

· 管理纵横 ·

# 基于大数据分析技术的联合基金项目 典型成果挖掘机制构建与应用

吴卫红<sup>1</sup> 张莉娜<sup>1</sup> 李志兰<sup>2</sup> 张爱美<sup>1\*</sup> 刘忠华<sup>1</sup>

1. 北京化工大学 经济管理学院, 北京 100029

2. 国家自然科学基金委员会 计划局, 北京 100083

**[摘要]** 如何利用大数据分析技术在海量科学基金资助成果中挖掘出具有代表性的典型成果是展示基金资助绩效、推动优秀成果应用贯通的关键。本文利用数据挖掘技术构建了大数据环境下基金项目典型成果的挖掘机制,以论文发表的年份及发表刊物信息为匹配检索点,按照标准筛选高水平论文,从法律、经济和技术三个维度构建包含11个指标的专利价值评估体系,利用第三方专利价值评价服务平台相关信息筛选高质量专利,建立典型经济效益成果的词库,实现典型经济效益的挖掘。在此基础上,以2002—2019年国家自然科学基金联合基金结题项目的成果信息为对象,进行了典型成果的挖掘分析。结果表明,上述典型成果挖掘机制可以分年度、分学部筛选高水平论文和专利;经济效益挖掘结果准确率与项目的实际应用效益吻合度较高。

**[关键词]** 典型成果;挖掘方法;大数据分析技术;联合基金

基础研究是整个科学体系的源头,是科技强国建设的根基。国家自然科学基金作为我国科技体制改革的重要产物,在国家基础研究和科技创新体系中发挥着重要的支撑作用<sup>[1]</sup>。30多年来,国家自然科学基金财政投入从1986年的8000万元增长到2019年的311.42亿元<sup>①</sup>,共资助项目56.4万项<sup>②</sup>,实现了跨越式的增长。国家自然科学基金委员会(以下简称“自然科学基金委”)作为我国支持基础研究的最大机构,在知识生产方面取得了突出的成效,促进了我国学科协调演化、全链条人才培养、研究前沿突破和源头创新供给<sup>[2]</sup>。

随着科技投入的持续增长,基金项目产出成果呈现指数型上升的趋势。在新时期科技体制改革不断深入的背景下,如何对丰硕的基金项目产出成果进行科学高效的管理已成为自然科学基金委面临的重点问题之一。国家自然科学基金“十三五”规划中提出“加强成果管理,完善科技报告制度,持续开展创新成果展示和科学传播”<sup>[3]</sup>,这意味着典型的



**张爱美** 博士,北京化工大学教授,主要从事公司财务、会计、环境经济、项目经济评价等方面的研究。主要参与了国家自然科学基金项目、教育部人文社会科学规划基金项目、北京市社会科学基金项目、北京市自然科学基金项目等多项国家级、省部级项目,发表学术论文50余篇。



**吴卫红** 博士,北京化工大学教授,主要从事创新管理、科技政策与管理等方面的研究。作为课题负责人先后承担了国家自然科学基金项目、教育部人文社会科学规划基金项目、北京市社会科学基金项目、北京市自然科学基金项目、科技部项目等多项国家级、省部级项目,发表学术论文80余篇。

识别和挖掘将成为新时期成果管理的重点工作。虽然国内外学者在关于论文、专利等的评价方面进行了比较深入的探索,但如何在巨大的基金项目成果池中迅速准确地挖掘出典型的成果还未形成一套科

收稿日期:2021-06-17;修回日期:2021-07-02

\* 通信作者,Email: zhangam@mail.buct.edu.cn

① 数据来源于国家自然科学基金委员会年度报告。

② 数据由国家自然科学基金共享服务网基金结题项目数据计算得出。

学且行之有效的的方法体系。典型成果挖掘方法的探索一方面有利于完善我国科学基金的成果管理体系,筛选出高质量的项目成果,丰富成果宣传与展示的内容,提高基金资助绩效;另一方面,高效的成果管理是成果应用贯通的前提,通过对不同类别典型成果的识别与挖掘,针对性地设计转化路线,可以更快地推动基础研究、应用研究到产业化的距离,发挥基金成果服务国家和社会的经济效能。基金项目典型成果的挖掘既是成果管理全过程的重要一环,也是成果应用贯通链条的前端工作,具有十分重要的意义。

在科学基金资助体系中,国家自然科学基金联合基金(以下简称“联合基金”)是由自然科学基金委和联合资助方共同出资设立,促进有关部门、企业、地区与高校和科研机构的合作,培养科学与技术人才,进而推动我国相关领域、行业、区域自主创新能力的提升。联合基金这一独特的“多元投入”资助体系,使得其更加注重与国家战略需求和实际问题的结合。识别与发现联合基金资助项目的典型成果,进行典型成果的共享与传播,促进典型成果的应用贯通,可以更好地发挥出基础研究助力社会经济发展的作用,同时加强全社会对于基础研究多元投入的认识,扩大基础研究的投入规模,具有非常好的示范作用。因此,本研究以联合基金项目为例构建典型成果挖掘机制,设计合理的成果筛选标准与挖掘方法,从海量的研究成果中挖掘出具有代表性的典型成果,为后期的成果管理和成果应用贯通提供理论支撑与决策依据。

## 1 概念界定与研究对象选择

随着国家对基础研究和科技创新的重视,国家自然科学基金的资助规模持续扩大,基金资助取得了丰硕的科研成果。根据《国家自然科学基金资助项目研究成果管理办法》(2015年)并结合基金项目结题报告成果数据统计表的栏目设置<sup>[4]</sup>,本文认为国家自然科学基金项目成果是指国家自然科学基金资助项目经过科学研究取得的论文、专著、软件、标准、重要报告、专利、数据库、标本库及科研仪器设备等有价值的科学技术产出以及项目人才培养、产生的经济效益等隐性成果<sup>[5]</sup>。而在科学基金资助产生的浩瀚成果中,只有高质量、具有重要应用前景和创新性的科研成果,才能解决我国经济发展中的关键科学问题,为提高我国自主创新能力提供科技动力和智力支持,该类成果即为本研究中的典型成果<sup>[6]</sup>。

联合基金项目作为国家自然科学基金的项目类型之一,其一大特点是紧密围绕区域和行业发展的重大关键科学问题开展大量基础性研究<sup>[7]</sup>。经过20多年的发展,联合基金资助的项目已经产生了丰硕的研究成果。如何识别发现联合基金项目的典型成果,并推动具有潜在经济价值的项目成果快速转化落地,促进实际生产关键问题的解决已是当下基金管理的重点工作。因此,本文以2002—2019年联合基金结题项目的成果为研究对象,探究网络时代下,如何利用大数据分析技术实现典型成果挖掘机制构建,从而推动典型成果的共享与传播,促进项目成果更快地应用转化。

根据自然科学基金委提供的2002—2019年联合基金项目结题报告中的成果统计信息,主要包括论文、获奖、专利、软件著作权、成果技术转移、标准等共90605项成果,其中期刊论文和专利产出占比高达80%。基于此,本研究主要围绕期刊论文和专利进行典型成果的挖掘。同时,结合联合基金设立的初衷,经济效益既是典型项目成果的主要体现形式,也是联合基金成立20年来资助绩效的重要衡量方式。基于以上考虑,本文典型成果的挖掘机制构建主要围绕高水平论文、高质量专利以及经济效益展开。

## 2 典型成果挖掘机制构建

典型成果挖掘机制的构建是在基金项目成果信息精准收集和分类管理的基础上,进一步挖掘论文、专利典型成果,明晰联合基金资助20年来共产生了多少高水平论文、有多少较高价值的专利,以及对我国相关领域、行业和区域的经济社会发展产生多少影响,典型的经济效益案例有哪些。主要目的是识别基金资助的典型成果并对其进行宣传和推介,激励科研工作者产出更多高质量的成果,并为相关后续研究的开展奠定基础,也可以在一定程度上检验联合基金资助战略目标的实现程度。

由于基金成果数据库非常庞大,难以对每一个项目的成果进行全面准确的评价,因此在构建典型成果挖掘体系前,需要明确科学基金资助项目成果挖掘的基本原则。有学者曾指出科研项目绩效评价应做到理论性和实用性相结合,完备性和简明性相结合<sup>[8]</sup>,即项目评价要以理论为指导进行实践,指标的构建要简洁且能全面涵盖评价的内容。基金项目典型成果的挖掘也应遵循上述原则。在构建典型成果挖掘机制时应以统计学理论为指导,设计的

指标体系既要全面反映成果的科研价值和经济价值,又需简洁实用、代表性强<sup>[9]</sup>。相反,如果把大量的人力、物力和财力花费到基金项目成果的评价和挖掘过程中,不仅会增加基金管理工作的压力,而且是另一种程度上的资源浪费。因此,本研究在构建基金项目高水平论文、高质量专利和经济效益的挖掘机制时,始终坚持以高效、科学的方式筛选出最有代表性的成果,以可行性和适用性作为重要标准,做到所构建的挖掘机制能够应用于基金成果管理的实践。

### 2.1 高水平论文挖掘标准

科技论文作为科研活动最基础的成果产出形式,在学术界广受关注。随着网络的发展和数据信息的共享,学者对论文的评价标准和方法也逐渐呈现出多元化的趋势。在评价标准方面, Garfield<sup>[10]</sup>于1955年提出论文被引频次指标具有科技评价的作用。不久后,美国科学情报信息研究所出版的《期刊引证报告》(InCites Journal Citation Reports, InCites-JCR)指出期刊影响因子指标也在一定程度上反映了论文的质量。此外,国内外研究者还基于引文网络、引文内容等构建论文的评价指标体系<sup>[11]</sup>,比如有学者通过测度论文的中心度、结构洞<sup>[12-14]</sup>等进行论文评价,也有学者利用论文元数据<sup>[15-17]</sup>,结合引用位置丰富论文的评价角度<sup>[18]</sup>。在关注论文学术影响力的同时,越来越多的研究者基于论文在社交媒体、学术网络等的使用量、浏览量、下载次数等衡量论文的社会影响力。随着论文评价标准的不断丰富,评价工具与方法也表现出多样性,包括层次分析法、主成分分析法、PageRank等<sup>[19]</sup>。

在对以往文献总结归纳的基础上,发现论文被引频次和期刊影响因子是最早被学者提出的评价标准,也得到了大多数学者的认可。论文被引频次指科学论文对文献的引用次数,是衡量一个国家科研文献被其他国家或机构的认可度的标志。影响因子既是测度期刊学术水平的主要标准,也是衡量论文质量的重要指标<sup>[20]</sup>。目前,已有诸多国外科研机构或者基金管理机构将上述论文评价标准应用于成果管理的实践中。例如,日本国立大学在第三方评价当中将学术期刊上的论文影响因子和论文引用指数作为成果评价的重要指标<sup>[21]</sup>;澳大利亚研究理事会在科研论文质量评价时也加入了引文分析和期刊影响因子两项指标,并将期刊论文影响因子以5%、20%、50%作为划分标准,评定为A\*、A、B、C四个等级<sup>[22]</sup>。参照国外成果管理和评估的经验,本文拟在联合基金结题项目成果库中筛选期刊影响因子前

10%同时被引频次前20%的文章作为高水平论文。此外,有学者指出基金项目绩效评价应遵循分类原则,针对不同科学部设计不同的评价指标<sup>[23]</sup>。同样地,由于联合基金项目的论文成果年限跨度长、涉及领域广,不同学科领域的期刊影响因子和被引频次指标存在显著差异,因此拟通过分年份、分科学部筛选联合基金结题项目中的高水平论文。

### 2.2 高质量专利挖掘标准

专利作为基金项目的重要成果之一,是“科学技术是第一生产力”的最直接体现。专利评估是实现知识产权产业化发展的重要环节。高质量的专利不仅是科研价值的表现,也是解决实际问题、促进社会生产的前提。虽然学术界还未构建出一套为大家公认的专利评价体系,但专利质量评价的基本准则是基本一致的,主要包括技术性评价标准、法律性评价标准和经济性评价标准。例如,李斌<sup>[24]</sup>采用技术性指标、经济性指标、社会性指标和市场性指标构建高校科技成果产业化筛选评价指标体系,并通过层次分析法(AHP法)对各指标赋予权重。吕晓蓉<sup>[25]</sup>从专利的技术价值、权利价值和市场价值方面考虑,构建了专利质量评级模型,并对1985—2013年获美国授权的专利进行了评价。许鑫等<sup>[26]</sup>提出了涵盖技术、经济、法律、战略四大层面的多维视角专利质量评价指标体系。此外,也有学者从数量维度、质量维度和价值维度构造专利价值评估体系,为专利的评价提供了新的理论尝试<sup>[27]</sup>。

在文献阅读的基础上,本文通过门户网站检索,查阅了地方网上技术市场和技术转移公共服务平台一些发展较成熟的技术评估方法,如浙江网上技术交易市场依据GBT 22900-2009《科学技术研究项目评价通则》构建了一套“标准化”+“专业化”的评价体系,加强对科技成果的评价<sup>[28]</sup>。基于以上分析并结合专家意见,本文从法律、经济和技术三个维度构建专利价值评估体系。其中,法律价值维度包括专利稳定性、专利侵权可判定性、专利有效性和专利自由度,技术价值维度包括技术先进性、技术发展趋势、适用范围、不可替代性和可实施性,经济价值维度包括市场应用情况和获益能力。拟选取专利评估排名前20%作为高质量专利。

由于法律价值和经济价值中的部分评价指标数据获取难度较大,例如权利要求数、专利市场需求度、专利技术利润率等。因此,本研究就目前我国在专利价值评估方面有着成熟做法的中国专利信息中心的专利价值评价服务平台进行了电话访谈和实地



调研。通过双方详细地磋商后,根据该平台现有的专利评估体系对已经构建的专利评估体系做出部分调整,并选用该平台分析软件作为高质量专利挖掘的主要工具。

### 2.3 典型经济效益挖掘方法

经济效益是项目成果转化与推广的一个重要体现,也是评价科学基金项目典型性及基金战略目标实现程度的重要指标。目前,已有研究对于科学基金成果产生的经济效益分析多采用典型案例研究或结题报告中的成果数据统计表中经济效益一栏。本研究在初步浏览若干份联合基金结题报告后发现,联合基金项目负责人对经济效益一栏的填写存在概念模糊、理解不一致等情况。同时,典型案例研究方法也无法应用于大批量的成果经济效益挖掘中。

基于上述分析,本文基于联合基金项目的结题报告,选用 Spyder 作为 python 集成开发环境,运用 jieba 分词词库,通过批量数据读写以及基于自定义词典的批量数据文本分词功能,实现对项目结题报告文本内容的挖掘,探索建立典型经济效益成果的分词词库。

为了进一步提升 python 数据批量读写的工作效率,本研究将项目结题报告文件格式由 PDF 转化为 txt 格式,在保留结题报告完整内容信息的基础上,有效地去除了数字水印的影响。考虑到指标体系构建的严谨性和科学性,本研究主要分为三个阶段来实现:

第一个阶段主要通过浏览最初从自然科学基金委获取的 26 份联合基金项目结题报告,初步确认了 23 个与经济效益有关的关键字段,并进行反向匹配以此检测所选经济效益字段的准确度。结果显示,已有的经济效益字段库可以选出具有经济效益的项目。但是,由于每个项目负责人写作习惯不尽相同,本研究设计经济效益词库无法达到百分之百的精确度,也会存在字段频次高但实际上并未产生经济效益的项目。针对上述问题,本研究进行了第二阶段的测试和完善。

第二个阶段主要是在首次尝试的基础上进一步完善经济效益关键词库并提高经济效益关键字段的准确度。为了解决这一问题,我们增加了 241 份结题报告,其中包括成果数据统计表中有经济效益填写的结题报告 98 份,无经济效益填写的结题报告 143 份。通过浏览随机增加的 241 份结题报告后,重新整理出一些新字段,对第一阶段的经济效益词库进行了扩充,增加了 57 个与经济效益有关的关键字段,共 70 个关键字段,进行了经济效益的再次挖

掘。此次结果显示,241 份结题报告中,存在经济效益描述的结题报告共 195 份。

第三阶段的主要工作是将经济效益词库中各关键词根据频次回溯到各个结题报告中以确认关键字段精确度。这一阶段主要是在第二阶段确定的 70 个经济效益字段和 195 份结题报告的基础上,对总计 13 650 个待确认数据即每一结题报告中的 70 个经济效益字段进行手工回溯以确认所选关键字段的精确度。经过人工查找与统计后,最终删除了 38 个精确度低于 61% 的关键字段,保留了 32 个精确度较高的经济效益字段作为联合基金结题项目经济效益筛选的最终词库(具体见图 1)。在所有经济效益的挖掘字段中,出现频次最高的为转让、产业化和示范,较准确地反映了基金项目成果的应用情况;其次是销售、产值和示范应用,出现频次均在 15 次以上。其他经济效益字段虽然出现频次有限,但在挖掘具有经济效益的项目时也具有一定的代表性。

## 3 联合基金项目典型成果挖掘分析

根据上述构建的基金项目典型成果挖掘标准和方法,本文对 2002—2019 年联合基金结题项目中的论文、专利和经济效益等典型成果进行识别,一方面有利于展示联合基金成立 20 年来的绩效和突出成果;另一方面将典型成果的挖掘结果和实际调研的案例相结合,进一步说明本文构建的基金项目典型成果挖掘机制的科学性和准确性。

### 3.1 高水平论文挖掘

#### (1) 基础数据获取和数据预处理

根据自然科学基金委提供的联合基金项目论文信息,共获得了 56 116 条论文原始数据,论文发表的年份主要分布在 2003—2020 年。经过数据初步清洗,去除重复错误数据,仅保留 SCI 检索的论文(本文以 SCI 论文为例进行分析),得到有效数据

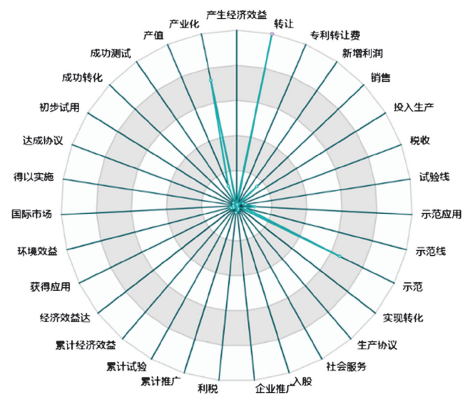


图 1 经济效益词频分布图

30 159 条。

### (2) 数据匹配分析

高水平论文挖掘需要获取每条论文数据的影响因子和被引频次,为此本研究引入第三方基础数据,即《期刊引证报告》和《科学引文索引》(Science Citation Index,SCI),该类数据库有比较公认的影响力,可以表明每年 SCI 检索期刊的影响因子情况和论文的引文索引情况。在此基础上,实现了论文成果列表数据的学术价值匹配分析工作,形成包含论文基础信息、SCI 影响因子和被引频次的标准化数据。

### (3) 高水平论文挖掘结果

根据 SCI 期刊影响因子和被引频次,本文筛选出联合基金结题项目所发表的高质量论文分别为 1 262 篇和 897 篇论文。基于前文构建的高水平论文挖掘标准,我们进一步识别了联合基金结题项目所发表的同时处于期刊影响因子前 10% 和被引频次前 20% 的论文,最终确定联合基金项目高水平论文产出为 315 篇,约占 29 311 篇有效数据的 1.07%,具体分布如图 2 所示。

从图 2 可以看出,高水平论文的产出主要集中在 2009—2020 年。在 2009—2015 年,高水平论文总数总体呈上升趋势,联合基金资助成效显著。2018—2020 年,高水平论文产出数下降较大,主要原因是截至统计时间,联合基金项目结题不完全,导致论文总数与前几年相差较大。从八大科学部分布来看,高水平论文主要集中于工程与材料科学部、数理科学部和化学科学部,管理科学部论文产出数相对较少。

高水平论文的挖掘不仅仅有助于了解联合基金成立近 20 年来的资助成效,更有利于促进科学知识的积累和成果的宣传与展示。高水平论文在一定程度上代表了某领域的学术前沿,更具创新性和突破

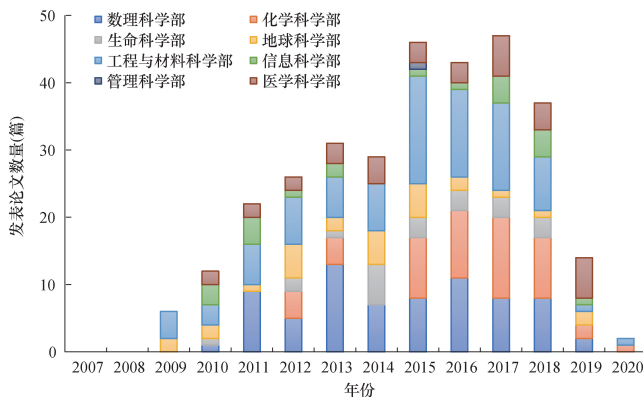


图 2 联合基金高水平论文分布情况

性。例如,赖剑煌教授团队承担的“面向视频监控的生物特征识别关键理论与技术研究(U0835005)”发表 SCI 检索论文 40 篇,基于敏感度在线学习算法的聚类算法获得 ICDM 2010 的 Honorable Mention Award 奖;张军教授承担“数字家庭智能化的基础算法理论与关键技术的研究(U0835002)”时,发表的论文“Adaptive Particle Swarm Optimization”入选 ESI 全球高被引 1% 论文;张道远教授先后承担联合基金项目“齿肋赤藓复水过程的分子机制—基于基因和蛋白质高通量表达的研究(U1170304)”和“基于转录组的荒漠藓类齿肋赤藓耐干相关基因深度挖掘研究(U1403302)”,发表 SCI 论文 35 篇,总被引频次达 289 次,被期刊评为高访问率论文(Highly Accessed Articles)。

## 3.2 高质量专利挖掘

### (1) 基础数据获取和数据预处理

根据自然科学基金委提供的联合基金结题项目 13 804 条专利成果数据,在剔除专利号信息有误、数据缺失等相关专利后,剩余的有效数据为 9 238 条。由于专利是否授权极大地影响了评估结果,故本研究仅对截至结题时已获得授权的 6 199 项专利进行挖掘与评估。此外,由于外观专利可获取信息较少,无法进行评价,本文的挖掘仅涉及发明专利和实用新型专利。基于联合基金结题项目已授权专利数据,在剔除了包括专利号格式错误和外观型专利等 2 434 条无效数据后,共剩余 3 765 条有效专利数据。

### (2) 高质量专利挖掘结果

根据前述专利挖掘标准,经过专利价值评价服务平台软件的分析,得到每一项专利的价值评价总分,并选取专利总分前 20% 作为高质量专利,最终筛选出高质量专利 753 项。高质量专利排名前 20% 情况如图 3 所示。

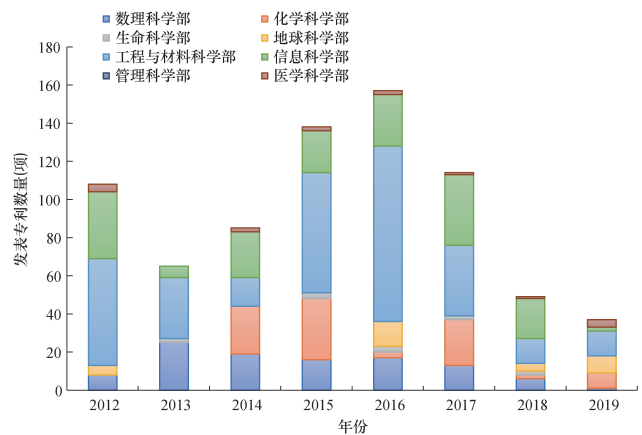


图 3 联合基金高质量专利分布情况

从图3可以看出,高质量专利主要分布在2012—2019年。2013—2016年期间,高质量专利成果数呈现逐年递增的趋势,这表明,随着我国知识产权保护力度的加大,联合资金资助项目成果产出的质量也在不断提升。2017—2019年,高质量专利成果数有所下降,这可能是由于部分专利转化应用需要较长的时间,在截止统计时间内未能完全发挥其社会经济效应,因此影响其价值评分。从科学部分布来看,工程与材料科学部的高质量专利产出数最多,占到了总数的42.6%。其次是信息科学部和数理科学部。生命科学部、医学科学部和管理科学部高质量专利产出数较少。

结合高质量专利评分表可以发现,经济价值在总分中的占比最高,其次是技术价值,这在一定程度上说明了高质量专利具有更广阔的应用前景和更高的创新力。例如,汤勇教授承担的“面向光电领域的微热管制造基础研究(U0834002)”项目已授权专利获得广东省科技进步奖一等奖、二等奖以及国家专利优秀奖;方攸同完成的“高速列车永磁牵引系统设计理论与方法研究(U1434202)”项目专利成果获国家技术发明奖二等奖1项、中国铁道学会科技进步一等奖1项,解决了热管理、电磁兼容、可靠性设计等难题,成功研发了国内首台高速列车永磁牵引电机;罗笑南教授的“可视媒体智能处理理论和传输方法研究(U0735001)”项目专利成果共开发出170余种数字家庭产品及八大类互动应用服务,带动了产业示范基地孵化数字家庭相关企业200余家企业的发展。

### 3.3 经济效益挖掘

由于基金项目经济效益挖掘体系是以部分联合基金结题报告为依据构建的,且由于客观因素本研究无法根据前文建立的挖掘体系实现联合基金全样本经济效益的统计和分析,因此,本文根据上述最终确定的32个经济效益词库的字段,从267份结题报告中挖掘出了77个产生经济效益的项目,占比为28.83%,准确率高达70%以上。可以看出,截至项目结题,联合基金项目的成果实现了部分前瞻性基础研究、引领性原创成果的有效转化应用,对社会经济发展起到了推动作用。例如,姚红宇教授承担的“机场跑道端飞机拦阻系统的动力学模型与实验验证项目”(U1233203)的研究成果已成功转化为特性材料拦阻系统(EMAS)产品,在云南省腾冲机场实现了首次工程应用,创造了高达400万元的产值,

在国内民用、军用机场以及国外机场具有广阔的应用前景<sup>[29]</sup>;陈曦教授项目团队完成“专用化学品联合装置的分布式协同调控与智能集成优化(U1509209)”后,浙江大学与浙江龙盛集团成立了“浙江大学—安诺智能制造联合研发中心”,企业投入一千万元进行“企业数字化及智能制造”研发;戴李宗教授承担的“基于超分子技术的多元素协同有机/无机纳米阻燃体系的可控构筑与相关基础科学问题研究(U1205113)”的研究成果——P、N协同阻燃聚醚多元醇已在涓洲湾氯碱工业有限公司实现产业化,为企业实现销售收入500万元,新增利税85万元,促进了社会经济的发展<sup>[30]</sup>。

## 4 结论与建议

在当今大数据时代,“保障科技成果供给质量,挖掘具有转化潜力的优秀科技成果,提高科技成果落地转化效率”已成为基金管理的主要工作之一<sup>[31]</sup>。这就意味着成果信息的科学管理在国家科技创新和经济社会发展中发挥着越来越重要的作用,基金项目典型成果挖掘机制的构建已成为新时期基金管理的必然趋势和要求。充分利用互联网技术、信息技术和大数据分析,设计合理的成果筛选标准与挖掘方法,从海量的成果库中挖掘出具有代表性的典型成果,既有助于完善自然科学基金委的管理制度和流程,又推动了基金成果的共享、传播及应用贯通。

本研究构建的基金成果挖掘机制的目标实现并不是一蹴而就的,需要以精准的数据收集和分类为基础。因此,本文提出如下建议:

### 4.1 规范项目结题报告填写,提高成果数据收集精度

高质量的数据收集既是成果管理的重要组成部分,也是典型成果挖掘机制实现的关键前提。在联合基金项目典型成果挖掘过程中,由于基金项目成果的原始数据存储信息的方式和填写准确性的限制,进行数据清洗的工作耗时耗力。基金管理部门需要规范基金项目结题报告的填写,给出明确的成果统计标准,优化成果收集方法,提升成果管理精确度,明确基金项目成果管理工作是发挥科学基金积极导向作用的重要基础。同时,项目负责人和依托单位应当积极配合自然科学基金委成果管理工作,做好项目成果原始记录的采集和保存工作,确保项目成果报告中科学数据的系统性、完整性和准确性。



## 4.2 依托大数据网络平台,加强成果分类管理

加强成果分类管理是科技成果推广和应用和产业化的迫切需求<sup>[32]</sup>,而成果挖掘机制是识别和发现可转化成果的重要前提。在基金项目成果数据收集的基础上,可以借鉴发达国家的经验,从制定政策和法律入手,注重成果数据的分类管理和系统集成,依托大数据技术,对基金成果准确分类并精准刻画。通过高标准的信息采集、信息编码、信息储存,有效提升成果的价值利用,进而为基金项目典型成果挖掘的实现奠定基础。

## 4.3 推动信息数据库联动,提高成果挖掘效率

根据基金项目典型成果的不同,挖掘机制所需的筛选标准也各异。而在构建的典型成果筛选标准基础上,需要利用外部数据库才能够实现典型成果的挖掘,在今后的成果管理工作中,自然科学基金委可以联合有关部门充分利用互联网和大数据分析技术的优势,推动各数据库之间的数据联动,提高典型成果筛选指标结果获取的便捷性和准确性,节省成果管理时间,提高成果挖掘效率,从而推动成果的共享、传播和应用贯通。

本文从优化基金项目成果管理的角度出发,构建的基金项目典型成果挖掘机制与以往研究最大的区别就是在已有成果数据库的基础上对不同类型的成果构建不同的评价标准和挖掘方法,进而筛选识别出典型成果,加强该类成果的宣传、展示、共享与传播、转化与应用,不但能充分发挥基金成果服务国家和社会的效能,也能不断优化基金的资助管理。本文虽然以联合基金结题项目成果为研究对象,但构建的典型成果挖掘机制对各种类型基金的成果具有普适性,对于各类基金项目的管理有一定的可推广性。

## 参 考 文 献

- [1] 国家自然科学基金委员会. 自然科学基金委会同有关部门关于印发《新形势下加强基础研究若干重点举措》的通知. (2020-05-11)/[2021-06-17]. <http://www.nsf.gov.cn/publish/portal0/tab434/info77851.htm>.
- [2] 李静海. 国家自然科学基金支持我国基础研究的回顾与展望. (2018-05-10)/[2021-06-17]. <http://www.nsf.gov.cn/publish/portal0/tab440/info73751.htm>.
- [3] 国家自然科学基金委员会. 国家自然科学基金“十三五”发展规划. (2015-09-17)/[2021-06-17]. <http://www.nsf.gov.cn/publish/portal0/xgk/043/info72249.htm>.
- [4] 国家自然科学基金委员会. 国家自然科学基金资助项目研究成果管理办法. (2015-09-08)/[2021-06-17]. <http://www.nsf.gov.cn/publish/portal0/tab475/info73668.htm>.
- [5] 冯勇, 刘权. 国家自然科学基金资助项目研究成果的管理实践与思考. 中国科学基金, 2017, 31(5): 466—470.
- [6] 国家自然科学基金委员会. 国家自然科学基金资助项目优秀成果选编一六. 北京: 科学出版社, 2016.
- [7] 朱蔚彤, 孟宪平. 国家自然科学基金联合基金设立与资助管理机制探讨. 中国科学基金, 2012, 26(1): 34—37.
- [8] 刘华海. 科研项目绩效评价模型和指标体系的构建. 科研管理, 2016, 37(S1): 19—24.
- [9] 王力. 自然科学基金项目绩效评价指标体系构建. 中国科技信息, 2017(Z1): 126—128, 130.
- [10] Garfield E. A new dimension in documentation through association of ideas. *Science*, 1955, 122(3159): 108—111.
- [11] 刘盛博, 王博, 丁堃. 科技论文评价研究综述. 情报理论与实践, 2016, 39(6): 126—130, 138.
- [12] 宋歌. 社会网络分析在引文评价中的应用研究. 图书情报工作, 2010, 54(14): 16—19, 115.
- [13] Burt RS. *Structural holes: the social structure of competition*. Cambridge: Harvard University Press, 2009.
- [14] Chen CM. Predictive effects of structural variation on citation counts. *Journal of the American Society for Information Science and Technology*, 2012, 63(3): 431—449.
- [15] Fu LD, Aliferis C. Models for predicting and explaining citation count of biomedical articles. *AMIA Annual symposium proceedings*. American Medical Informatics Association, 2008, 6: 222.
- [16] Yogatama D, Heilman M, O'Connor B, et al. Predicting a scientific community's response to an article. *Proceedings of the 2011 conference on empirical methods in natural language processing*, 2011: 594—604.
- [17] Livne A, Adar E, Teevan J, et al. Predicting citation counts using text and graph mining. *Proceeding the iConference 2013 workshop on computational scientometrics: Theory and applications*, 2013: 1—4.
- [18] Sombatsompop N, Kositchaiyong A, Markpin T, et al. Scientific evaluations of citation quality of international research articles in the SCI database: Thailand case study. *Scientometrics*, 2006, 66(3): 521—535.
- [19] 高锡荣, 杨娜. 基于社会网络分析方法的论文评价指标体系构建. 情报科学, 2017, 35(4): 97—102, 144.
- [20] 林德明, 姜磊. 科技论文评价体系研究. *科学学与科学技术管理*, 2012, 33(10): 11—17.
- [21] 聂虹, 魏翔, 佟方. 英国、美国、日本科技成果评价比较及启示. *中国矿业*, 2017, 26(S2): 45—48.

- [22] Australian Research Council. ERA 2018 evaluation handbook. [2021-01-28]. <https://www.arc.gov.au/excellence-research-australia/key-documents>.
- [23] 罗骏, 周小丁, 黄云生. 青年科学基金项目绩效评价指标体系研究与探讨. 中国科学基金, 2016, 30(4): 336—339.
- [24] 李斌. 高校科技成果产业化筛选评价指标探讨. 中国高校科技与产业化, 2010(8): 46—48.
- [25] 吕晓蓉. 专利价值评估指标体系与专利技术质量评价实证研究. 科技进步与对策, 2014, 31(20): 113—116.
- [26] 许鑫, 赵文华, 姚占雷. 多维视角的高质量专利识别及其应用研究. 现代情报, 2019, 39(11): 13—22, 45.
- [27] 李振亚, 孟凡生, 曹霞. 专利三维评价指标体系研究. 情报科学, 2010, 28(10): 1569—1573.
- [28] 中关村巨加值科技评价研究院. 科技成果(专利技术)第三方评价系统. (2020-05-11)/[2021-06-17]. <http://www.tpste.cn/web/home/kjpg.jsp>.
- [29] 中国民航科学技术研究院. 航科院 EMAS 基金重点项目通过结题验收. (2017-05-17)/[2021-06-17]. [http://www.castc.org.cn/dtxw/kjdt/201705/t20170517\\_6387.html](http://www.castc.org.cn/dtxw/kjdt/201705/t20170517_6387.html).
- [30] 彭海媛, 杨志鹏, 黄初升. 促进海峡两岸科技合作联合基金项目资助绩效分析与对策研究. 中国科学基金, 2019, 33(1): 71—78.
- [31] 科技部. 2019 年科技成果直通车工作推进会在京召开. (2019-06-28)/[2021-06-17]. [http://www.most.gov.cn/kjbgz/201906/t20190628\\_147331.htm](http://www.most.gov.cn/kjbgz/201906/t20190628_147331.htm), 2019-06-28.
- [32] 科技部. 《关于加强国家科技计划成果管理的暂行规定》. (2003-07-09)/[2021-06-17]. [http://www.most.gov.cn/xinwzx/xwzx/xwfb/200312/t20031214\\_10452.htm](http://www.most.gov.cn/xinwzx/xwzx/xwfb/200312/t20031214_10452.htm).

## Construction and Application of the Typical Achievements Mining Mechanism of the Joint Fund Projects Based on Big Data Analysis Technology

Wu Weihong<sup>1</sup> Zhang Lina<sup>1</sup> Li Zhilan<sup>2</sup> Zhang Aimei<sup>1\*</sup> Liu Zhonghua<sup>1</sup>

1. School of Economics and Management, Beijing University of Chemical Technology, Beijing 100029

2. Bureau of Planning, National Natural Science Foundation of China, Beijing 100083

**Abstract** How to use big data analysis technology to mine representative typical achievements from massive National Natural Science Foundation of China(NSFC) funded achievements is the key to show the performance of NSFC funded and promote the application of excellent achievements. In this paper, data mining is used to build the mining mechanism of typical achievements of fund projects in the big data environment. The year of paper and paper information are used as matching search points screen high-quality papers by criteria. From the three dimensions of law, economy and technology, the patent value evaluation system including 11 indicators is constructed, and the relevant information of the third-party patent value evaluation service platform is used to screen high-quality patents and the thesaurus of typical economic benefits is established to realize the mining of typical economic benefits. Taking the achievement information of the projects concluded by the National Natural Science Foundation of China from 2002 to 2019 as the object, this paper analyzes the typical achievements. The results show that the above typical results mining mechanism can select high-level papers and patents by year and department; the accuracy rate of economic benefit mining results is highly consistent with the actual application benefits of the project.

**Keywords** typical results; mining methods; big data analysis technology; joint fund

(责任编辑 刘敏)

\* Corresponding Author, Email: zhangam@mail.buct.edu.cn