

# 文献计量视角下中国基于 NLP 的大模型研究综述



张楠, 卢居辉\*

鹏城实验室, 广东深圳 518000

**摘要:** 【目的】随着 ChatGPT 的发布及应用, 大模型训练相关技术开启了新一轮的研究热潮, 而这一次则被称为最接近 AI 实现的人工智能。通过对基于 NLP 的大模型研究现状进行全面的分析, 可以使研究者们对目前中国该领域的研究有一个全面的了解。【方法】通过使用文献计量和内容分析等方法, 对选取期刊论文的外部特征和内容进行分析, 通过对各年度发文、期刊被引量、核心作者、核心机构、研究关键词、研究热点等维度进行分析, 对中国大模型领域研究有了一个较为全面的认知, 同时也总结了一些中国该领域研究的特点, 也对接下来的研究方向做了相应的预测。【结果】中国对大模型相关的研究已经深入到社会的多个领域, 目前正处于蓬勃发展阶段。【结论】目前中国该领域研究正处于科学文献增长的第二阶段, 且随着大模型的基础理论不断完善, 以及相关应用的落地, 接下来基于 NLP 的大模型研究将会进入一个高速发展时期。

**关键词:** NLP; 大模型; 文献计量; 内容分析

**DOI:** [10.57237/j.cst.2023.04.002](https://doi.org/10.57237/j.cst.2023.04.002)

## A Review of Research Based on NLP Foundation Models in China from the Perspective of Bibliometrics

Zhang Nan, Lu Juhui\*

Peng Cheng Laboratory, Shenzhen 518000, China

**Abstract:** [Objective] With the release and application of ChatGPT, large model training technologies have started a new round of research boom, and this time is called the closest to AI AI. Through a comprehensive analysis of the current situation of the large model research based on NLP, the researchers can have a comprehensive understanding of the current research in this field in China. [Methods] Through the use of bibliometric and content analysis, analyzes the external characteristics and content of journal papers, through the annual post, journal cited, the core authors, core institutions, research keywords, research dimensions, China big model research has a more comprehensive cognition, but also summarizes some of the characteristics of the research in the field of China, also for the following research direction.[Results] China's research on large models has penetrated into many fields of society and is currently in the stage of vigorous development. [Conclusion] At present, the research in this field in China is in the second stage of the growth of scientific literature. With the continuous improvement of the basic theory of large model and the implementation of related applications, the research of large model based on NLP will enter a period of rapid development.

**Keywords:** NLP; Foundation Models; Bibliometric; Content Analysis

\*通信作者: 卢居辉, [brucelujhui@163.com](mailto:brucelujhui@163.com)

## 1 引言

2022 年 11 月 30 日 OpenAI 对外发布了 ChatGPT 聊天机器人程序[1]，颠覆了传统搜索引擎的使用方式，以一种全新的人机对话的方式迅速占领用户，也拉开了自然语言处理大模型大战的序幕。2023 年 5 月 28 日，在 2023 中关村论坛上《中国人工智能大模型地图研究报告》发布[2]，该报告指出中国大模型正呈现蓬勃发展态势，在领域分布上，自然语言处理仍是目前大模型研发最为活跃的重点领域。2023 年 7 月 13 日国家网信办联合国家发展改革委、教育部、科技部、工业和信息化部、公安部、广电总局七部联合公布了《生成式人工智能服务管理暂行办法》[3]，弥补了 NLP 大模型法律监管上的不足之处。这一系列的动作无不证明中国对大模型相关领域研究的重视，以及对该领域研究成果所产生价值的认可。

因此，本文从所选期刊的外部特征和文献内容两个角度出发，对所选期刊的年代发文、文献被引量、核心期刊、核心机构、核心作者群等维度，通过借助 CiteSpace、Excel 等分析工具对上述罗列维度进行详细的分析，并基于分析的结果来明确下面几个疑惑：当前中国该领域的研究到底处于科学文献增长的逻辑曲线[4]的哪个阶段？该领域期刊在年代发文及被引率上呈现哪些特点？该领域核心作者群、核心机构群之间合作关系如何？该领域这些年来的研究热点具体有哪些？接下来该领域的热点研究方向有哪些？

## 2 数据来源

本文所有期刊数据均选用中国知网数据库，出于对选文的权威性及高可靠性考虑，数据库来源选择 SCI、EI、北大核心、CSSCI、CSCD、AMI，采用高级检索方式，确定检索式为：“（主题=NLP+自然语言处理+大模型）OR（篇名=NLP+自然语言处理+大模型）OR（关键词=NLP+自然语言处理+大模型）OR（篇名=NLP+自然语言处理+大模型）”（精确查找）。将搜索时间设为 2000 年—至今（截至 2023 年 7 月 29 日，下同）检索共得到 4756 条结果，通过排除与分析主题无

关等结果之后，最终得到 4641 条可用数据。

## 3 大模型研究期刊年度发文及作者特征分析

通过分析选文期刊的各年度发文、期刊被引量、核心作者群、核心机构等维度进行分析，有助于我们从宏观上把握中国基于 NLP 的大模型研究领域（简称“大模型研究”，下同）的发展进程与现状。

### 3.1 年度发文特征分析

#### 3.1.1 各年度期刊累积发文分析

通过对各年度的期刊累积发文分析可以探索出该领域的研究到底处于哪个研究阶段，以及各年度的研究热度如何，因此本文借助 Excel 工具，对选取的 4641 篇期刊论文进行年度发文分析（见图 1）。

通过观察发现，2000-2003 年时间段，该领域各年度的发文量均呈现出较为明显的下降趋势，4 年累积发文量为 924 篇，占选取总期刊发文量的 19.91%，结合科学文献增长的逻辑曲线[4]进行对比分析，可知此时该领域研究正处于科学研究的起步阶段；2004-2016 年时间段，该领域整体发文量为 2286 篇，占选取总期刊发文量的 49.25%，但其中有几个比较明显的区间段，即 2005-2008 时间段累积发文量处于上升阶段，2009-2011 整体处于下降阶段，2012 年短暂爆发之后，2013-2016 区间段一直处于下降趋势。这一区间段为何会出现这种不稳定的发文现象？本文将在下文结合关键词时序图探析其原因；2017-2022 区间段，整体发文累积量为 1261 篇，占总体发文量的 27.17%，此时间段发文整体呈上升趋势，表明大模型研究相关的课题开始稳定增加，而且截至 2023 年 7 月 29 日，2023 年的累积发文量已达 170 篇，根据这一趋势，笔者预测，该领域 2023 年全年期刊发文量将会超过 239 篇（2023 年全年期刊发文量的增长趋势如图 1 中虚线部分所示）。综上分析可以得出，当前大模型领域研究正处于蓬勃发展阶段。

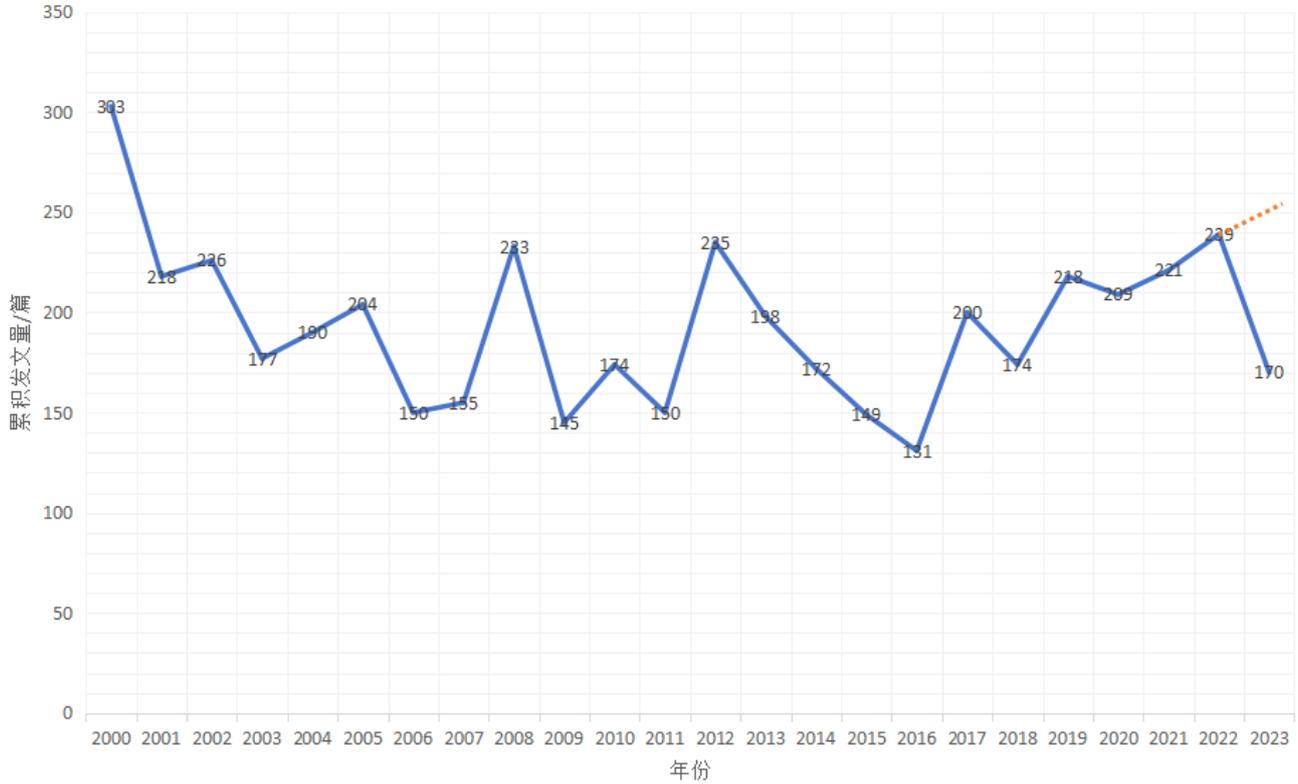


图 1 2000-2023 各年度累积发文章量

Figure 1 Cumulative publication of papers over the years of 2000-2023

### 3.1.2 选文期刊被引量分析

通过分析各年度期刊被引量，可以找出影响该领域研究的发文集中在哪一年或者哪一时间段，这样更

容易捕捉到该领域研究的重要发展节点。因此，本文对 2000-2023 年各年度的期刊文献被引量做了详细的统计，详见表 1。

表 1 2000-2023 年各年度累积期刊发文被引量

Table 1 Cumulative citation number of journal articles in 2000-2023

年份	累积发文章量/篇	被引频次>0 次发文总量/篇	被引频次/次	被引频次总占比/%
2000	303	37	857	0.90
2001	218	43	1478	1.55
2002	226	58	2057	2.16
2003	177	57	3437	3.61
2004	190	92	4697	4.94
2005	204	117	4324	4.55
2006	150	92	4985	5.24
2007	155	130	3851	4.05
2008	233	125	2613	2.75
2009	145	130	2174	2.29
2010	174	139	2955	3.11
2011	150	148	3660	3.85
2012	235	127	1802	1.89
2013	198	164	5258	5.53
2014	172	148	2844	2.99
2015	149	136	2942	3.09
2016	131	116	8172	8.59
2017	200	194	11364	11.95

年份	累积发文量/篇	被引频次>0 次发文总量/篇	被引频次/次	被引频次总占比/%
2018	174	158	6815	7.17
2019	218	194	7187	7.56
2020	209	177	5980	6.29
2021	221	151	4566	4.80
2022	239	137	784	0.82
2023	170	61	291	0.31
总计	4641	2931	95093	100.00

由图 2 可以看出：研究者在 2000-2006 的 7 年间，发表该领域的相关研究文献为 1623 篇，占总发文量的 34.97%，但累计被引次数仅 25686 次，占总选文期刊被引量的 27.01%，即 7 年间研究者在该领域的发文总量占整体发文量的三分之一，但被引率却不高。造成该现象的原因可能有研究者们对该领域的研究还处于起步探索阶段，在一些重要的基础概念上并没有达成统一共识；2007-2015 年，此时间段累积期刊发文 1611 篇，占总发文量的 34.71%，累计被引次数 28099 次，占总选文期刊被引量的 29.54%，进一步分析可以发现 2013 年的累积被引量虽然上涨，但是该时间段整体被引率处于负增长的趋势，直到 2016-2017 年“深度学习”、“卷积神经网络”等概念的提出，总被引频次又开始了新的增长；2016-2017 年间虽然累积发文量只有 331 篇，仅占总发文量的 7.13%，但总被引频数高达

19536 次，占总选文期刊被引量的 20.54%，证明这两年的发文在基础理论研究上有了重大突破，众多学者对这一时间段的发文有极大的认同感，同时也表明这一阶段的研究成果对整个大模型领域的理论研究起到了举足轻重的作用；2018-2023 年，此时间段累积期刊发文 1231 篇，占总发文量的 26.52%，累计被引次数 25623 次，占总选文期刊被引量的 26.95%，该时间段被引率也是处于下降趋势，这可能是由于 2016-2017 年提出的理论依然能指引此后数年的研究导致的。

后来研究者们可以通过研读该领域的高被引论文来快速了解到该领域已经达成共识的一些基础理论。因此，本文将 2000-2023 年各年度的高被引论文进行了整理，篇幅所限，只将被引频次在 500 次以上的论文列举出来，如表 2。

表 2 2000-2023 年高被引期刊论文列表

Table 2 List of highly cited journal papers 2000-2023

序号	篇名	作者	刊名	发表时间	被引频次/次
1	卷积神经网络研究综述	周飞燕；金林鹏；董军	计算机学报	2017	5063
2	深度学习的昨天、今天和明天	余凯；贾磊；陈雨强；徐伟	计算机研究与发展	2013	2011
3	知识图谱构建技术综述	刘峤；李杨；段宏，等	计算机研究与发展	2016	1842
4	基于 HowNet 的词汇语义倾向计算	朱嫣岚；闵锦；周雅倩，等	中文信息学报	2006	1303
5	神经网络七十年：回顾与展望	焦李成；杨淑媛；刘芳，等	计算机学报	2016	1261
6	本体论研究综述	李善平；尹奇韡；胡玉杰，等	计算机研究与发展	2004	974
7	信息抽取研究综述	李保利；陈玉忠；俞士汶	计算机工程与应用	2003	800
8	自然语言处理中主题模型的发展	徐戈；王厚峰	计算机学报	2011	729
9	深度卷积神经网络的发展及其在计算机视觉领域的应用	张顺；龚怡宏；王进军	计算机学报	2017	693
10	一种结合词项语义信息和 TF-IDF 方法的文本相似度量方法	黄承慧；印鉴；侯昉	计算机学报	2011	652
11	面向自然语言处理的深度学习研究	奚雪峰；周国栋	自动化学报	2016	554
12	图卷积神经网络综述	徐冰冰；岑科廷；黄俊杰，等	计算机学报	2019	542
13	教育人工智能（EAI）的内涵、关键技术与应用趋势——美国《为人工智能的未来做好准备》和《国家人工智能研发战略规划》报告解析	闫志明；唐夏夏；秦旋，等	远程教育杂志	2017	537
14	话题检测与跟踪的评测及研究综述	洪宇；张宇；刘挺，等	中文信息学报	2007	514

通过对表 2 的分析不难发现：在这 14 篇高被引论文中仅发表在 2016-2017 年间的就有 6 篇，而 2008-2010 年间则一篇也没有。这一现象证明

2016-2017 这两年的论文对大模型研究领域起到了理论奠基作用。分析发文年份也可以进一步印证上文对各时间段发文情况的分析。

### 3.1.3 核心期刊及其发文研究

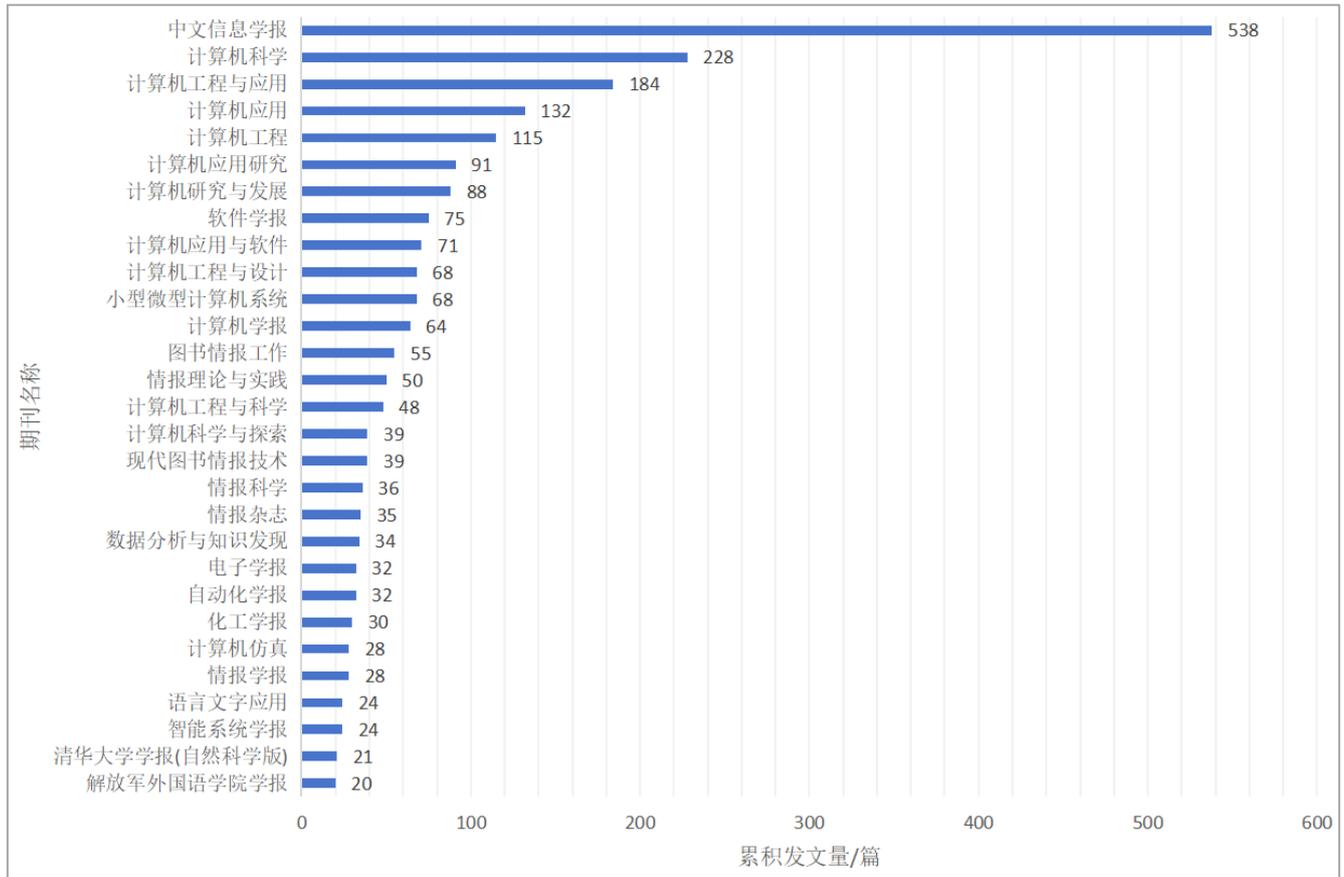


图 2 2000-2023 年度期刊文献论文研究重要期刊分布

Figure 2 Research on the distribution of important journals in 2000-2023

为了方便后来研究者们可以快速定位大模型研究领域的期刊资料来源，本文对刊载这 4641 篇论文的期刊进行了统计分析。同时，依据布拉德福定律分析出了该领域的核心期刊群[5]。图 2 列举出了大模型研究领域累积发文量在 20 篇及以上的期刊名称及其累积发文量。

此外，结合布拉德福的定律[6]分析发现，所选该领域的期刊载文量基本符合相对集中又高度分散的特征。

## 3.2 期刊文献作者相关分析

### 3.2.1 核心作者及其机构分析

通过使用 CiteSpace 工具提取所选期刊各年度发文

累积量前 10%的作者，由此得到高产作者知识图谱(如图 3 所示)。在符合条件的 189 位作者中，其相互之间代表有合作关系的连线只有 76 条，这表明该领域作者之间的合作关系并不密切。

通过使用德里克·普赖斯公式(1) [7]，可以计算出大模型研究领域的核心作者数量。

$$M = 0.749 \times \sqrt{N_{\max}} \quad (1)$$

有(1)式计算可得，M 值约为 4.5，由此可以统计出 218 位符合条件的高产作者。通过对高产作者进行分析，可以使研究者们找到该领域的专家。本文将发文在 20 篇及以上的高产作者定位核心作者，如表 3。



图 3 高产作者知识图谱

Figure 3 Knowledge map of productive authors

表 3 2000-2023 年高产作者列表

Table 3 List of prolific authors in 2000-2023

作者	作者机构	累积发文量/篇
朱巧明	苏州大学计算机科学与技术学院	37
周国栋	苏州大学计算机科学与技术学院	37
俞士汶	北京大学计算语言学教育部重点实验室	28
刘挺	哈尔滨工业大学计算机学院信息检索研究中心	26
咎红英	郑州大学信息工程学院	23
李培峰	苏州大学计算机科学与技术学院	23
林鸿飞	大连理工大学计算机科学与技术学院	22
李生	哈尔滨工业大学计算机学院信息检索研究室	22
孙茂松	清华大学智能技术与系统国家重点实验室	21
李寿山	苏州大学计算机科学与技术学院	21
姬东鸿	武汉大学计算机学院	20

分析表 3 可以发现，核心及高产作者的累积发文量为 1857 篇，占所选期刊总发文量的 40.01%，这表明大模型研究领域核心研究圈子已经形成，该领域的主要研究将在这些核心及高产作者的带领下不断向前探索发展。

通过对选文机构进行分析可以找出当前中国该领域研究机构之间的合作关系是怎样的。利用

CiteSpace 分析工具，对本文所选取的 2000-2023 年间的 4661 篇论文中每年发文量 TOP 10 的研究机构进行分析，构建出如图 4 的大模型研究机构知识图谱。图谱共有节点 137 个，连线 17 条，其网络密度为 0.0018。连线反馈了机构间合作的寡众，机构对外连线越多，表示与其合作的机构越多，反之表示与其合作的机构越少。



图 4 研究机构知识图谱

Figure 4 Knowledge atlas of research institution

结合上文,依据普赖斯定律[8]的计算以及结合图 4、表 3 可以发现,中国该领域的机构间合作力度还是非常低的,基本上处于各自为战的状态,建议核心作者所在的机构牵头,组织一些高水平的学术论坛,通过加强学术交流,积极探索课题合作上的可能性。

### 3.2.2 期刊文献作者合作分析

同样,通过对选文期刊的作者进行分析来确认中国该领域学者间的合作力度。本文通过对所选的 4641 篇期刊文献作者进行建模分析,得到如下图 5 所示。

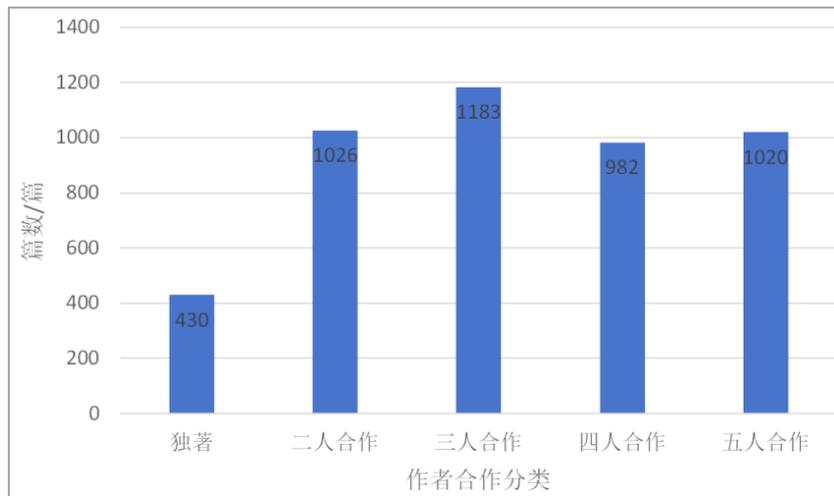


图 5 作者合作分析

Figure 5 Authors cooperation analysis

结合图 4、图 5，可以总结出，虽然中国大模型研究领域机构与机构之间的合作力度极低，但是中国该领域的研究学者之间的合作力度却很高，占比高达 90.73%。这表明该领域的研究并不局限于机构，学者之间已经形成了比较稳定的研究团体。

## 4 大模型期刊文献内容分析

通过使用词频分析法[9]，对大模型研究领域的关键词进行分析，更好的聚焦该领域近年来的研究热点都有哪些。

CiteSpace, v. 5.8.R5 (64-bit)  
 July 31, 2023 8:14:59 PM CST  
 C:\Users\myuser\Desktop  
 Timespan: 2000-2023 (Slice Length=1)  
 Selection Criteria: Top 20 per slice  
 Network: N=191, E=211 (Density=0.0116)  
 Pruning: MST

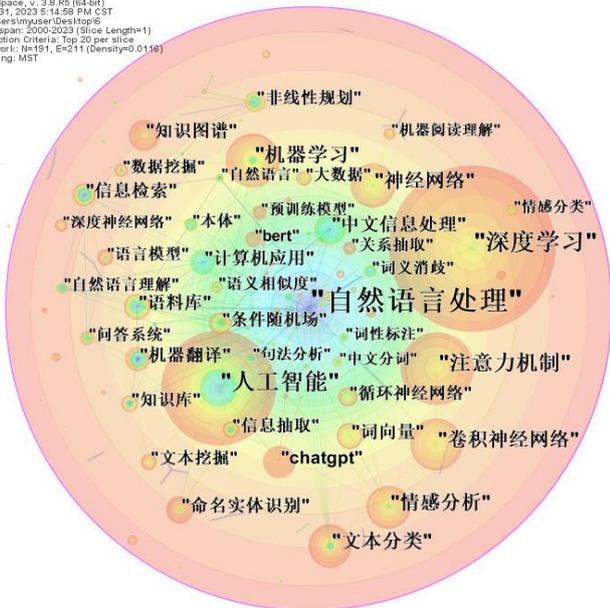


图 6 关键词共词分析

Figure 6 Keywords and Co-words Analysis

通过对所选 4641 篇期刊，按照“KeyWords”进行共词分析，取 2000-2023 时间段前 20%的关键词，得到如图 6 所示的关键词知识图谱，图谱共有 191 个节点，211 条连线，网络密度为 0.0116。节点间的连线代表着节点间的紧密度。

通过使用高频低频词界分公式(2) [10]计算，可以清晰的梳理出历年来中国该领域研究的主要热点词汇有哪些，通过分析这些热点词，可以对该领域研究的整体生态有一个较为清晰的认知。

$$T = (-1 + \sqrt{1 + 8I_1}) / 2 \quad (2)$$

根据公式(2)计算得出， $T \approx 78.91$ ，因此该领域

2000-2023 时间段的高频词汇共 10 个。但为了更加全面的反馈该领域的研究热点，本文对关键的统计选择公式为“词频  $\geq 50$ ”，具体如表 4 所示。

表 4 高频词统计分析

Table 4 High-frequency word statistical analysis

关键词	词频/次	关键词	词频/次
自然语言处理	1307	命名实体识别	72
深度学习	365	语料库	72
人工智能	260	ChatGPT	71
注意力机制	162	机器翻译	70
机器学习	134	信息检索	67
卷积神经网络	118	词向量	64
文本分类	107	计算机应用	62
神经网络	106	条件随机场	59
情感分析	104	非线性规划	58
中文信息处理	80	信息抽取	52
知识图谱	75	知识库	50

为了更好地了解 2000-2023 年时间段研究者们在该领域的研究热点和发展动态，本文还利用 Citespace 软件分析，绘制出了该研究领域关键词时序图(见图 7)，通过图 7 可以总结出以下几点：

- (1) 2000-2003 年“自然语言处理”[11, 12]、“人工智能”、“机器学习”、“中文信息处理”等热词被大量提及，后续的许多重要研究都是在此基础上展开进一步研究的；
- (2) 2004-2016 年此时间段可以发现，该领域的研究基本上是围绕前几年提出的热点词展开研究，这其中除了 2008 年“关系抽取”、“语义相似度”、“本体”，2012 年“情感分类”、“轨迹优化”等词汇开始被大量提及后，引发一波研究热潮，其他年度整体还是围绕着之前的热点词汇在进行着研究；
- (3) 2017-2018 年发表的核刊论文被引率占了整个所选的 4641 篇期刊总被引量的 20.54%，经过进一步分析发现，这两年频繁提及了“注意力机制”、“卷积神经网络”、“知识图谱”、“词向量”、“循环神经网络”、“大数据”、“深度学习”这几个热点词，结合上面的高被引论文也可以发现这阶段提出的理论研究直接奠定了接下来数年的研究方向；
- (4) 2019-2023 年研究的基本围绕在前几年提及的“神经网络”、“深度学习”等热词展开的，特别是 2023 年初横空出世的“ChatGPT”使该领域的研

究推向了一个高潮，而且这个影响还在持续扩大中。这些现象说明上述提及的热词将会在接下来的几年里成为各领域研究者的研究热点，而该领域研究也将以这些热点向周边衍生、发散并逐渐构成一个完整的生态体系。

根据上述分析得出的结论，结合该领域最新的一些研究成果，如杨春等[13]，通过全面地调研不同角度下模型量化相关技术现状，并且深入地总结归纳不同方法的优缺点，发现量化技术目前仍然存在的问题。梁斌等[14]，提出一种基于多注意力卷积神经网络的特定目标情感分析方法，该方法通过整合多种注意力

机制，克服了仅仅依赖于内容层面注意力机制的局限性。这使得模型在不依赖外部知识（如依存句法分析）的情况下，能够获取更深层次的情感特征信息，并有效地识别不同目标的情感极性。李舟军等[15]，通过对面向自然语言处理的预训练技术进行研究分析提出未来预训练技术可能重点在于如何处理常识和推理问题，如何改善生成任务，以及如何降低训练成本等。同时分析这些具有权威性、前沿性的核刊论文也可以发现，除了上述总结的研究热点外，“预训练模型”[16]、“生成式人工智能”[17]、“AIGC（人工智能生成内容）”[18]也将会成为该领域研究的热点及趋势。

CiteSpace, v. 5.8.R5 (64-bit)  
July 31, 2023 5:24:41 PM CST  
C:\Users\myuser\Desktop  
Timespan: 2000-2023 (Slice Length=1)  
Selection Criteria: Top-10 per slice  
Network: N=86, E=80 (Density=0.0219)  
Pruning: MST  
Modularity Q=0.6388  
Mean Silhouette=0.3268

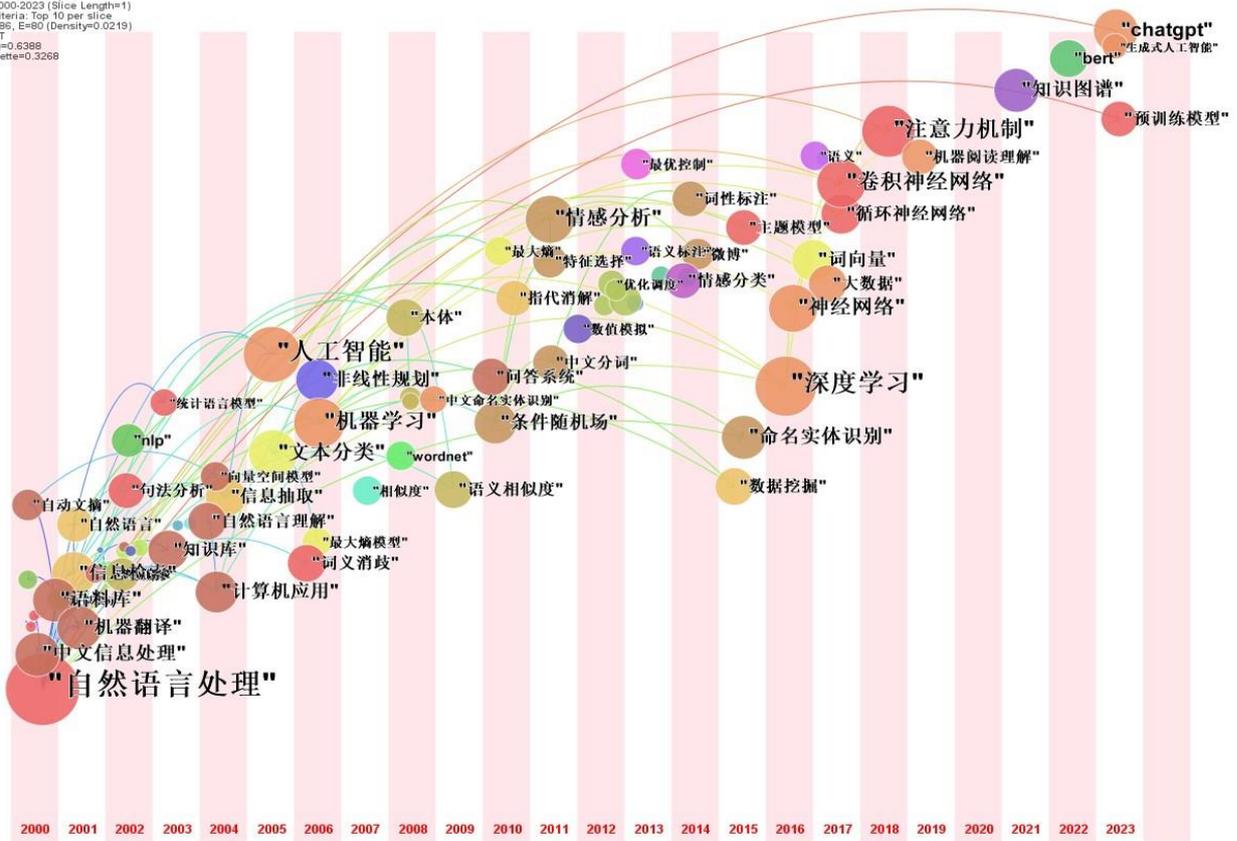


图 7 关键词时序图

Figure 7 Keywords timing diagram

### 5 总结

本文通过对所选期刊的年代发文、文献被引量、核心期刊、核心机构、核心作者群等维度进行分析发现：

(1) 大模型领域研究，在接下来的几年将处于高速发展状态，且随着时间的推移，该领域的研究

将会逐步拓展到社会的方方面面；

- (2) 通过对被引量及高被引论文分析发现，目前该领域研究的高被引论文主要集中在 2016-2017 年度，说明该时间段的发文对该领域研究有重要的推动作用，后来研究者们可以重点研究一下该时间段的论文；
- (3) 通过对大模型研究领域的发文期刊累积载文量

进行分析可以发现，目前该领域研究者们发表论文已经形成了一个以中文信息学报、计算机工程与应用、计算机工程、计算机应用等期刊为核心的核心期刊群，后来研究者通过上述核心期刊群，快速找到该领域权威且高质量的论文。

- (4) 通过对大模型研究领域作者及机构的统计分析可以发现，该领域作者之间的研究主要以团队合作为主，但是机构之间的合作却比较少。这表明接下来各机构在该领域的研究合作上还有很大的提升空间。
- (5) 通过对大模型研究领域期刊发文中的关键词进行共词分析可以发现，目前该领域研究是与时代前沿科技紧密贴合的，随着 5G 技术的迅猛发展[19]，以及“大模型”[20]概念的密集提及，接下来对基于上述热词的基础理论、应用，构建并完善该领域的整个生态体系研究势必进入新一轮的研究热潮。

## 参考文献

- [1] 蒋华林. 人工智能聊天机器人对科研成果与人才评价的影响研究——基于 ChatGPT、Microsoft Bing 视角分析 [J]. 重庆大学学报(社会科学版), 2023, 29(02): 97-110.
- [2] 中国人工智能大模型呈现蓬勃发展态势. [2023-05-30] <https://baijiahao.baidu.com/s?id=1767319212555686796&wfr=spider&for=pc>
- [3] 生成式人工智能服务管理暂行办法. [2023-07-14] [http://www.whwx.gov.cn/wlzl/zljg/202307/t20230714\\_2231690.shtml](http://www.whwx.gov.cn/wlzl/zljg/202307/t20230714_2231690.shtml)
- [4] 林辉, 林伟. 科学文献的增长规律和老化规律及其新的一般模型 [J]. 情报杂志, 2010, 29(04): 22-25.
- [5] 邱均平. 信息计量学(四)第四讲文献信息离散分布规律——布拉德福德定律 [J]. 情报理论与实践, 2000 (04): 315-314+316-320.
- [6] 何峻, 蔡蓉华. “北大中文核心期刊”是如何评出的 [J]. 中国出版, 2009 (07): 19-24.
- [7] 董国豪, 潜伟. 普赖斯与科学史定量研究 [J]. 科学学研究, 2017, 35(05): 667-675+680. DOI: 10.16192/j.cnki.1003-2053.2017.05.003.
- [8] 姚雪, 徐川平, 李杰等. 基于普赖斯定律和二八定律及在线投稿系统构建某科技期刊核心作者用户库 [J]. 编辑学报, 2017, 29(01): 64-66. DOI: 10.16811/j.cnki.1001-4314.2017.01.021.
- [9] 王米雪, 张立国. 我国智慧教育领域的研究热点与发展趋势分析——基于词频分析法、共词聚类法和多维尺度分析法 [J]. 现代教育技术, 2017, 27(03): 1-48.
- [10] 刘奕杉, 王玉琳, 李明鑫. 词频分析法中高频词阈值界定方法适用性的实证分析 [J]. 数字图书馆论坛, 2017 (09): 42-49.
- [11] 张蕾, 李学良, 刘小冬. 自然语言处理中的逻辑词 [J]. 小型微型计算机系统, 2000 (02): 42-46.
- [12] 孙宏林, 俞士汶. 浅层句法分析方法概述 [J]. 当代语言学, 2000 (02): 74-83+124.
- [13] 杨春, 张睿尧, 黄龙等. 深度神经网络模型量化方法综述 [J]. 工程科学学报, 2023, 45(10): 1613-1629. DOI: 10.13374/j.issn2095-9389.2022.12.27.004.
- [14] 梁斌, 刘全, 徐进等. 基于多注意力卷积神经网络的特定目标情感分析 [J]. 计算机研究与发展, 2017, 54(08): 1724-1735.
- [15] 李舟军, 范宇, 吴贤杰. 面向自然语言处理的预训练技术研究综述 [J]. 计算机科学, 2020, 47(03): 162-173.
- [16] 卢经纬, 郭超, 戴星原等. 问答 ChatGPT 之后: 超大预训练模型的机遇和挑战 [J]. 自动化学报, 2023, 49(04): 705-717. DOI: 10.16383/j.aas.c230107.
- [17] 戚凯. ChatGPT 与数字时代的国际竞争 [J]. 国际论坛, 2023, 25(04): 3-23+155. DOI: 10.13549/j.cnki.cn11-3959/d.2023.04.001.
- [18] 李白杨, 白云, 詹希旒等. 人工智能生成内容 (AIGC) 的技术特征与形态演进 [J]. 图书情报知识, 2023, 40(01): 66-74. DOI: 10.13366/j.dik.2023.01.066.
- [19] Adam M M, Zhao L, Wang K, et al. Beyond 5G Networks: Integration of Communication, Computing, Caching, and Control [J]. China Communications, 2023, 20(07): 137-174.
- [20] 何赛克, 张培杰, 张玮光等. 大模型时代下的决策范式转变 [J]. 中国地质大学学报(社会科学版), 2023, 23(04): 82-91. DOI: 10.16493/j.cnki.42-1627/c.20230619.001.