

NSF 材料科学十年一基金项目分布及趋势变化分析^{*}

陈 挺¹ 李国鹏¹ 姜 山² 王小梅^{*,1}

(1. 中国科学院科技战略咨询研究院, 北京 100190; 2. 中国科学院武汉文献情报中心, 武汉 430071)

摘 要:结合文本挖掘技术与定量分析方法,以美国国家科学基金会(NSF)材料学部 2007~2016 年十年基金资助项目为分析数据,从项目的资助机构分布、重点资助方向、资助类型等角度开展统计分析,展示了学科总体资助趋势变化与重点资助项目构成。同时利用隐形语义模型和 K-means++ 聚类算法对项目申请书摘要进行主题分析,并使用回归拟合分析项目主题的投入金额变化趋势,直观展现了项目主题十年间的发展变化趋势。通过基于数据的统计分析与文本挖掘分析,客观地呈现了 NSF 材料学部资助项目的发展轨迹、现状与未来发展趋势。研究发现,NSF 材料学部的资助金额近年来呈持续下降的趋势,2011 年之后几乎不再资助大型研究中心与大型仪器设备;在通过聚类分析方法发现的 47 个材料领域项目主题中,NSF 对材料相关仪器设备的投入不论是资助金额还是资助数量均远高于其他项目主题,尤其重视射线光源和显微镜成像等材料相关表征设备。从主题的资助金额时序中可以看出,“晶体生长”和“界面研究”两个主题在过去几年内的资助金额增长最为迅猛,可能是材料领域近期的资助热点。

关键词:项目分布;材料研究;NSF;文本聚类

中图分类号:G350

文献标识码:A

doi:10.16507/j.issn.1006-6055.2017.09.005

Past Decade of NSF Material Science: An Analysis of Layout and Trend of Funded Projects^{*}

CHEN Ting¹ LI Guopeng¹ JIANG Shan² WANG Xiaomei^{*,1}

(1. Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190, China;

(2. Wuhan Branch of National Science Library, Chinese Academy of Sciences, Wuhan 430071, China)

Abstract: This paper combines text mining technology and quantitative analysis methods to analyze the funding data of National Science Foundation (NSF) projects between 2007 and 2016. Research is conducted from the perspective of funding priority, funding distribution, topic structure and discipline to analyze the NSF scientific research layout and developmental trend in the field of materials science, revealing the direction and key points of scientific development. From an objective view of point, observing the distribution of leading scientific research, Latent Semantic Analysis (LSA) and K-means++ clustering algorithm are used to cluster project abstract applications to further analyze research trends. Regression analysis is also used show the change in the trend of research investments, offering an in-depth view of the developments and changes of research topics over the past 10 years. The research shows the amount of funding provided by NSF materials division has declined in the past 5 years, the funding for research centers and large equipment also has reduced since 2011; 47 research topics are discovered by using clustering algorithm. Further analysis discovers that NSF invested more money in equipment-related topics, particularly X-ray light source and microscope. As observed from the trend of funding, funding of 2 topics, i. e. "Crystal growth" and "interface research", increased rapidly in the past 5 years. It is likely that these two are to be the hot topics in materials research.

Key words: funding layout; material research; NSF; text clustering

1 引言

国立科研资助机构的资助项目是国家科技研发工作的具体部署,体现了国家科技发展战略和科学

研究活动的重点与方向。而科研项目申请书正是这样一种载体,不仅能从微观上反映个体研究的内容和设想,还能从宏观上反映未来一定时期国家科技发展的部署。

国内外不少学者利用科研基金项目数据开展了分析研究,常见的有以下四种:

第一类是利用基金项目自身提供的标注数据进行统计分析,如陈丽贞等^[1]利用项目的承担机构数据统计中国自然科学基金依托单位的学科分布,分

2017-07-18 收稿,2017-09-15 接受,2017-09-22 网络发表

^{*} 国家自然科学基金(71173211),中国科学院科技战略咨询研究院青年基金(Y7X1161Q01)资助

^{**} 通讯作者,E-mail:wangxm@casipm.ac.cn

析重点资助机构;秦嘉杭^[2]通过统计社科基金项目包含的关键词,发现特定领域资助热点词;马凌远^[3]统计了社科基金实际研究时间,发现不同学科领域下的基金研究时长差异;国外学者 Eckhouse^[4]统计对比了不同国家的项目资助来源,发现不同国家地区的科研基金资助行为差异;Dorsey 等^[5]统计分析了 2003 ~ 2008 年 NIH 对生物医药研究的资助金额变化趋势;Gonzalez^[6]统计分析了 2003 ~ 2012 十年美国科学基金会(NSF)对科学、技术、工程和数学(STEM)相关教育资助的金额变化情况。

第二类是结合基金项目与论文数据开展的投入产出分析,基于基金产出的研究论文的质量与数量,对比分析了不同国家的科研基金投入产出现状。这类研究是现阶段基金分析的主要方向,国内有王贤文等发表的关于全球主要国家基金资助与论文产出的对比分析^[7,8];张诗乐等^[9]分析了 NSFC 资助的论文、其他基金资助论文与无基金资助论文的被引差异;宋志红等^[10]通过基金与论文产出数据,对比了受到资助的学者与未受资助学者的学术表现;国外学者如 Auranen 与 Nieminen^[11]利用基金与论文产出对比世界多所大学投入产出情况;Wang 与 Shapira^[12]利用论文中致谢数据,关联了基金资助与论文产出,分析了各国投入产出与基金资助研究的合作关系;Ebadi 和 Schiffauerova^[13]分析了加拿大研究人员的性别、所属团队规模与基金投入与论文产出之间的关系。

第三类研究是利用基金资助量化基础研究竞争力,如华子春与王雨轩^[14]利用基金相对资助率构建了反应机构或者国家科研竞争力的指标;马廷灿等^[15]使用各省获得国家自然科学基金资助信息,设计了基础研究综合竞争力指数,对我国大陆 31 个省市的基础研究竞争力进行对比分析。类似的研究还有李瑞楠等^[16]利用国家自然科学基金资助情况对比分析我国城市基础研究竞争力;国外学者 CE Garcia 等^[17]早在 2004 年开始便设计基于基金资助竞争力指标,展开了一系列利用基金资助量化基础研究竞争力的研究。

而第四类是利用项目申请书摘要进行文本挖掘,从而发现资助布局。近年来,国内外学者逐渐从基于项目数据的统计分析转向发现基金主题布局研究。笔者所在的研究团队曾对比分析 NSF 与欧盟框架计划对大型强子对撞机(LHC)相关基金项目布局差异^[18]。国内外多位学者采用 LDA 主题模型

发现资助主题,如王效岳等^[19]学者使用 LDA 主题模型和网络结构探测基金研究前沿主题探测;De Miguel Molina^[20]同样利用演化主题模型,发现相关利用领域主题及其演化。

综合上述研究发现,之前的科研基金分析往往受制于资助项目数据,要么针对资助机构提供数据简单统计,要么结合产出论文进行文献计量分析,分析维度往往过于宏观,不能有效揭示科研主题分布及研究内容的细致差异。而基于主题模型的项目主题发现研究往往聚焦在词主题的发现及其变化,未能针对项目本身进行深度分析。

本文综合利用文本挖掘技术和专家定性分析,基于 NSF 资助机构的资助项目申请书摘要,从主题结构、资助重点、发展演变等角度开展比较研究,利用聚类与回归算法分析美国 NSF 的科研项目分布和发展趋势,揭示科学发展的方向与重点。由于材料科学在人类生产活动中举足轻重的重要地位,许多发达国家和新兴经济体都高度重视材料科学及其相关学科的研究与应用,这也是各个国立科研资助机构的重点资助方向。同时,材料研究因为其研究涉及学科、环节较多,很难用一个简单统一的分类体系进行细致分类,因此本文针对 NSF 材料学部资助的项目重点分布进行比较分析,以揭示材料科学领域的研究方向分布、结构和差异,分析每个研究方向中的总体资助规模、发展趋势等情况,从客观角度为我国决策者和科研工作者提供一种新视角观察世界科技发展部署。

2 分析数据与方法说明

2.1 数据来源与清洗

NSF 每年收到超过 30000 项资助申请,资助约 9000 个项目^[21],其中材料学部每年资助约 500 项研究项目。作为美国国立科研资助机构,NSF 一直是政府机构数据开放透明的典范。为了响应奥巴马政府发表的“透明开放的政府”备忘录,NSF 于 2010 与 2012 年两次发布“开放政府计划”,在机密与隐私的约束范围内最大限度地开放资助科研项目产生的数据,且均为机器可读格式。如今 NSF 官方网站中提供了历年资助统计数据、发布相关报告 PDF 文档下载,和 1959 年至今所有资助项目的 XML 数据下载,包括题名、项目承担机构、承担人、项目类型、所属学部、申请书摘要等多个字段信息。笔者构建的科技决策知识服务系统^[22]全球基金项目资助数

数据库收集了美国、欧盟、中国等 11 个国家 28 个资助机构 450 万条资助项目。本文从中筛选了由美国 NSF 材料学部 2007 ~ 2016 年的全部资助项目 4824 条,数据中包括资助项目的题名、摘要、资助起始年、资助对象、资助金额等分析字段。

为了提高文本聚类算法的准确性,从而更准确地发现 NSF 的项目科研重点分布,本文去除了研究项目中资助纯学术会议、科普、教学等非研究项目,以及摘要字段缺失的项目,最终只保留 4281 项资助项目。为了提高聚类分析的准确性,尽可能保留申请书文本中最核心的描述研究科研目标、方法的文本,本文进一步对申请书文本摘要文本内容进行清洗,删除了文本中跟研究内容相关性不强的文本内容,数据清洗规则如下:删除项目申请书摘要首段中资助来源介绍、承担机构等背景介绍内容;删除申请书摘要中“boarder impact”和“non-technical”相关内容,其中“boarder impact”包含科研项目完成后更广泛的科普、培养研究生等方面的社会影响,“non-technical”文本包含了项目的研究内容对公众、对社会的重要性。然后,删除文中小于两个字母的单词和停用词;删除所有的标点符号;对剩下的文本所包含的单词进行词干化处理。本研究中的项目申请书数据清洗环节均采用计算机程序根据规则自动批量清洗。自动清洗与词干化流程如下:

- 1) 去除会议、科普、教学类的非实质性研究项目,通过项目题名中“Conference”、“Workshop”、“Summer School”、“Consortium”、“Summit”、“Symposium”等关键词删除非相关项目;
- 2) 去除摘要字段缺失的数据;
- 3) 删除项目申请书摘要资助来源介绍、承担机构背景、师资力量介绍等内容,通常为三段文本的首段;
- 4) 删除项目申请书摘要中“boarder impact”部分,通常为最后一段的最后一句话;
- 5) 删除项目申请书摘要中“non-technical”部分,通过 non-technical 作为关键字判断。

函数 Clean_NSF_abs 伪代码

```
函数 Clean_NSF_abs():NSF 申请书只要清洗函数
输入: NSF 原始项目申请书摘要文本
输出: 清洗并词干化的项目申请书文本
function Clean_NSF_abs( list raw_NSF_abs ):
    for i,abs in enumerate( raw_NSF_abs ):
        #全部字母小写,替换有意思的特殊符号
        abs.lower().strip().replace( 'dr.', 'dr' ).replace( 'ég.', 'ég' )...
        #删除背景介绍
```

```
if abs has 3 or more phrase:
    abs.remove( first phrase ) # 一般第一段为背景介绍
else
    abs.remove( first sentence )
#删除结尾广泛影响
if last phrase contains "broader impact":
    abs.remove( content after "broader impact" )
else
    abs.remove( content from "broader impact" to first full stop )
#删除非技术部分文本
abs.remove( sentence contains "nontechnical" )
#删除非摘要中的 HTML 标签,标点符号和小于 3 个字母的词
abs.remove( html tag )
abs.remove( punctuation )
abs.remove( len( word ) < 3 )
#删除停用词,文本词干化
abs.remove( stop word )
abs.stem()
return abs
```

2.2 项目主题发现方法

与论文数据相比,项目数据没有论文的引用数据,主题发现分析往往采用文本特征聚类方法。本研究使用团队之前的基金项目申请书文本聚类方法^[18],采用隐性语义模型(Latent Semantic Analysis, LSA)^[23]表征项目申请书文本特征,与 Tf-idf 空间向量模型相比,LSA 将词和文档映射到潜在语义空间,有效降低矩阵的维度,去除一些“噪音”,同时能有效改善 Tf-idf 空间向量无法有效处理的一词多义与一意多词问题。聚类算法则采用 K-Means 改进算法 K-Means + + 算法进行文本聚类。K-means + + 聚类算法因为其简单高效,是应用最为广泛的聚类算法,但它最大局限性是需要在聚类前人工指定聚类个数 K,即多维空间内中心点个数^[24],本研究采用聚类轮廓系数^[25]辅助确定初始聚类数 K,进而调节 K 值大小,并结合专家判读的方式确定最佳聚类个数 K。为了进一步分析项目主题的时序变化趋势,本研究采用基于最小二乘估计法改良的岭回归(RidgeRegression)算法^[26]对相对离散的时序分布进行多次、不同时间段的线性拟合,通过线性拟合的回归线斜率判断十年、五年的项目主题资金投入趋势变化。本文所涉及的算法都使用 scikit-learn^[27]机器学习库实现。

3 资助项目概况分析

3.1 项目数量与资助金额变化趋势

2007 ~ 2016 年,美国 NSF 材料学部共资助了纯

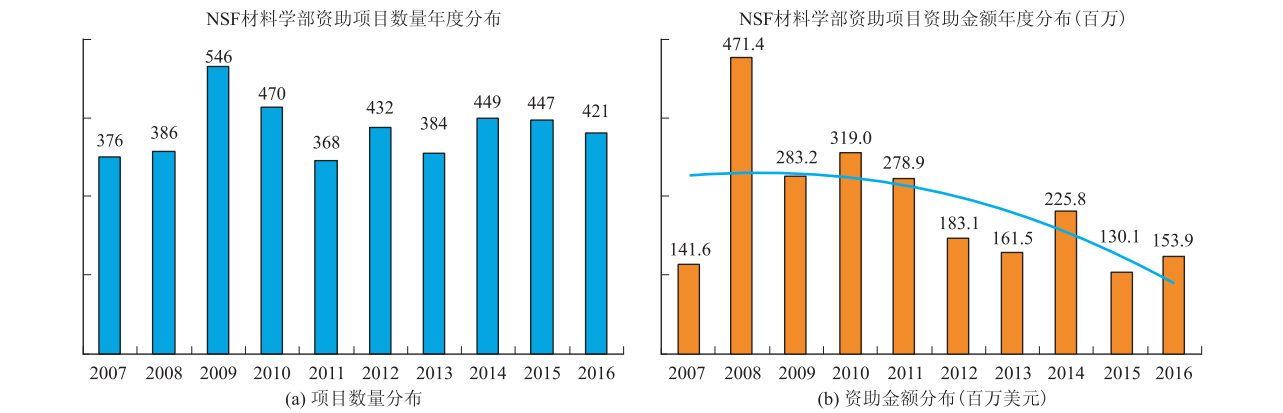


图 1 NSF 材料学部资助项目数量与金额年度趋势变化

学术研究项目 4281 项,平均每年约 428 项,资助立项数量与资助金额年度分布情况如图 1 所示。通过年度立项数与立项资助金额趋势可以明显发现,NSF 材料学部资助年度数量没有明显的变化,但年度资助金额却大幅下降。资助金额从 2010、2011 年的每年平均近 2.9 亿下降到了 2012~2015 年平均近 1.6 亿,项目平均资助金额也有持续下降的趋势。在材料领域内的资助经费消减,与美国整体科研经费下降的趋势相符。在金融危机和债务危机的影响下,美国国会于 2011 年通过《预算控制法》,要求联邦政府开支在 2013 财政年度削减 1090 亿美元。从 NSF 发布的 2015 财年的报告^[28]中看出,从 2011 年开始,联邦政府为了对应金融危机,也大幅消减了科学和工程项目的资助,联邦政府对于科研的资助从 2011 年开始逐年下降。NSF 从 2013 年开始资助金额同样呈下降趋势(表 1)。

表 1 美国联邦政府与 NSF 资助金额的年度变化(百万美元)^[28]

	2011	2012	2013	2014	2015
联邦政府总资助	40,766	40,140	39,444	37,961	37,877
NSF 资助	5,140	5,276	5,393	5,125	5,114

3.2 项目承担机构分析

统计 NSF 材料学部资助项目承担机构后发现,项目承担机构主要以美国本土大学为主。项目资助金额前 20 的承担机构如表 2 所示,均是美国传统名校,2007~2016 年获得 NSF 材料学部资助最多的机构为康奈尔大学和佛罗里达州立大学,虽然获得资助总项目数并不是最多,但不论是项目总金额还是平均资助金额均远远超过其他承担机构。康奈尔大学和佛罗里达州立大学获得的资助金额都是第三名的 3 倍,平均资助金额超过第三名的 3~5 倍。通过分析这两所大学获得资助项目数发现:佛罗里达州

立大学排名第二名,但所获得的资助项目中有一个 2008 年的大型实验室项目 High Magnetic Field Laboratory,资助额高达 1.6 亿,除去该项目后,佛罗里达州立大学其他 47 个资助项目仅获得不到 2000 万的资助金额。康奈尔大学则相继获得了建设多个大型设备和材料研究与创新卓越中心的项目资助。3~20 名的承担机构与前两名相比差距较大,但各承担机构获得的资助金额没有太大差别。

表 2 承担机构获得总项目资助金额前 20 列表

机构名称	项目数	金额 (万 \$)	资助强度 (万 \$/项)
Cornell University	69	20055.7	290.7
Florida State University	48	17636.7	367.4
Pennsylvania State Univ University Park	89	6335.7	71.2
University of Wisconsin-Madison	64	5805.6	90.7
University of California-Santa Barbara	75	5697.7	76.0
University of Pennsylvania	64	5506.3	86.0
Massachusetts Institute of Technology	66	5117.0	77.5
Northwestern University	91	5106.8	56.1
Harvard University	46	4915.4	106.9
University of Minnesota-Twin Cities	63	4717.1	74.9
Princeton University	41	4356.6	106.3
University of Illinois at Urbana-Champaign	97	4021.2	41.5
University of Chicago	40	3876.4	96.9
University of Michigan Ann Arbor	63	3806.4	60.4
Yale University	52	3401.9	65.4
Georgia Tech Research Corporation	59	3122.7	52.9
University of Massachusetts Amherst	47	3106.7	66.1
University of Colorado at Boulder	40	3000.0	75.0
Columbia University	53	2776.9	52.4
Rutgers University New Brunswick	62	2715.9	43.8
University of Pennsylvania	64	5506.3	86.0
Massachusetts Institute of Technology	66	5117.0	77.5
Northwestern University	91	5106.8	56.1
Harvard University	46	4915.4	106.9
University of Minnesota-Twin Cities	63	4717.1	74.9

3.3 项目资助类型分析

2007~2016 年 NSF 资助项目包括了 4 种类型,Standard Grant、Continuing Grant、Cooperative Agreement 和 Interagency Agreement。从表 3 可以看出,资

助项目数与资助金额最多的类型是 Continuing Grants, 2635 项;其次是 Standard Grant,1595 项;Co-operative 和 Interagency Agreement 分别为 46 项和 3 项。虽然后两种类型的资助项目数非常少,但资助强度却远超其他两种类型的资助项目,尤其是 Co-operative Agreement,每个项目平均资助金额超过 1500 万。

表 3 项目类型与资助金额分析

类型	项目数	资助金额 (万 \$)	资助强度 (万 \$/项)
Standard Grant	1595	65689.5	41.2
Continuing Grant	2635	97864.3	37.1
Cooperative Agreement	46	71965.4	1564.5
Interagency Agreement	3	1344.1	448.0

3.4 重大项目分析

表 4 为资助金额前 20 的项目列表,资助金额最高的项目全部为 Cooperative 和 Interagency Agreement 项目(1 项),资助资金额均超过 1000 万美金。其中仅 2008 年一年,NSF 材料学部资助了大型项目 8 个,2010~2011 两年间再次集中资助了 8 项,2013

~2014 年资助了少量大型项目后就不再资助大型项目了。前 20 的高资助额项目中有 15 个项目为建立实验室或者研究中心,包括了 5 个材料研究与创新卓越中心(CEMRI-Center of Excellence for Materials Research and Innovation,表中标灰项目);资助金额最高的项目是 NSF 于 2008 年资助的高磁场实验室的持续建设项目,资助金额高达 1.5 亿美金;第二第三的重大项目为 NSF 持续资助康奈尔大学的高强度 X 光射线光源(CHESS)^[29]大型设备项目(标题名称标深灰),两年的资助额超过 1 亿美金。通过表中项目起始年字段不难发现,NSF 的资助大型项目和研究中心集中在 2008~2011 年,2011 年之后较少资助。从过去 10 年内 NSF 资助筹建材料研究与创新卓越中心列表(表 5)可以看出,全部 7 个创新卓越中心的资助起始年均均为 2011 年。

4 NSF 材料资助项目主题分布分析

4.1 项目主题发现

通过申请书摘要文本 LSI 模型形成项目申请书

表 4 资助金额前 20 资助项目列表

项目名称	承担机构	资助类型 ¹⁾	资助金额 (万 \$)	起始年
National High Magnetic Field Laboratory Renewal	Florida State University	CA	15672.5	2008
Operation of the Cornell High Energy Synchrotron Source (CHESS)	Cornell University	CA	7494.5	2010
OPERATION OF CORNELL HIGH ENERGY SYNCHROTRON SOURCE (CHESS)	Cornell University	CA	3777.0	2014
IMR;Phase Ib Energy Recovery Linac (ERL) Technology R&D	Cornell University	CA	3094.8	2010
Princeton Center for Complex Materials	Princeton University	CA	2010.0	2008
MIT Materials Research Science and Engineering Center (MRSEC)	Massachusetts Institute of Technology	CA	1999.0	2008
Materials Research Science and Engineering Center	University of Chicago	CA	1739.8	2008
University of Minnesota Materials Research Science and Engineering Center	University of Minnesota-Twin Cities	CA	1699.4	2008
Center of Excellence for Materials Research and Innovation	University of Pennsylvania	CA	1629.7	2011
Center of Excellence for Materials Research and Innovation at UCSB	University of California-Santa Barbara	CA	1542.1	2011
Cornell Center for Materials Research-CEMRI	Cornell University	CA	1514.5	2011
UW CEMRI on Structured Interfaces	University of Wisconsin-Madison	CA	1500.0	2011
NSEC on Templated Synthesis and Assembly at the Nanoscale	University of Wisconsin-Madison	CA	1470.0	2009
Materials Research Science and Engineering Center on Polymers	University of Massachusetts Amherst	CA	1465.5	2008
MRSEC;Center for Nanoscale Science	Pennsylvania State Univ University Park	CA	1373.0	2008
The NIST/NSF Center for High Resolution Neutron Scattering	National Institute of Standards and Technology	IA	1321.4	2010
CEMRI;Multifunctional Nanoscale Material Structures	Northwestern University	CA	1241.5	2011
Materials Research Science and Engineering Center	Harvard University	CA	1224.0	2008
Center for Integrated Quantum Materials	Harvard University	CA	1204.0	2013
NSEC on Molecular Function at the Nano/Bio Interface	University of Pennsylvania	CA	1185.0	2009

1) CA, Cooperative Agreement; IA, Interagency Agreement。

表 5 NSF 资助筹建研究中心项目列表

项目名称	起始年	资助金额(万\$)
Center of Excellence for Materials Research and Innovation	2011	1630
Center of Excellence for Materials Research and Innovation at UCSB	2011	1542
Cornell Center for Materials Research-CEMRI	2011	1514
UW CEMRI on Structured Interfaces	2011	1500
CEMRI: Multifunctional Nanoscale Material Structures	2011	1242
Triangle Center for Excellence for Materials Research and Innovation; Programmable Assembly of Soft Matter	2011	1140
CEMRI for Photonic and Multiscale Nanomaterials	2011	1117

特征向量空间,再使用 K-means + + 聚类算法聚出代表不同研究方向的 47 个项目主题(项目聚类簇),并使用自然语言分析工具提取聚类簇关键词与概念词辅助专家为聚类簇命名。聚类前计算了不同 K 值的聚类轮廓系数,当 K 为 38 或 47 时轮廓系数最高,表明聚类效果理论上最佳。再结合情报人员分析判读,认为 47 个项目主题数较为合理,主题内部所含资助项目研究内容相对一致,同时各主题间研究内容重合较少。

材料领域研究涉及材料功能、组成、性质等多个

研究维度,很难用一个简单统一的分类体系细致地分类。从本文的聚类主题名称中也不难发现,材料资助项目主题既包括石墨烯、量子点、超导体、太阳能电池、有机半导体等新兴材料,也包括各种氧化物、陶瓷、玻璃、金属、合金、膜、聚合物、纳米材料、半导体等传统材料研究。相比新兴材料,传统材料主题涵盖了更多的材料类型。另一些项目主题则是材料研究的共性技术,如界面研究、缺陷研究、相变研究、表征、晶体学、动力学等,这类领域也十分宽泛,几乎大部分材料学研究都可能涉及。图 2 为 47 个项目主题的所含项目数,表 6 为情报人员人工判读后各项目主题命名。

4.2 项目主题资助强度分析

发现资助项目主题后,进一步通过资助数量、资助总金额和平均资助金额分析主题资助强度。从表 7 看出,NSF 材料学部十年间资助超过 5000 万美金的主题有:“界面研究”6821.4 万、“量子点、量子性质”7300.8 万、“表征(显微镜、扫描等)”6747.9 万、“表征(X 射线、光谱等)”2 亿、“磁性及磁性材料”2.4 亿、“研究中心、培训等”3.3 亿、“聚合物 2”501.7 万、“太阳能电池”5136.8 万、“材料建模、模拟”

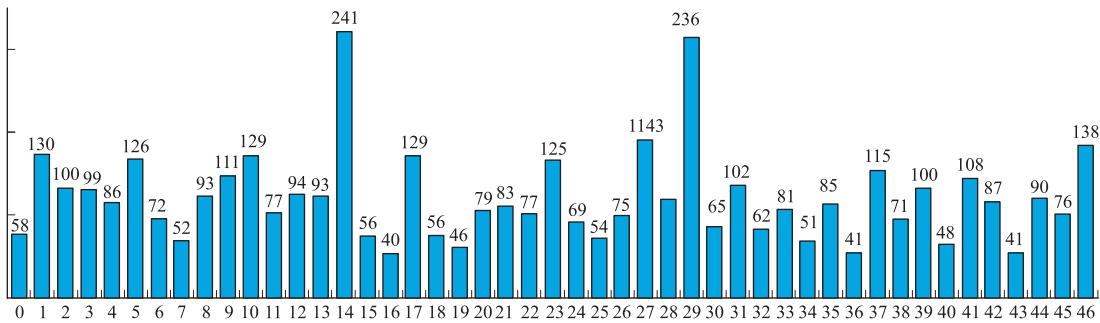


图 2 47 个受资助项目主题的项目数量

表 6 项目主题名称

编号	主题名称	编号	主题名称	编号	主题名称
0	石墨烯	1	关联、自旋轨道耦合、费米	2	各种氧化物
3	界面研究	4	有机半导体	5	量子点、量子性质
6	陶瓷	7	缺陷研究	8	相变等性质研究
9	表征(显微镜/扫描)	10	涉及金属相关研究	11	玻璃
12	超导体、超导性质	13	拓扑性质	14	磁性、磁性材料
15	液晶	16	膜	17	生物(干细胞)
18	聚合物 1	19	晶体研究	20	动力学
21	自旋	22	纳米颗粒	23	研究中心、培训等
24	纳米结构	25	纳米线	26	生物(DNA)
27	量子性质	28	金属/合金/钢铁	29	聚合物 2
30	蛋白质	31	薄膜	32	生物(药物输运)
33	电解质	34	胶体研究	35	太阳能电池
36	碳纳米材料	37	光电	38	共聚物
39	合金	40	流体性质	41	半导体
42	表面材料	43	氢	44	表征(X 射线/光谱等)
46	晶体生长	47	材料建模、模拟等		

5400 万。从投入经费统计上看,NSF 材料学部非常重视研究中心筹建与仪器设备研发。研究中心、培于训主题共资助了 125 个研究项目,总资助额度超过 3 亿,平均资助金额高达 266.4 万,资助总额和平均资助在所有主题中都排名第一。此外,有两个关资助材料相关仪器设备研发主题:“表征(显微镜、扫描等)”和“表征(X 射线、光谱等)”,资助金额都超过 5000 万,其中“表征(X 射线、光谱等)”主题资助额更是高达 2 亿,平均资助金额 226.2 万,涉及 X 射线光源、激光光源、显微镜,核磁共振等相关材料表征设备建设与相关技术研发。“磁性与磁性材料”主题在过去的十年内资助项目最多,共计 241 项,总投入高达 2.4 亿,其中 1.6 亿是 2008 年资助给佛罗里达州立大学的一个项目:National High Magnetic Field Laboratory Renewal,但即便如此,主题内其他项目总投资也超过 8000 万。而“液晶”、“自旋”、“生物(DNA)”、“胶体研究”和“晶体生长”5 个项目主题虽然项目数量较少,资助总额较低,但平均资助金额较高,均超过 50 万美金。全部项目主题的资助数量、资助金额和变化趋势详见附录表 10。

表 7 主题资助数量与资助金额列表¹⁾

主题名称	项目数	资助金额 (万\$)	平均资助 (万\$/项)
3 界面研究 *	99	6821.4	68.9
5 量子点、量子性质 *	126	7300.8	57.9
9 表征(显微镜、扫描等) *	111	6747.9	60.8
14 磁性及磁性材料 *	241	24328.7	100.9
15 液晶 **	56	2792.4	50
21 自旋 **	83	4935.1	59.5
23 研究中心、培训等 *	125	33295.2	266.4
26 生物(DNA) **	75	4513.5	60.2
29 聚合物 2 *	236	9501.7	40.3
34 胶体研究 **	51	3227.5	63.3
35 太阳能电池 *	85	5136.8	60.4
44 表征(X 射线、光谱等) *	90	20362.4	226.2
45 晶体生长 **	76	3911.4	51.5
46 材料建模、模拟等 *	138	5400.1	39.1

1) *:总资助超过 5000 万美金的项目主题; **:总资助额低于 5000 万,平均、资助金额超过 50 万的项目主题。

4.3 主题时序分析

为了进一步分析项目主题的时序变化趋势,本文采用岭回归算法,对较为离散的时序分布进行线性拟合,利用回归线斜率判断趋势变化。通常来说,斜率大于 0.17 就表示趋势是处于上升状态,大于 0.5 表示高速增长,反之小于 -0.17 则为下降趋势。本文分别对各主题的项目数量和项目金额变化进行 3 次回归拟合,包括 10 年总趋势、前 5 年趋势和最近 5 年趋势。图 3 为 47 个主题资助金额趋势图,绝

大部分的主题资助金额处于下降趋势,近 5 年资助金额下降幅度更大,只有少量主题处于上升趋势,这也符合前章节概况分析中 NSF 材料学部资助金额持续下降的结论。主题中项目数量年度变化趋势见附录图 4。

本文对主题的 10 年资助金额变化回归线斜率进行排序,进一步总结出资助金额上升和下降趋势最明显的项目主题。如表 8 所示,资助金额增长最快的项目主题有:“晶体生长”、“拓扑性质”、“材料建模、模拟等”、“界面研究”和“电解质”。其中,“晶体生长”主题不论项目数量变化趋势还是资助金额变化趋势都呈现快速增长,近 5 年资助金额变化回归线斜率高达 1.96。“晶体生长”是材料制备领域的基础性问题,绝大多数无机材料的制备都离不开晶体生长环节,从资金统计分析中可以看出,NSF 近年来对晶体生长相关制备研究有非常重视,有非常显著的加大投入的趋势。近 5 年来“晶体生长”主题中受资助最多的机构是宾夕法尼亚州立大学(表 9),共获得了 3 项资助,资助总额超过 1000 万美金,远超排在第二第三的加州大学河滨分校与西北大学。此外“界面研究”主题的项目数量与资助金额同样呈现全面高速增长的趋势,近 5 年资助

表 8 主题资助金额上升与下降趋势最快主题列表¹⁾

主题名称	10 年	5 年	10 年	5 年
	项目数回归斜率	项目数回归斜率	投入金额回归斜率	投入金额回归斜率
45 晶体生长	0.32	0.4	0.57	1.96
13 拓扑性质	2.22	1.8	0.53	-0.05
46 材料建模、模拟等	1.19	-0.1	0.49	-0.55
3 界面研究	0.28	0.8	0.47	1.9
33 电解质	1.42	2.2	0.4	-0.02
24 纳米结构	-0.87	-0.5	-0.43	-0.34
29 聚合物	-0.68	0.6	-0.93	-0.22
44 表征(X 射线、光谱等)	-0.39	0.1	-1.45	0.32
23 研究中心、培训等	-0.12	1.0	-6.21	-0.39
14 磁性及磁性材料	-2.38	-1.8	-7.4	-0.53

1) 按 10 年资助金额变化趋势排序。

表 9 主题 45、3、23、14 中近五年中获得资助金额前 3 的机构列表

项目主题	机构名称	项目数	金额(万)
45 晶体生长	Pennsylvania State Univ University	3	1079
	University of California Riverside	2	109
	Northwestern University	3	107
3 界面材料	Cornell University	2	950
	University of Wisconsin-Madison	3	295
	University of Florida	4	238
23 研究中心、培训等	University of Chicago	1	675
	University of Minnesota-Twin Cities	1	605
	Massachusetts Institute of Technology	1	565
14 磁性及磁性材料	Florida State University	7	335
	Pennsylvania State Univ University Park	2	159
	Ohio State University	3	149

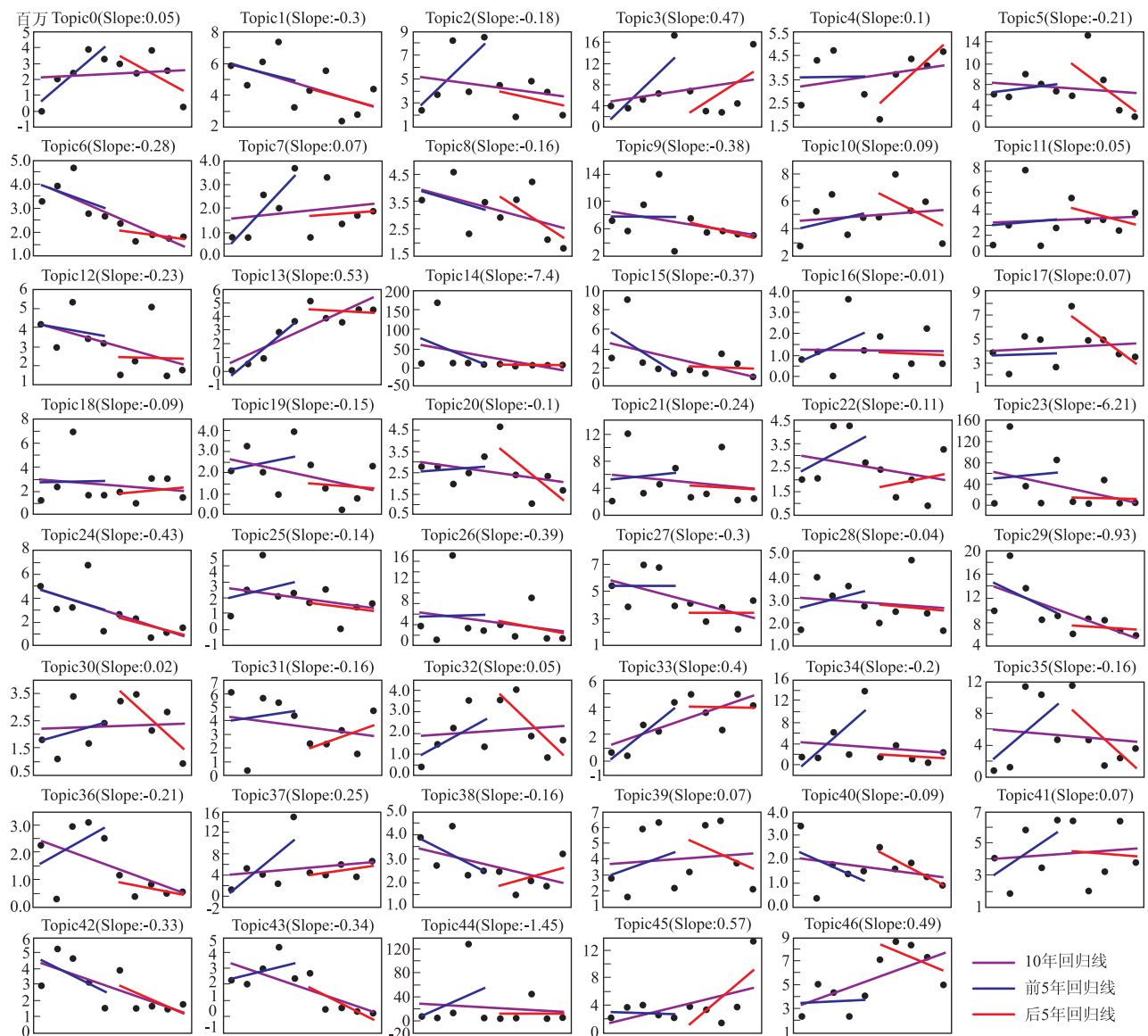


图3 主题中项目资助金额年度变化趋势图

金额变化回归线斜率也高达 1.9,近年来国际上对能源研究的大手笔投入引发了化学能源方面的研究热潮,尤其在催化研究中,催化剂的表面结构与性质是决定催化效率、活性和稳定性的重要因素,因此界面项目主题领域基金的增加,可能与近年来能源领域相关的催化剂研究升温有关。“界面研究”中近 5 年来受资助金额最高的是康奈尔大学,2 个项目 950 万美金,同样远超威斯康辛大学麦迪逊分校和佛罗里达大学。

另一方面,“磁性及磁性材料”、“研究中心、培训等”、“表征(X 射线、光谱等)”、“纳米结构”和“纳米结构”五个主题的资金投入呈大幅度下降趋势,尤其是“磁性及磁性材料”和“研究中心、培训等”主题,十年资金投入变化回归线斜率分别为

-6.21 和 -7.4。磁性材料研究,特别是钕铁硼等永磁材料,由于研究相对比较成熟,研究热度逐渐下滑。近年来,磁性材料领域几乎很少发现具有重大意义的新型材料,而在应用方面,传统磁性材料的应用领域并未得到进一步拓展,这也可能导致相关基础研究基金的缺失。从机构获得的资助金额中也可发现,排名第一的佛罗里达州立大学近五年共获得 7 项资助,但总金额只有 335 万美金,远低于“晶体生长”与“界面材料”研究。另外,虽然 NSF 对材料研究基础设施建设的支持一直比较重视,但近年来 NSF 也降低了对研究中心、实验室的资助,而转为资助利用计算机建模、模拟进行材料研究。造成相关基金支持下滑的原因可能在于,研究中心的建设周期和使用周期较长,而早期 NSF 投资建立的一些大

型中心由于资助周期到期或政策原因退出了资助计划,而新的计划尚未设立,导致相关资金有所下滑。近五年获得资助最高的机构均只获得 1 项资助,但与其他主题项目相比,“基础设施建设”项目资助经费相对较高,前三的机构都获得超过 500 万美金的资助。

5 结语

本文的分析方法比较直观地展示了美国国家科学基金会材料学部资助的项目主题分布,揭示了不同主题的资助趋势,从而了解了资助机构的研究重点以及发展趋势。通过定量分析结果结合科技情报分析专家解读,得出以下结论:

1) 总体来看,NSF 材料学部资助项目数量每年变化不大,但资助金额呈大幅下降的趋势。

2) 2008 ~ 2012 年前后,NSF 材料学部集中资助筹建了多家实验室和研究中心,包括 7 个材料研究与创新卓越中心。

3) 通过文本聚类算法,聚出了 47 个材料项目主题,分析发现 NSF 对材料相关仪器设备的投入不论是资助金额还是资助数量均远高于其他项目主题,尤其重视射线光源和显微镜成像等材料相关表征设备。另外对“界面研究”、“量子点、量子性质”、“聚合物”、“太阳电池”、“材料建模、模拟”5 个主题也投入了高额的研究经费。

4) “晶体生长”和“界面研究”两个主题在过去几年内资助金额增长最为迅猛,可能是近期材料领域的资助热点;作为传统材料的磁性材料可能不再是 NSF 主要资助方向;另外,NSF 也降低了对研究中心实验室的资助,转为资助利用计算机建模、模拟进行材料研究。

本文的研究方法尚有不足,目前仅使用美国 NSF 一个资助机构的基金项目进行实验分析,所有的数据清洗、布局发现方法都是针对单一数据来源,无法对比多个资助机构的布局差异。多个资助机构基金项目主题分布对比面临两个难题:第一,各个资助机构组织架构各不相同,资助项目所属领域很难人工相互匹配;第二,多数据源主题分布差异比较缺少量化方法。这两个问题也是笔者的研究团队未来的研究方向。如能解决这两个难题,便有可能揭示各国家、各资助机构项目布局的特征和差异,可能对科研管理与决策者审视宏观科研布局、前瞻科技发展方向起到更大的辅助作用。

参考文献

- [1] 陈丽贞,李洁,郑世珠,等. 国家自然科学基金项目依托单位的分布统计研究[J]. 科研管理研究,2008,28(3):106-108.
- [2] 秦嘉杭. 图书馆、情报与文献学热点研究主题分布及其发展趋势—基于 2011-2013 年国家社科基金项目分析[J]. 图书与情报,2013,157(6):112-115.
- [3] 马凌远. 国家社科基金项目的实际研究时间及其分类特征—基于 1994-2014 年国家社科基金项目数据的统计分析[J]. 郑州航空工业管理学院学报:社会科学版,2017,36(2):10-16.
- [4] ECKHOUSE S, LEWISON G, SULLIVAN R. Trends in the Global Funding and Activity of Cancer research [J]. Molecular Oncology, 2008,2(1):20-32.
- [5] DORSEY E R., DE ROULET J, THOMPSON J P, et al. Funding of US biomedical research, 2003-2008 [J]. Jama, 2010, 303(2):137-143.
- [6] GONZALEZ H B. An Analysis of STEM Education Funding at the NSF: Trends and Policy Discussion [R]. Washington D C: Congressional Research Service, 2012
- [7] 王贤文,刘则渊,侯海燕. 全球主要国家的科学基金及基金论文产出现状:基于 Web of Science 的分析[J]. 科学学研究,2010,28(1):62-66.
- [8] 孙金伟,刘迪,王贤文,等. 科学基金资助与 SCI 论文产出:对 10 个国家的比较分析[J]. 科学学研究,2013,31(1):36-42.
- [9] 张诗乐,盖双双,刘雪立. 国家自然科学基金资助的效果—基于论文产出的文献计量学评价[J]. 科学学研究,2015(4):507-515.
- [10] 宋志红,郭艳新,李冬梅. 科学基金资助提高科研产出了吗? [J]. 科学学研究,2016,34(1):116-121.
- [11] AURANEN O, NIEMINEN M. University Research Funding and Publication Performance—An International Comparison [J]. Research Policy, 2010,39(6):822-834.
- [12] WANG J, SHAPIRA P. Funding acknowledgement analysis: an enhanced tool to investigate research sponsorship impacts: the case of nanotechnology [J]. Scientometrics, 2011, 87(3):563-586.
- [13] EBADI A, SCHIFFAUEROVA A. How to boost scientific production? A statistical analysis of research funding and other influencing factors [J]. Scientometrics, 2016, 106(3):1093-1116.
- [14] 华子春,王雨轩. 基金相对资助率—反映国家自然科学基金竞争能力的一个新指标[J]. 中国科学基金,2009(1):50-55.
- [15] 马廷灿,曹慕昆,王桂芳. 从国家自然科学基金看我国各省市基础研究竞争力[J]. 科学通报,2011,56(36):3115-3121.
- [16] 李瑞婧,马廷灿,岳名亮. 我国城市基础研究竞争力对比分析—基于国家自然科学基金资助情况的考察[J]. 中国科学基金,2017(4):412-416.
- [17] GARCÍA C E, SANZ-MENÉNDEZ L. Competition for funding as an indicator of research competitiveness [J]. Scientometrics, 2005, 64(3):271-300.
- [18] 陈挺,韩涛,李泽霞,等. 科研项目布局差异对比方法研究—以 NSF 和 EUPF 项目为例 [J]. 现代图书情报技术,2015,31(7/8):89-96.
- [19] 王效岳,刘自强,白如江,等. 基于基金项目数据的研究前沿主题探测方法 [J]. 图书情报工作,2007,61(13):87-98.

[20] DE-MIGUEL-MOLINA B, CUNNINGHAM S W, PALOP F. Analyzing Funding Patterns and Their Evolution in Two Medical Research Topics[J]. International Journal of Innovation and Technology Management, 2017, 14(2): 1740010.

[21] LIU S, CAO N, LV H. Interactive visual analysis of the NSF funding information[C]. Kyoto: Visualization Symposium, PacificVIS '08. IEEE Pacific. IEEE, 2008: 183-190.

[22] 科技决策知识服务系统[EB/OL]. 2016-09-20. <http://strategy.las.ac.cn/strategy>.

[23] DUMAIS S T. Latent semantic analysis[J]. Annual review of information science and technology, 2004, 38(1): 188-230.

[24] ZOUBI M B A, RAWI M A. An efficient approach for computing silhouette coefficients[J]. Journal of computer science, 2008, 4(3): 252-252.

[25] ROUSSEUW P J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis[J]. Journal of Computational and Applied Mathematics, 1987, 20: 53-65.

[26] HOERL A E, KENNARD R W. Ridge regression: Biased estimation for nonorthogonal problems[J]. Technometrics. 1970, 12(1): 55-67.

[27] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. Scikit-learn: Machine learning in Python[J]. Journal of Machine Learning Research, 2011, 12(10): 2825-2830.

[28] Ronda B. Universities Report Fourth Straight Year of Declining Federal R&D Funding in FY 2015[R]. Alexandria, Virginia: National Science Foundation, National Center for Science and Engineering Statistics. 2015.

[29] BATTERMAN B W, ASHCROFT N W. CHESS: The new synchrotron radiation facility at Cornell[J]. Science, 1979, 206(4415): 157-161.

附录

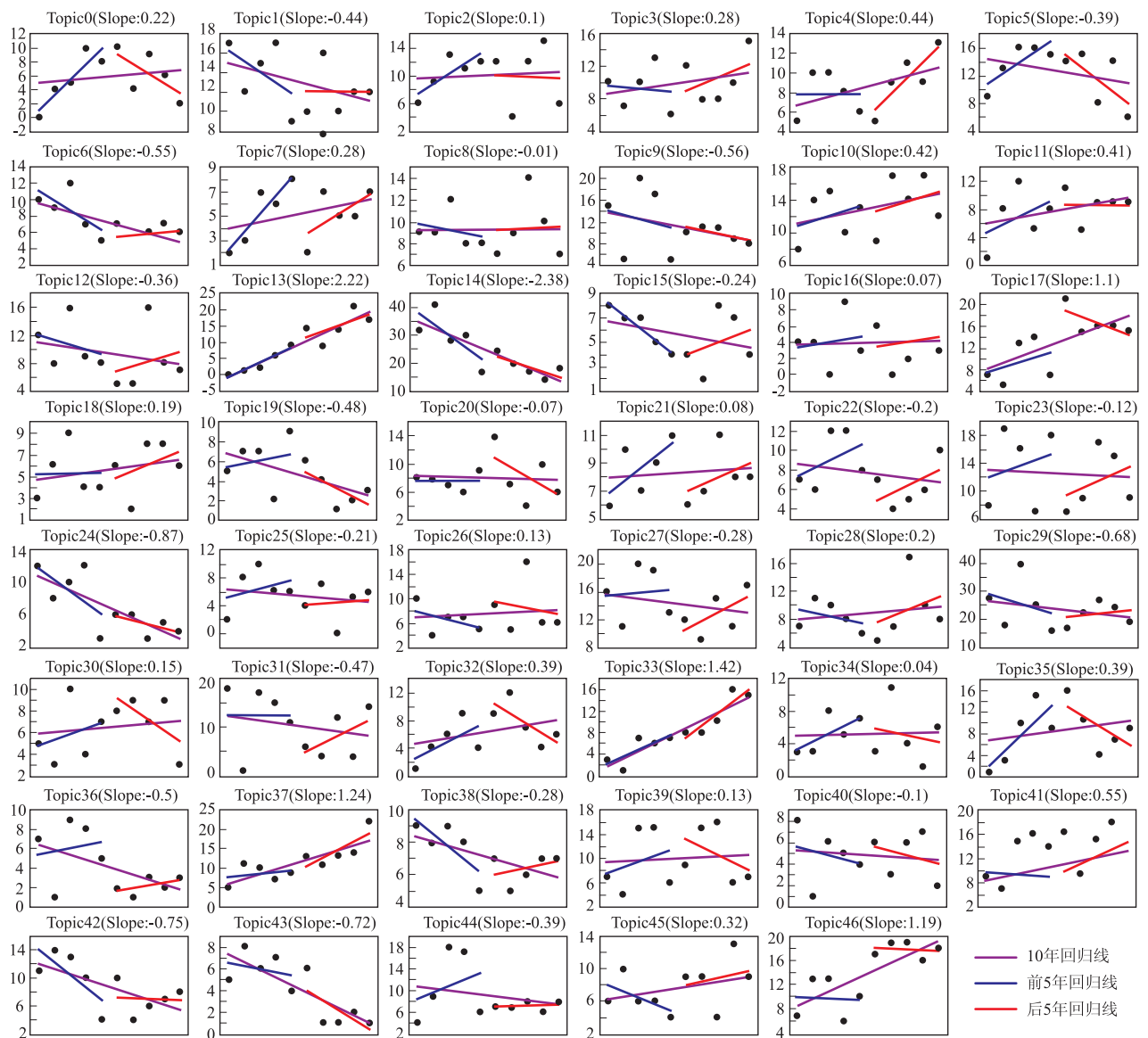


图4 主题中项目数量年度变化趋势图

表 10 主题资助数量、资助金额和变化趋势列表

主题名称	项目数	资助金额 (万 \$)	资助强度 (万 \$/项)	10 年项目 回归斜率	5 年项目 回归斜率	10 年金额 回归斜率	5 年投入 回归斜率
0 石墨烯	58	2324.5	40.1	0.22	-1.4	0.05	-0.53
1 关联、自旋轨道耦合、费米	130	4624.4	35.6	0.42	0.6	0.09	-0.58
2 各种氧化物	100	4360.5	43.6	0.41	0.0	0.05	-0.37
3 界面研究	99	6821.4	68.9	-0.36	0.7	-0.23	-0.02
4 有机半导体	86	3635.0	42.3	2.22	1.8	0.53	-0.05
5 量子点、量子性质	126	7300.8	57.9	-2.38	-1.8	-7.4	-0.53
6 陶瓷	72	2658.1	36.9	-0.24	0.5	-0.37	-0.05
7 缺陷研究	52	1869.0	35.9	0.07	0.3	-0.01	-0.03
8 相变等性质研究	93	3213.6	34.6	1.1	-1.1	0.07	-0.97
9 表征(显微镜、扫描等)	111	6747.9	60.8	0.19	0.6	-0.09	0.11
10 涉及金属相关研究	129	4978.1	38.6	-0.48	-0.8	-0.15	-0.06
11 玻璃	77	3444.6	44.7	-0.44	0.0	-0.3	-0.27
12 超导体、超导性质	94	3095.6	32.9	-0.07	-1.3	-0.1	-0.6
13 拓扑性质	93	2945.0	31.7	0.08	0.5	-0.24	-0.11
14 磁性及磁性材料	241	24328.7	100.9	-0.2	0.8	-0.11	0.14
15 液晶	56	2762.4	49.3	-0.12	1.0	-6.21	-0.39
16 膜	40	1202.0	30.0	-0.87	-0.5	-0.43	-0.34
17 生物(细胞、干细胞)	129	4322.7	33.5	-0.21	0.2	-0.14	-0.13
18 聚合物 1	56	2425.5	43.3	0.13	-0.5	-0.39	-0.56
19 晶体研究	46	1897.0	41.2	-0.28	1.2	-0.3	-0.01
20 动力学	79	2530.7	32.0	0.2	0.9	-0.04	-0.07
21 自旋	83	4935.1	59.5	-0.68	0.6	-0.93	-0.22
22 纳米颗粒	77	2509.7	32.6	0.1	-0.1	-0.18	-0.29
23 研究中心、培训等	125	33295.2	266.4	0.15	-1.0	0.02	-0.52
24 纳米结构	69	2710.3	39.3	-0.47	1.6	-0.16	0.4
25 纳米线	54	1955.9	36.2	0.39	-1.4	0.05	-0.71
26 生物(DNA)	75	4513.5	60.2	1.42	2.2	0.4	-0.02
27 量子性质	143	4389.2	30.7	0.04	-0.4	-0.2	-0.15
28 金属/合金/钢铁	89	2771.5	31.1	0.39	-1.8	-0.16	-1.81
29 聚合物 2	236	9501.7	40.3	-0.5	0.3	-0.21	-0.11
30 蛋白质	65	2287.6	35.2	1.24	2.1	0.25	0.44
31 薄膜	102	3572.1	35.0	-0.28	0.2	-0.16	0.19
32 生物(药物运输)	62	2070.9	33.4	0.13	-1.3	0.07	-0.44
33 电解质	81	2988.0	36.9	0.28	0.8	0.47	1.9
34 胶体研究	51	3227.5	63.3	-0.1	-0.4	-0.09	-0.35
35 太阳电池	85	5136.8	60.4	0.55	1.2	0.07	-0.09
36 碳纳米材料	41	1439.5	35.1	-0.75	-0.1	-0.33	-0.44
37 光电	115	5175.1	45.0	-0.72	-0.9	-0.34	-0.52
38 共聚物	71	2689.5	37.9	-0.39	0.1	-1.45	0.32
39 合金	100	4004.6	40.0	0.32	0.4	0.57	1.96
40 流体性质	48	1627.7	33.9	1.19	-0.1	0.49	-0.55
41 半导体	108	4320.1	40.0	0.44	1.6	0.1	0.61
42 表面材料	87	2774.6	31.9	-0.39	-1.7	-0.21	-2.22
43 氮	41	1806.0	44.0	-0.55	0.2	-0.28	-0.1
44 表征(X射线、光谱等)	90	20362.4	226.2	0.28	0.8	0.07	0.06
45 晶体生长	76	3911.4	51.5	0.01	0.1	-0.16	-0.37
46 材料建模、模拟等	138	5400.1	39.1	-0.56	-0.6	-0.38	-0.53