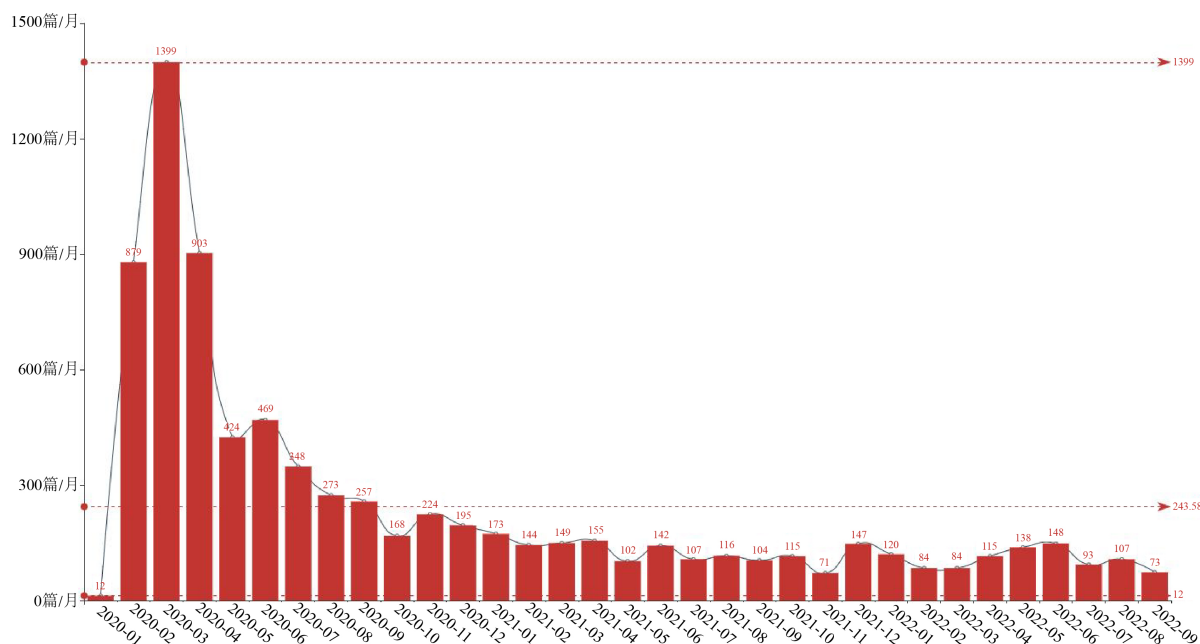


Figure 1. Monthly distribution of COVID-19 related literature

图 1. 新冠相关文献的月发文量分布图

使用 VOSviewer 软件进行关键词共现分析, 考虑到 VOSviewer 软件支持的文件格式存在局限, 故使用 Python 语言编程将文献数据改写成 Refworks 格式并导入到软件中进行关键词共现分析, 删去“新冠疫情”、“新冠肺炎”等检索词, 合并“疫情防控”、“防控”等同近义词, 最终获取到关键词排序, 统计频次为 1 的关键词数量为 8913 次, 由 1973 年 Donohue 根据齐普夫第二定律提出的高低频词界定公式[14]

$$T = \frac{1}{2}(\sqrt{8 * l_1 + 1} - 1)$$

式中: l_1 为频次为 1 的关键词数量; T 为高低频词分界值。

由此可以得到高低频词分界值约为 133, 对提取出的高频关键词按照关联强度排序得到表 1。从该表中可以看出, 词频与关联强度最高的词是“作用机理”, 其次是“中医药”、“网络药理学”、“疫情防控”等, 与新冠疫情下国民的关注点非常吻合。此外, 高频关键词中出现的“现实群体”、“心理健康”等字眼, 提醒我们在面对突发公共卫生事件, 要全面关注社会中的不同群体, 针对产生心理应激反应的群体推出更多更精准的心理干预, 这对于维护社会稳定、全面防控疫情也是至关重要的[15]。

文献关键词共现网络图如图 2 所示, 其中元素的大小取决于节点的度、连线的强度等, 元素的颜色代表其所属的聚类, 不同的聚类用不同的颜色表示, 通过该图可以发现研究热点的结构分布和关联情况[16]。例如, 与“疫情防控”关键词相关联的有“护理管理”、“应急管理”、“青少年”、“隔离”、“输入”等关键词, 而这些关键词中又以“应急管理”的关联程度最高, 由此可以推断出与“疫情防控”

40



Figure 2. Keyword co occurrence network diagram of COVID-19 related literature
图 2. 新冠相关文献关键词共现网络图

文献关键词共现热力图如图 3 所示, 图谱上每一点都会根据该点周围元素的密度来填充颜色, 密度越大, 越接近红色, 密度大小依赖于周围区域元素的数量以及这些元素的重要性, 从图 3 中可以看出, 除了表 1 中列出的 10 个关键词, “医护人员”、“流行特征”、“数据挖掘”等关键词的重要性也比较高, 反映出新冠疫情波及范围之广、牵涉领域之多。

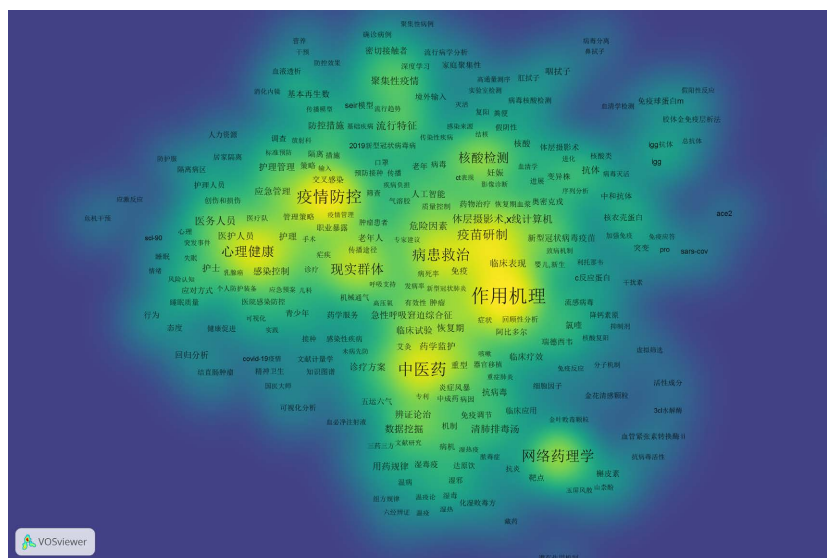


Figure 3. Keyword co occurrence thermodynamic diagram of COVID-19 related literature

图 3. 新冠相关文献关键词共现热力图

Table 1. Top 10 high-frequency keyword ranking of COVID-19 related literature
表 1. 新冠相关文献高频关键词排序 Top 10

序号	关键词	词频	关联强度	序号	关键词	词频	关联强度
1	作用机理	435	735	6	现实群体	152	265
2	中医药	298	449	7	临床特征	196	247
3	网络药理学	222	413	8	疫苗研制	137	199
4	疫情防控	303	372	9	心理健康	170	194
5	病患救治	223	335	10	核酸检测	147	149

3.3. 学科主题及类别分析

文章的学科主题指的是该文章所属的领域，根据《中国图书馆分类法》，用 A-Z 中的 22 个字母代表 22 个大类，表示军事、经济、医药卫生等领域[17]；学科类别则代表对应的研究方向，是学科主题内部更为细致的划分。通过对文献涉及的学科主题和类别进行统计分析，获得关联强度(此处的关联强度用每个学科主题或学科类别下的文献数量来衡量)前 10 的学科主题和学科类别如表 2、表 3 所示。从表 2 中可以看出，新冠相关文献在医药、卫生领域的关联强度最高，高达 8031，其次就是经济、工业技术等领域，与现实情况吻合。从表 3 中可以看出，在学科类别方面，呼吸系统疾病这一学科的关联强度最大，达到 3108，之后是预防医学与卫生学、中医学等相关学科，通过观察可以发现这些学科类别都是与医药卫生相关的学科，反映出新冠疫情对国内医药卫生领域的影响之大。

Table 2. TOP10 subjects in major disciplines of COVID-19 related literature
表 2. 新冠相关文献主要学科主题分布 Top 10

序号	学科主题	关联强度	序号	学科主题	关联强度
1	医药、卫生	8031	6	社会科学总论	69
2	经济	176	7	哲学、宗教	53
3	工业技术	166	8	生物科学	28
4	文化、科学、教育、体育	161	9	政治、法律	18
5	数理科学和化学	74	10	环境科学、安全科学	13

Table 3. Top 10 major discipline categories of COVID-19 related literature
表 3. 新冠相关文献主要学科类别分布 Top 10

序号	学科类别	关联强度	序号	学科类别	关联强度
1	呼吸系统疾病	3108	6	基础医学	491
2	预防医学与卫生学	1173	7	中药学	326
3	中医学	944	8	生物学	260
4	医药卫生方针政策与法律法规研究	731	9	药理学	257
5	临床医学	659	10	儿科学	238

根据时间顺序对每个月关联强度前 5 的学科类别进行统计，如图 4 所示，该图反映了不同时间段国内民众和学者对不同学科关注度的变化。从图中可以发现，呼吸系统疾病这一学科每个月的关注度都比

较高, 其余学科的关注度则存在些许变动, 但总体来看, 人们重点关注的学科始终围绕在医药卫生方面。

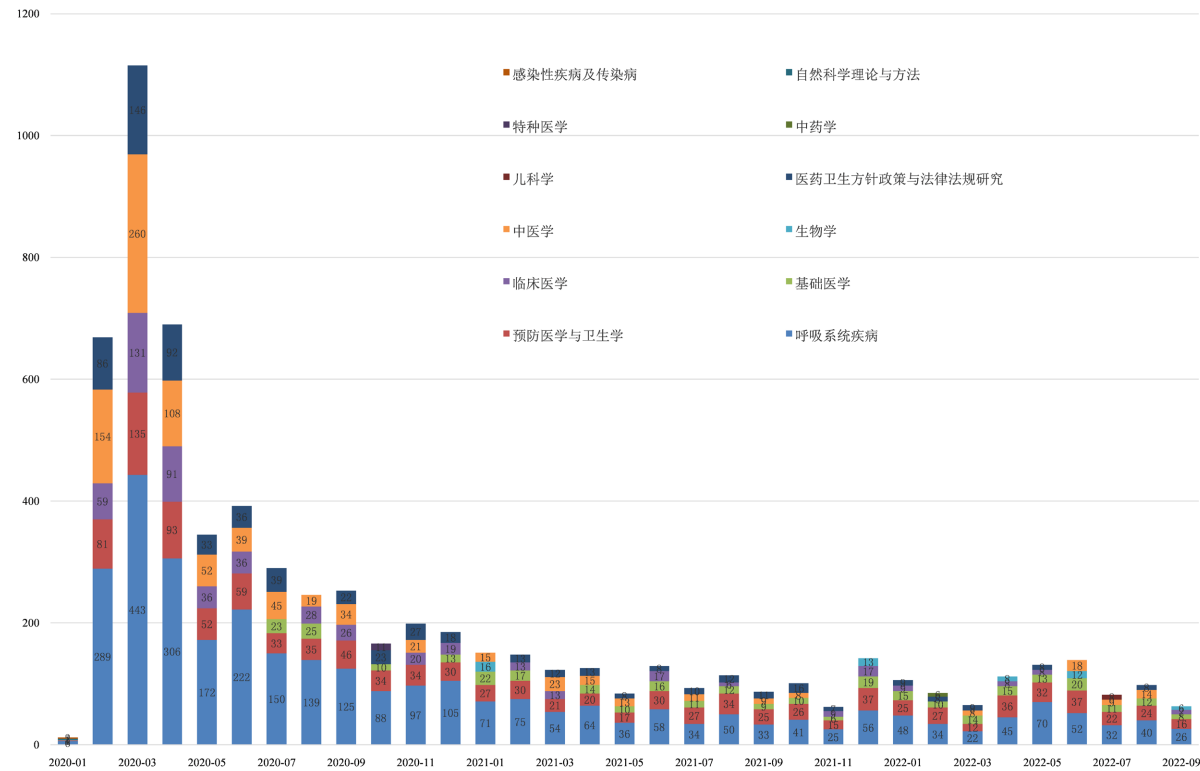


Figure 4. Monthly attention distribution of subject category of COVID-19 related literature
图 4. 新冠相关文献学科类别月关注度分布图

3.4. 刊物数量及发文机构分析

刊物数量分析能获得不同刊物刊登的文献数据, 可以从侧面衡量疫情环境下该刊物的影响力。发文机构分析能获得发文量较多的机构排名, 可用来衡量各机构在新冠疫情方面的学术和科研成就高低。使用 BiComb 软件统计刊文数前 10 的刊物和发文量前 10 的机构, 结果见表 4、表 5。从表 4 中可以看出, 刊文数前 10 的刊物都是医药卫生领域的, 值得注意的是, 武汉大学学报(医学版)是唯一一个在榜的高校学报刊物。在发文机构中, 发文量最多的是华中科技大学和武汉大学, 一方面是地理位置原因, 另一方面也反映出其学术科研水平较高。统计每篇文献的发文机构数量, 发现单机构独立发文 2637 篇, 占比 32.81%, 多机构联合发文 4361 篇, 占比 54.25%, 表明疫情期间国内各机构间的合作非常频繁。

Table 4. Top 10 sources of major journals of COVID-19 related literature
表 4. 新冠相关文献主要期刊来源分布 Top 10

序号	刊物名称	刊文数/篇	占比/%	序号	刊物名称	刊文数/篇	占比/%
1	中华流行病学杂志	179	2.23	6	病毒学报	104	1.29
2	中华医学信息导报	177	2.20	7	中华预防医学杂志	97	1.21
3	中华医院感染学杂志	164	2.04	8	广东医学	92	1.14
4	武汉大学学报(医学版)	138	1.72	9	世界科学技术-中医药现代化	91	1.13
5	中华危重病急救医学	118	1.47	10	中华传染病杂志	90	1.12

Table 5. Top 10 major publishing institutions of COVID-19 related literature
表 5. 新冠相关文献主要发文机构分布 Top 10

序号	发文机构	发文量/篇	占比/%	序号	发文机构	发文量/篇	占比/%
1	华中科技大学	958	11.92	6	四川大学	408	5.08
2	武汉大学	640	7.96	7	浙江大学	300	3.73
3	首都医科大学	476	5.92	8	中南大学	254	3.16
4	北京大学	476	5.92	9	上海交通大学	252	3.13
5	复旦大学	421	5.24	10	北京中医药大学	238	2.96

3.5. 刊物数量及发文机构分析

对文献的检索类型进行统计分析，得到高水平论文的产出情况，如表 6 所示。由该表可以看出，新冠相关文献中核心论文共有 4018 篇，约占总文献数的 50%，SCI 论文共有 186 篇，占核心文献数的 4.6%。对每个月不同检索类型的文献数量进行统计，如图 5 所示，从图中可以看出 2020 年 3 月产出的核心论文数最多，有 600 多篇，2020 年 5 月的 SCI 论文数最多，有 39 篇，2020 年 2、4 月的核心论文数在 400 多篇，2020 年 5、6、7 三个月的核心论文数都在 200 篇左右，其余月份的核心论文数在 100 篇左右。

Table 6. Distribution table of each retrieval type of COVID-19 related literature
表 6. 新冠相关文献各检索类型分布表

检索类型	文献数/篇	占比/%	检索类型	文献数/篇	占比/%
北大核心	4018	49.99	SCI	186	2.31
CSCD	3849	47.89	EI	91	1.13
CSSCI	193	2.40			

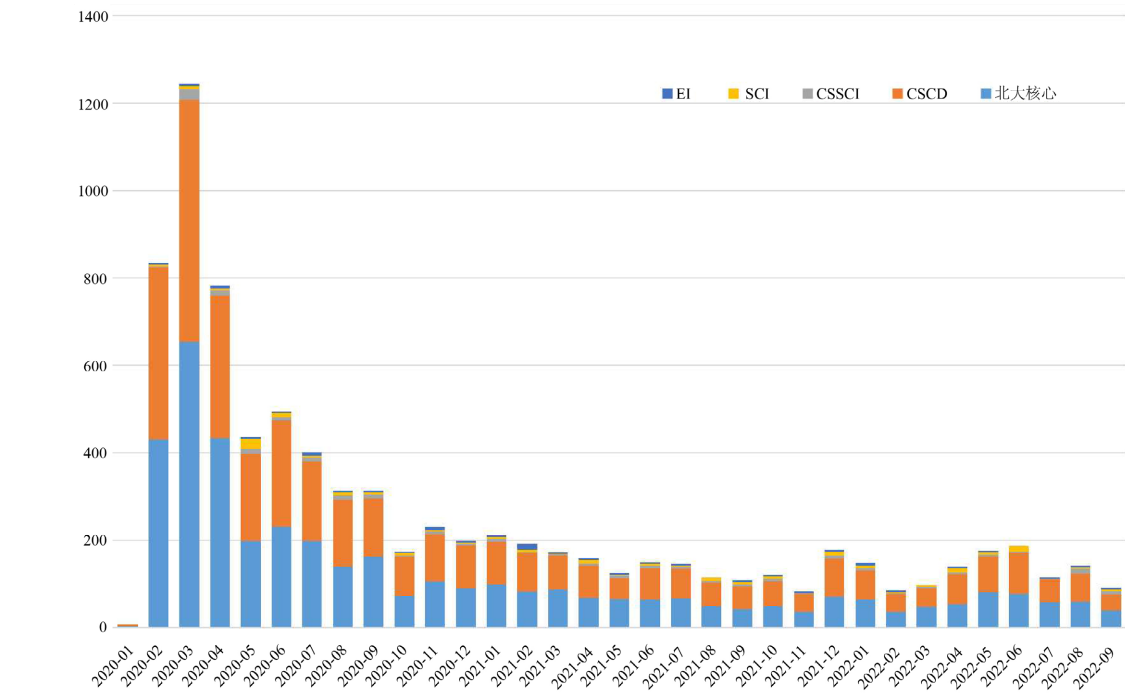


Figure 5. Monthly distribution map of papers by search type of COVID-19 related literature
图 5. 新冠相关文献各检索类型论文月分布图

3.6. 刊物数量及发文机构分析

对新冠相关文献的资助基金进行分析,发现总计 8038 篇文献中,有 6089 篇(占比 75.75%)文献有基金资助,其中 4973 篇(占比 61.87%)文献获得了多项基金资助,而剩余的 1949 篇(占比 24.81%)文献未有基金资助,如表 7 所示。经详细统计,所有文献共有资助基金 10249 项,这其中资助频次较高的几种基金如表 8 所示,可以看出国家自然科学基金的资助频次最高(971 次),国家重点研发计划的资助频次紧随其后(353 次),后面依次是省市自然科学基金(289 次)、国家科技重大专项(144 次)等。这些数据表明国家和政府在新冠疫情的防控和研究方面投入了大量的财力物力,很直观地反映出国家对新冠疫情的重视程度。

Table 7. Schedule of fund subsidies of COVID-19 related literature

表 7. 新冠相关文献基金资助情况一览表

文章是否有基金资助				数量/篇	占比/%
资助数量		数量/篇	占比/%		
有基金资助	单项基金资助	1116	18.33	6089	75.75
	多项基金资助	4973	81.67		
	无基金资助			1949	24.25

Table 8. Distribution of sub-vented funds by category of COVID-19 related literature

表 8. 新冠相关文献资助基金类别分布表

基金类别	资助频次/篇	占比/%	基金类别	资助频次/篇	占比/%
国家自然科学基金	971	9.48	国家科技重大专项	144	1.41
国家重点研发计划	353	3.44	国家社会科学基金	55	0.54
省市自然科学基金	289	2.82	其它类型资助基金	8437	82.32

4. 研究总结与思考

在围绕新冠疫情开展的学术科研环境下,本文从文献分析这一角度入手,对中国知网新冠知识平台的中文文献数据进行分析,从新冠疫情相关文献发文量、关键词共现、学科主题及类别、刊物数量及发文机构、检索类型、资助基金等多方面展开了相关内容的撰写,使用数据统计、共现分析、可视化展示等方式将此次疫情带来的各方面影响较为全面的展示在各位读者面前。与此同时,国内许多专家学者也开展了相关研究,相比而言,本研究的内容虽然只局限于国内新冠疫情研究现状,但文献时间跨度长,覆盖面广,具有针对性,所以最终结果和相关结论也有较大的参考价值。

通过本研究的数据挖掘和分析,总结了新冠相关文献研究的规律和走势,可以了解到新冠疫情环境下国内民众和学者对新冠研究的热点和关注度变化,对目前国内缺少相关研究的现状起了一个补充的作用,也使得新冠疫情研究这个框架变得更加丰富。透过相关数据和图表,也有些地方值得我们关注,例如由关键词共现网络图能发现,与重大公共卫生事件关联程度高的关键词有心理健康、现实群体等,因此在出现重大公共卫生事件时合理妥善的安置不同群体并给予相应的心理干预是非常必要的,笔者在这里未对其加以论述,但在后续的研究中会进行深入探讨。同时在疫情研究的学术科研框架下,文献计量学领域的论文数量较少,所以期待国内有更多的学者能在该领域内发表更新更有深度的学术成果。由于对新冠肺炎的相关研究不断更新,研究结果可能是初步的、不全面的,但通过本研究取得的一些结论仍

旧可以在一定程度上反映出新冠研究热点及领域发展情况, 可为此话题的相关领域提供参考。

基金项目

国家自然科学基金资助项目(11701434); 武汉工程大学研究生创新基金(CX2021386)。

参考文献

- [1] 陈莹, 郭怡博, 郭然, 等. 基于文献计量学的新型冠状病毒肺炎(COVID-19)研究可视化分析[J]. 中国中药杂志, 2020, 45(10): 2239-2248.
- [2] 张荣娜, 纪瀚然, 梁作如, 等. 新型冠状病毒相关研究的文献计量学分析——基于 Web of Science 核心合集[J]. 疾病监测, 2022, 37(6): 734-739.
- [3] 陈锐锋, 谭春林. 大数据视域下知网首发的新冠肺炎专题中文论文文献计量学分析[J]. 科技传播, 2020, 12(19): 10-14.
- [4] 郭婷, 张思琪. 2019 年 12 月~2020 年 3 月新型冠状病毒肺炎主题研究的文献计量学分析[J]. 医学信息, 2020, 33(12): 1-4.
- [5] 陈玉恒, 刘敬伟, 龙志强, 赵平. 基于 CNKI 和 PubMed 数据库的新型冠状病毒肺炎相关研究的文献计量学分析[J]. 中国社会医学杂志, 2020, 37(2): 115-118.
- [6] 李逸雯, 刘玥, 吕诚, 等. 新型冠状病毒肺炎的中医药临床研究: 基于文献计量学分析的解读与思考[J]. 科学通报, 2021, 66(26): 3385-3394.
- [7] 彭文雅, 肖祖林, 吴驻林, 彭立生. 中医药防治新型冠状病毒肺炎研究的文献计量学分析[J]. 中国中医药现代远程教育, 2020, 18(5): 221-223.
- [8] 张生太, 宣雅迪, 仇沛毅, 刘娜. 基于社会网络分析和文献计量学的新冠肺炎学术研究现状与特点研究[J]. 北京邮电大学学报(社会科学版), 2020, 22(5): 87-98, 124.
- [9] 孙佳琦, 梁会营. 药物重定位研究热点的文献计量学分析[J]. 中国现代医生, 2022, 60(19): 71-76.
- [10] 柴国荣, 汪佳颖. 基于 CiteSpace 的卫生统计研究进展及趋势分析[J]. 统计与决策, 2022, 38(18): 64-68.
- [11] 高风, 郭丽君, 张宏伟, 王亦菲, 任高灿, 马晓昌. 基于 VOSviewer 和 CiteSpace 中医药领域细胞焦亡的知识图谱分析[J/OL]. 中国中药杂志: 1-12.
<https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CAPI&dbname=CAPILAST&filename=ZGZY20221013002>, 2022-10-18.
- [12] 尚建国, 鲍茹君. 国内三语习得研究热点与发展趋势——基于 Bicom 的可视化分析[J]. 英语广场, 2022(23): 36-39.
- [13] 尹相旭, 张更平, 李晓菲. 基于关键词统计的情报学研究现状分析[J]. 情报杂志, 2009, 28(11): 38-41.
- [14] Donohue, J.C. (1973) Understanding Scientific Literature: A Bib-Liographic Approach. The MIT Press, Cambridge.
- [15] 詹昱新, 李素云, 刘义兰, 等. 综合医院应对新型冠状病毒肺炎疫情的护理应急管理[J]. 护理研究, 2020, 34(5): 744-748.
- [16] Ranjbari, M., et al. (2022) Waste Management beyond the COVID-19 Pandemic: Bibliometric and Text Mining Analyses. *Gondwana Research*. <https://doi.org/10.1016/j.gr.2021.12.015>
- [17] 中国图书馆分类法编辑委员会. 中国图书馆分类法[M]. 第四版. 北京: 北京图书馆出版社, 1999: 1.