

# 中美国家自然科学基金大数据技术领域项目分析

姜 丹

吉林省科学技术信息研究所, 吉林 长春

收稿日期: 2022年8月7日; 录用日期: 2022年8月29日; 发布日期: 2022年9月13日

## 摘 要

为了探究大数据技术领域的总体发展态势, 通过对中美国家自然科学基金的大数据全球科研项目对比分析, 运用数量统计和文本挖掘法对2012~2021年科研立项数据进行统计分析, 揭示大数据技术领域的科研总体态势、优势学科、核心机构、热点主题等。结果表明: 近10年大数据技术领域的科研项目自2018年后呈相对缓慢下降态势, 以美国自然科学基金的计算机与信息科学及工程学部、中国自然科学基金的信息科学部为主, 核心力量是中国清华大学和美国卡内基梅隆大学, 以研究项目为主导。我国科研投入主要分布在北京、上海、江苏, 重点研发方向以大数据、数据挖掘、深度学习为主。

## 关键词

大数据, 科研项目, 自然科学基金, 研究前沿

# Analysis of Big Data Technology Projects of China-US Household Science Fund

Dan Jiang

Jilin Provincial Information Institute of Science and Technology, Changchun Jilin

Received: Aug. 7<sup>th</sup>, 2022; accepted: Aug. 29<sup>th</sup>, 2022; published: Sep. 13<sup>th</sup>, 2022

## Abstract

To explore the overall development trend of big data technology, through the comparative analysis of the big data global scientific research projects of the China-US National Science Fund, the statistical analysis of the 2012~2021 scientific research project data is carried out by using the quantitative statistics and text mining methods to reveal the overall scientific research situation, advantageous disciplines, core institutions, hot topics, etc. in the field of big data technology. The results show that the scientific research projects in the field of big data technology in the past 10 years have shown a relatively slow decline since 2018. The computer and information science and

Engineering Department of the American Science Foundation and the information science department of the China Science Foundation are the main forces. The core forces are Tsinghua University in China and Carnegie Mellon University in the United States, with research projects as the leading force. China's scientific research investment is mainly distributed in Beijing, Shanghai and Jiangsu. The key research and development directions are big data, data mining and in-depth learning.

## Keywords

Big Data, Research Projects, Self-Study Fund, Research Frontier

Copyright © 2022 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

随着大数据产业的飞速发展，大数据技术不断突破与创新，大数据是现代信息技术发展的产物，具有存储量大、数据复杂、应用空间广等特点[1]。当代大数据技术已经被广泛应用于计算机、互联网、人工智能等诸多行业，发挥着重要的社会经济价值。国家科研基金通过资助机构、科学家，大力推进了大数据技术科技创新水平。尤其在中美国家自然科学基金自设立以来，有力地推动了大数据技术在自然科学领域的研究与发展。针对国际科研项目的学科或机构竞争力分析、学科前沿布局规划、重点领域遴选、科技资源投入产出等分析，把握国际项目学科研究趋势、发现不同国家的项目态势、探究不同机构的项目动态，深入挖掘国际项目资助热点、获知各机构科研项目竞争能力，为科技决策、机构发展、学科规划等提供信息服务，从而提升科技信息服务的效率与效果。

近年来，科学计量作为专业的研究方法得到了快速发展，是科技情报人员应对大数据情景的重要工具。马海韵[2]、陈航宇[3]、伍之昂[4]等利用文献计量方法进行大数据技术或管理的研究热点分析；孙鸿飞[5]、王雪[6]、乔楠[7]、冯鑫[8]等针对国内外大数据的科技文献进行计量分析和对比分析，挖掘其前沿技术、主题演进等；麻桂新[9]、张怡宁[10]、雷水旺[11]、缪瑞生[12]等分别从农业大数据、大数据政策、图书情报领域大数据的文献计量分析，从行业领域探究大数据的研究趋势；都为大数据技术科技创新起到了积极推动作用。

当前我国大数据技术领域正全速高质量发展，为了准确把握我国大数据技术领域的科研现状和发展态势，了解科研项目动态，对推动大数据技术的科技创新发展至关重要。但缺少其国际科研项目态势研究，因此，本文采用数量统计和文本挖掘等情报分析方法，对 2012~2021 年中美国家自然科学基金大数据技术领域的科研项目进行了统计分析，以期为我国大数据产业的科研能力与水平进行客观评价，并通过定量分析探究其总体态势、优势学科、核心机构，进一步分析我国大数据技术领域的研究热点和攻关方向，为我国大数据科研机构准确把握科研选题以及科技管理部门相关科研攻关方向、政策指导等提供参考。

## 2. 数据来源与研究方法

数据来源：在科研基金的选择方面，为了尽可能全覆盖以及权威性，统计源遴选万方数据科慧平台 (Sci-Fund) 全球科研基金项目库。在大数据技术的检索策略方面，项目题名或关键词中包含大数据主题词的科研项目，专业检索式即：(项目题名：“大数据” OR “big data” OR “large data” OR “the great data” OR

“mega data” OR “big data analysis” OR “data mining” OR “large-scale data” OR “big data era”) OR (关键词: “大数据” OR “big data” OR “large data” OR “the great data” OR “mega data” OR “big data analysis” OR “data mining” OR “large-scale data” OR “big data era”) AND (资助机构团体: “中国国家自然科学基金委” OR “美国国家自然科学基金”) 立项年份: [2012 TO 2021], 检索时间 2022 年 5 月 25 日。经人工筛选和数据清洗, 最终获取项目数据集 8578 条, 其中, 美国国家自然科学基金 4543 项, 中国国家自然科学基金委 4035 项。数据集包括项目名称、关键词、摘要、承担机构等描述性字段。

研究方法: 通过分析数据集, 从资助项目的立项时间、承担机构、项目经费、学科类别、项目类型、关键词等特征属性, 综合采用了统计、计量、共现等分析方法[13], 按数据、表格、图型等多角度多维度直观揭示年度、学科、资助机构、承担机构, 为大数据技术领域的科技决策、机构发展、学科规划、科研投入提供信息服务和知识服务。

### 3. 中美国家自然基金大数据技术领域科研项目态势分析

#### 3.1. 总体态势分析

在中美国家自然科学基金大数据主题项目中, 按项目级别均为国家级 8578 项, 按项目有无产出分为中国有产出 3030 项、中国无产出 1005 项、美国无产出 4034 项, 按项目状态分为在研项目 5334 项(其中, 中国 2218 项、美国 3116 项)、结项 2732 项(其中, 中国 1305 项、美国 1427 项)、中国未知状态 512 项。从大数据技术领域科研项目年度分布(见图 1)可以看出, 大数据技术领域科研项目总体呈波动态势, 2012~2017 年处于快速上升态势, 在 2017 年立项数达到峰值 1571 项, 2018 年后出现大幅下降, 在 2020 年立项数最低 378 项, 2021 年科研项目呈现缓慢上升态势。其中, 美国自然科学基金大数据技术领域科研项目态势呈波动态势, 在 2015 年立项数达到峰值 634 项; 而我国自然科学基金大数据技术领域科研项目态势呈波动态势, 在 2017 年立项数达到峰值 981 项, 自 2018 年后出现大幅下降趋势。数据说明, 大数据技术领域 2014~2017 年处于快速成长期, 自 2018 年后出现下降趋势。

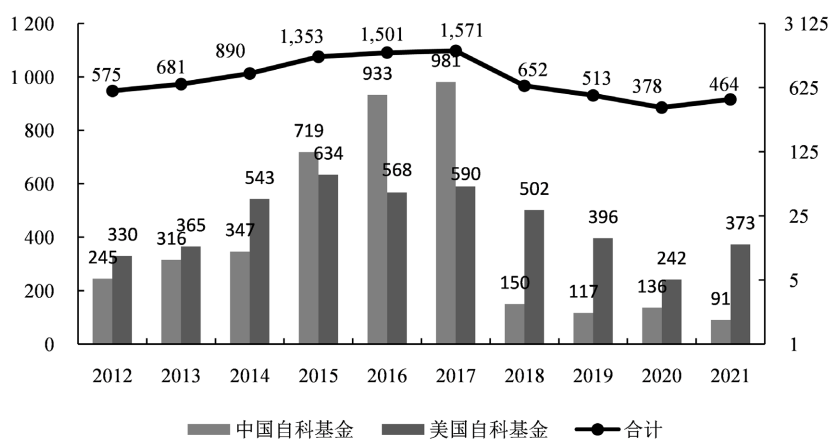


Figure 1. The figure of the trend of the number of big data research projects of China-US household science funds

图 1. 中美国家自然科学基金大数据科研项目数量态势图

#### 3.2. 优势学科分析

大数据技术领域科研项目按学科分布(见图 2), 总体排名前 3 学科分别是信息和计算机科学 2409 项、工程学 532 项、医学与生命科学 270 项, 累计占项目总数的 37.43%, 并且, 信息和计算机科学、工程学、

医学与生命科学均在 2017 年立项最多分别为 554 项、163 项、68 项。其中，中国自然科学基金的信息和计算机科学 1580 项、工程学 468 项、医学与生命科学 248 项等学科的科研项目最多，美国自然科学基金的信息和计算机科学 829 项、数学 168 项、教育 102 项等学科的科研项目最多。数据说明，中美国家自然科学基金大数据技术领域的优势学科主要侧重于信息和计算机科学领域，另外我国自然科学基金还侧重于工程学、医学与生命科学，美国自然科学基金还侧重于数学和教育学。

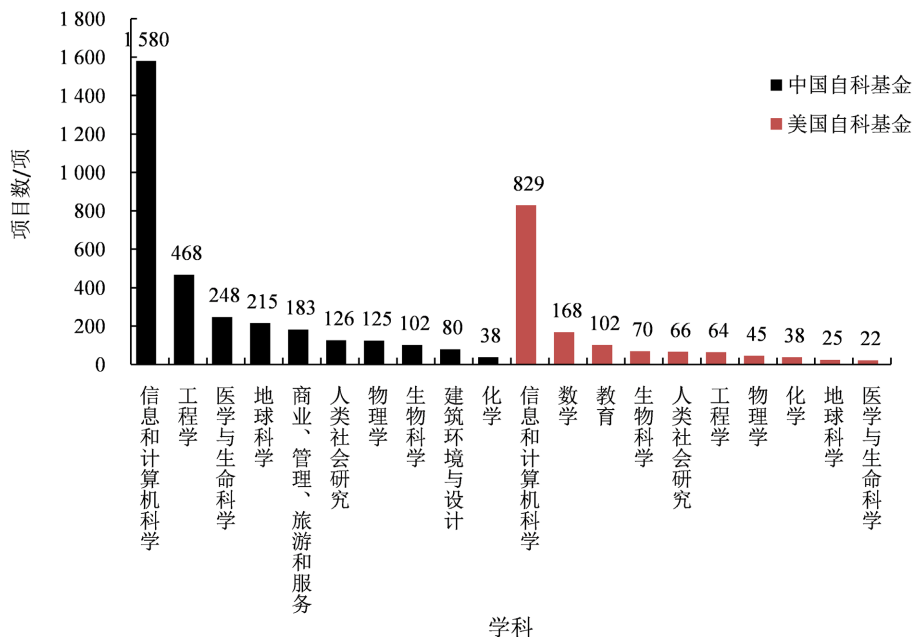


Figure 2. Statistics by number of projects on the subject distribution of big data research projects of the China and the United States Natural Science Foundation

图 2. 中美国家自然科学基金大数据科研项目学科分布按项目数统计

### 3.3. 主要资助机构分析

大数据技术领域科研项目按资助机构统计(见表 1)，计算机与信息科学及工程学部、信息科学部、数学与物质科学部的资助项目最多，共资助项目 4774 项，占总项目数的 55.65%，其中，美国自然科学基金计算机与信息科学及工程学部资助项目 2304 项，占总项目数的 26.86%。从主要资助机构所属国家可以看出，我国有 6 个资助机构最多，即信息科学部 1657 项、基金委管理科学部 718 项、基金委工程与材料科学部 286 项、基金委地球科学部 269 项、基金委数理科学部 258 项、自然科学基金委 249 项，另外我国自然科学基金的医学科学部、国际合作局、生命科学部、化学科学部资助机构科研项目也相对较多；其次美国资助机构有 4 个信息科学部 2304 项、数学与物质科学部 813 项、工程学部 465 项、教育与人力资源部 255 项，另外美国自然科学基金的行为与经济科学部、生物科学部、地球科学部等资助机构科研项目也相对较多。

### 3.4. 主要承担机构分析

大数据技术领域科研项目按承担机构统计(见表 2)，排名前 10 的主要承担机构有 9 所高等院校和 1 所科研机构，共承担科研项目 825 项，占总项目数 9.62%。其中，中国和美国各 5 个机构，我国主要承担机构的项目数共 392 项，美国主要承担机构的项目数共 433 项。

**Table 1.** Statistics by project number of funding agencies for big data research projects of the China and the United States Natural Science Foundation**表 1.** 中美国家自然科学基金大数据科研项目资助机构按项目数统计表

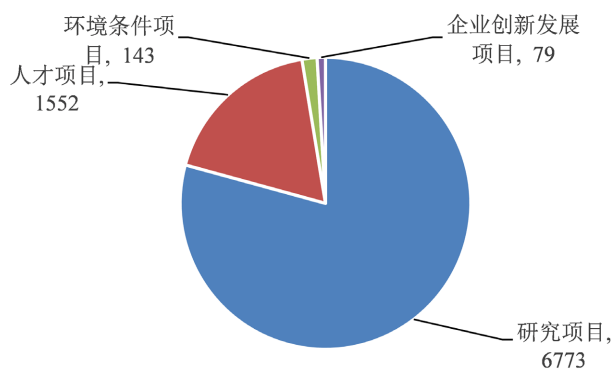
序号	资助机构	所属国家	项目数(项)
1	计算机与信息科学及工程学部	美国自然科学基金	2304
2	信息科学部	中国自然科学基金	1657
3	数学与物质科学部	美国自然科学基金	813
4	基金委管理科学部	中国自然科学基金	718
5	工程学部	美国自然科学基金	465
6	基金委工程与材料科学部	中国自然科学基金	286
7	基金委地球科学部	中国自然科学基金	269
8	基金委数理科学部	中国自然科学基金	258
9	教育与人力资源部	美国自然科学基金	255
10	自然科学基金委	中国自然科学基金	249

**Table 2.** China-US home science fund big data research project authorities by project statistics**表 2.** 中美国家自然科学基金大数据科研项目承担机构按项目数统计表

序号	承担机构	所属国家	项目数(项)
1	清华大学	中国	129
2	Carnegie-Mellon University (卡内基梅隆大学)	美国	89
3	武汉大学	中国	82
4	University of Illinois at Urbana-Champaign (伊利诺伊大学香槟分校)	美国	80
5	Georgia Tech Research Corporation (乔治亚理工研究公司)	美国	76
6	上海交通大学	中国	76
7	University of Minnesota-Twin Cities (明尼苏达大学双城分校)	美国	74
8	北京大学	中国	74
9	Pennsylvania State Univ University Park (宾夕法尼亚州立大学公园)	美国	73
10	浙江大学	中国	72

### 3.5. 项目类型分析

大数据技术领域科研项目按项目类型统计(见图 3), 主要有研究项目 6773 项、人才项目 1552 项、环境条件项目 143 项、企业创新发展项目 79 项, 研究项目最多占比达 78.96%; 研究项目中清华大学 102 项、卡内基梅隆大学 89 项、武汉大学 66 项, 人才项目清华大学 19 项、武汉大学 13 项, 环境条件项目中清华大学 7 项、武汉大学 3 项。按中国自然科学基金的原始项目类型划分, 其中, 我国自然科学基金的项目主要有面上项目 1635 项、青年科学基金项目 1281 项、国际(地区)合作与交流项目 195 项、地区科学基金项目 181 项、联合基金项目 155 项、重大研究计划 134 项、重点项目 133 项等, 美国自然科学基金的项目主要有 Standard Grant 3508 项和 Continuing grant 797 项。



**Figure 3.** Disciplinary distribution of big data research projects of China-US Household Science Fund

**图 3.** 中美国家自然科学基金大数据科研项目学科分布

## 4. 我国自然科学基金大数据技术领域科研项目态势分析

### 4.1. 省市分布情况

在大数据技术领域科研项目中，我国科研项目共 4035 项，从省市分布情况看(见表 3)，按项目数量划分，主要分布在北京 1007 项、上海 339 项、江苏 321 项、广东 288 项等省市；按项目经费划分，主要分布在北京 14612.96 万美元、广东 4847.99 万美元、上海 3732.9 万美元、江苏 2772.84 万美元等省市。数据说明，我国大数据技术领域的核心区域集中在北京、上海、江苏、广东等东部发达地区。

**Table 3.** Statistics on the distribution of big data research projects in provinces and cities of China self-research  
**表 3.** 中国自然科学基金大数据科研项目省市分布按项目数统计表

序号	省市	项目数(项)	项目经费(万美元)
1	北京	1 007	14612.96
2	上海	339	3732.90
3	江苏	321	2772.84
4	广东	288	4847.99
5	湖北	246	2540.83
6	湖南	173	1579.32
7	陕西	171	1975.60
8	浙江	164	1775.41
9	辽宁	147	1565.46
10	四川	139	1578.71

### 4.2. 承担机构分析

在大数据技术领域科研项目中，清华大学、武汉大学、上海交通大学的项目数居前 3 (见表 4)。从承担机构类型看，有 7 所综合性大学和 3 所专业性大学；从所属省市看，北京、上海、武汉均有 2 所高等院校。其中，清华大学在 2017 年立项 29 项最多，工程学立项 18 项，研究项目 102 项；武汉大学在 2016 年立项 18 项最多，信息和计算机科学立项 49 项，研究项目 66 项；上海交通大学在 2017 年立项 23 项最多，信息和计算机科学立项 33 项，研究项目 56 项。数据说明，我国大数据技术领域的核心机构主要是

清华大学、武汉大学、上海交通大学、北京大学等综合性高等院校。

**Table 4.** Distribution of undertaking institutions for big data research projects of China Science and Technology Fund by project statistics

**表 4.** 中国自然科学基金大数据科研项目承担机构分布按项目数统计表

序号	承担机构	所属省市	项目数(项)
1	清华大学	北京	129
2	武汉大学	武汉	82
3	上海交通大学	上海	76
4	北京大学	北京	74
5	浙江大学	浙江	72
6	华中科技大学	武汉	70
7	哈尔滨工业大学	黑龙江	65
8	中南大学	湖南	58
9	中山大学	广东	57
10	复旦大学	上海	57

### 4.3. 资助主题分析

在大数据技术领域科研项目中，主要有大数据、数据挖掘、深度学习 3 大主题，项目数分别为 586 项、368 项、142 项。其中，大数据资助主题主要分布在北京 155 项、上海 51 项、江苏 50 项；数据挖掘资助主题主要分布在北京 99 项、江苏 26 项、上海 25 项；深度学习资助主题主要分布在北京 36 项、广东 13 项、湖北 11 项。

## 5. 结语

通过对近 10 年中美自然科学基金大数据技术领域的科研项目数据对比分析，进一步深入研究我国的大数据技术领域的科研态势，研究发现：1) 在总体态势上，近 10 年中国和美国的大数据技术领域的科研投入处于快速成长时期，2018 年后相对有所下降态势；2) 在学科主题分布上，中国和美国的大数据技术领域的学科布局均是信息和计算机科学领域，其次重点侧重于工程学、医学与生命科学领域；3) 在资助机构分布上，大数据技术领域的科研资助基金以美国自然科学基金的计算机与信息科学及工程学部、中国自然科学基金的信息科学部为主；4) 在承担机构分布上，大数据技术领域的核心力量是中国清华大学和美国卡内基梅隆大学，另外美国的伊利诺分大学香槟分校、乔治亚理工研究公司和明尼苏达大学双城分校等，以及我国的武汉大学、上海交通大学、北京大学和浙江大学等；5) 在项目类型上，中美国家自然科学基金的大数据技术领域以研究项目为主导，其中美国主要是 Standard Grant 和 Continuing grant，而我国主要是面上项目和青年科学基金项目；6) 我国大数据技术领域自 2012~2017 年处于调整发展态势，2018 年后出现明显的下降态势，科研基金投入主要分布在北京、上海、江苏、广东等省市，代表性的科研机构是清华大学、武汉大学、上海交通大学，大数据技术领域的重点研发方向以大数据、数据挖掘、深度学习为主，其中北京、上海的大数据和数据挖掘的科研投入最多。

## 基金项目

2022 年吉林省科学技术信息研究所基本科研经费项目(JX-JBKY-2022-03)。

## 参考文献

- [1] 董恩康. 大数据的计量数据采集与应用分析[J]. 网络安全技术与应用, 2020(11): 75-76.
- [2] 马海韵, 袁园. 大数据驱动下公共治理变革研究综述——基于 2013-2020 年 CSSCI 及核心期刊来源的文献计量分析[J/OL]. 北京交通大学学报(社会科学版): 1-12. <https://doi.org/10.16797/j.cnki.11-5224/c.20220420.006>
- [3] 陈航宇, 杨鑫, 罗子江. 基于文献计量的大数据技术研究现状与趋势[J]. 产业与科技论坛, 2022, 21(2): 57-58.
- [4] 伍之昂, 赵新元, 黄宾, 等. 基于文献计量的大数据管理决策研究热点分析[J]. 管理科学学报, 2021, 24(6): 117-126.
- [5] 孙鸿飞, 张海涛. 基于文献计量与可视化方法的国内外大数据领域研究动态研究[J]. 情报科学, 2018, 36(11): 169-176.
- [6] 王雪. 基于文献计量的国内外大数据研究对比分析[J]. 高校图书馆工作, 2017, 37(4): 49-54.
- [7] 乔楠. 我国政务大数据研究回顾——基于 CSSCI 的文献计量分析[J]. 信息技术与信息化, 2020(1): 165-167.
- [8] 冯鑫, 李佳培, 吴晔, 等. 近 10 年大数据研究热点演进及平台发展——以智慧教育领域 1469 篇文献计量分析为例[J]. 中国高校科技, 2021(Z1): 43-47.
- [9] 麻桂新. 农业大数据研究的文献计量分析[J]. 农业经济, 2020(10): 7-9.
- [10] 张怡宁, 李健. 贵州省大数据政策计量分析[J]. 江苏科技信息, 2020, 37(24): 25-31.
- [11] 雷水旺. 国内图书情报领域大数据研究的文献计量分析[J]. 新世纪图书馆, 2017(8): 91-96.
- [12] 缪瑞生, 马海群. 国内图书情报领域大数据研究的文献计量分析[J]. 情报科学, 2017, 35(3): 93-97.
- [13] 王贤文. Altmetrics: 大数据时代的科学计量学[J]. 图书情报知识, 2019(2): 4.