



开放科学
(资源服务)
标识码
(OSID)

美国科学基金资助热点布局及对我国海洋领域规划的借鉴

曹京晶 王莹 王烨 陈皓 张雪 张晓夏 朱晓晨

科学技术部科技人才交流开发服务中心 北京 100045

摘要: [目的/意义] 从世界强国的发展经历看, 要想崛起于世界, 必先强盛于海洋。我国海洋科技起步较晚, 海洋科技创新能力基本处于全球海洋科技竞争格局中的第二梯队, 与第一梯队还存在一定的差距。本文旨在通过分析美国国家科学基金会(NSF)对海洋领域的资助情况, 对海洋领域的热点方向及布局开展分析研究, 以期对我国海洋领域的规划布局提供参考。[方法/过程] 本文主要通过对NSF近四十年来项目资助数量、资助金额、资助时间等数据维度进行简要说明, 并通过NSF在海洋领域的样本数据, 尝试将LDA模型与BERT模型相结合, 首先采用基于Bert的模型将文本向量化, 然后基于LDA(Latent Dirichlet Allocation)算法对文本聚类, 以解决提高聚类准确度的文本自动分类任务。尝试分析出海洋领域的重点关键技术及研究热点方向, 为我国未来对海洋领域的规划及布局提供支撑。[结果/结论] 通过NSF在海洋领域的样本数据, 尝试分析出海洋领域的重点关键技术及研究热点方向, 为我国未来对海洋领域的规划及布局提供参考。

关键词: 文本挖掘; NSF; 海洋科学; 前沿识别

中图分类号: G35; TP391

基金项目 科技创新战略研究专项(ZLY202111)。

作者简介 曹京晶(1987-), 硕士研究生, 助理研究员, 研究方向为科技人才政策、科技重点领域人才分布研究等; 王莹(1985-), 副研究员, 研究方向为科技人才政策、科技人才评价研究等, 通讯作者, E-mail: wangy@sttc.net.cn; 王烨(1981-), 工学博士, 副研究员, 研究方向为科技人才评价, 数据挖掘分析等; 陈皓(1995-), 管理学硕士, 研究实习生。研究方向为科技人才信息获取及分析等; 张雪(1987-), 硕士研究生, 助理研究员, 研究方向为科技人才政策、科技重点领域人才分布研究等; 张晓夏(1988-), 助理研究员, 研究方向为科技人才政策、数据挖掘分析等; 朱晓晨(1990-), 实习研究员, 研究方向为数据挖掘、数据模型等。

引用格式 曹京晶, 王莹, 王烨, 等. 美国科学基金资助热点布局及对我国海洋领域规划的借鉴[J]. 情报工程, 2021, 7(5): 62-74.

Distribution of NSF-funded Projects and its Referential Experience to Marine Research Planning of China

CAO Jingjing WANG Ying WANG Ye CHEN Hao ZHANG Xue ZHANG Xiaoxia ZHU Xiaochen

Exchange, Development & Service Center for Science & Technology Talents The Ministry of Science & Technology, Beijing 100045, China

Abstract: [Objective/ Significance] Judging from the development experience of world powers, if they want to rise in the world, they must first become stronger in the ocean. my country's marine science and technology started relatively late, and the innovation capability of marine science and technology is basically in the second echelon of the global marine science and technology competition, and there is still a certain gap between it and the first echelon. The purpose of this article is to analyze the US NSF's funding for the marine field, and conduct analysis and research on the hotspot direction and layout of the marine field, in order to provide a reference for the planning and layout of my country's marine field. [Methods/Process] This paper mainly introduces the data dimensions of NSF's project funding amount, funding amount, funding time and other data in the past 40 years, and tries to combine the LDA model with the BERT model through the sample data of the NSF in the marine field. , Firstly use the Bert-based model to vectorize the text, and then cluster the text based on the LDA (Latent Dirichlet Allocation) algorithm to solve the task of automatic text classification to improve the accuracy of clustering. Try to analyze the key and key technologies and research hotspots in the marine field to provide support for my country's future planning and layout in the marine field. [Results/Conclusions] Through the sample data of NSF in the marine field, try to analyze the key key technologies and research hotspots in the marine field, so as to provide reference for my country's future planning and layout of the marine field.

Keywords: Text Mining; NSF; marine science; frontier identification

引言

从世界强国的发展经历看,要想崛起于世界,必先强盛于海洋。进入 21 世纪,海洋愈加成为国际社会关注的焦点,在全球竞争中的战略地位日趋突出。海洋上的政治、军事、经济、科技的现实竞争与战略博弈呈愈演愈烈趋势,争夺的焦点是控制更广阔的海洋,进而掌握更多的资源,获得更大生存发展空间,深海、极地、太空、网络成为战略新疆域。新时代海洋科技发展,正迎来“从水面到水下、从浅海到深海、从近海到远海、从机械化到智能化”的飞速跨越。

相较于美国、英国、日本、俄罗斯等传统海洋强国,我国海洋科技起步较晚,但近年来发展很快,已经成为世界海洋科技网络中的重要力量,海洋科技创新能力基本处于全球海洋科技竞争格局中的第二梯队,与第一梯队还存在一定的差距。本文旨在通过分析美国国家科学基金会对海洋领域的资助情况,对海洋领域的热点方向及布局开展分析研究,以期对我国海洋领域的规划布局提供参考。

美国国家科学基金会(National Science Foundation, United States, 以下简称 NSF)是美国独立的联邦机构,由美国国会于 1950 年创

立。该机构支持除医学领域外所有基础科学、工程学领域的研究和教育（负责医学的同类机构为国家卫生研究院），在美国近代科学技术发展过程中发挥了重要作用^[1]。NSF 重点支持基础科学研究，与我国自然科学基金在资助方向上最为接近。

樊春良等^[2]在研究美国国家科学基金中考察了融合研究的内涵和特点，以及美国国家科学基金会 (NSF) 融合研究的概念；根据 NSF 十大投入计划和新的战略规划 (2018—2022 年)，指出发展融合研究在 NSF 中的战略地位；分析 NSF 支持融合研究的基础和融合项目的特点，并对第一批融合项目的实施作了介绍。该文章的研究思路及数据获取方式为本文的研究提供了参考思路。Greengrove C et^[3]教授在美国自然科学基金关于海洋领域的研究部署具有较为深入的研究，研究结果表明美国 NSF 重点关注海洋深海领域探测、地形地貌研究以及资源勘探等领域，为本文在后续的实证研究部分提供了重点领域方向的指导。

本文主要通过对 NSF 近四十年来项目资助数量、资助金额、资助时间等数据维度进行简要说明，并通过 NSF 在海洋领域的样本数据，尝试分析出海洋领域的重点关键技术及研究热点方向，为我国未来对海洋领域的规划及布局提供支撑。

1 数据来源与分析

1.1 数据获取来源

从美国国家科学基金会获取自 1980—2019

年之间近 40 年的已公开基金资助项目数据，并对获取的数据根据时间、学科进行切片处理，形成基于时序标签以及国标学科的数据集合。

1.2 数据预处理与清洗

在获取 NSF 资助数据后，对项目名称、项目单位或项目摘要中的文本数据进行纯文本格式转换，对项目资金中的金额为零的数据进行剔除，对标点符号以及科技文献中无意义的高频词进行过滤，以达到提升准确度的目标。

1.3 基于短文本聚类模型 BertPlusLDA 的主题识别方法

在自然语音处理领域，近年来深度学习取得了迅速发展，不少专家学者尝试了多种文本特征提取算法。Zhang Y et^[4]介绍了一种基于长短时记忆 (LSTM) 递归神经网络的预测模型，该模型以语言文本特征为输入，将输出设置为固定的指标判断结果，具有较好的预测准确率。王瑞等^[5]提出一种基于 Labeled-LDA 模型的文本特征提取方法，根据 TF-IDF 算法实现对文本类别的关键词进行提取，通过文本 simhash 算法对提取出的主题词与关键词进行相似度计算，找到文本隐含主题类别并提取特征词。张卫卫等^[6]提出一种基于 Word2Vec 和 LDA 主题模型的 Web 服务聚类方法。该方法首先将 Wikipedia 语料库作为扩充源，使用 word2vec 对 Web 服务描述文档内容进行扩充，再将扩充后的描述文档利用主题模型进行特征建模，将短文本主题建模转化为长文本主题建模，更准确地实现服务内容主题表达。严佩敏等^[7]在研究中发现 BERT 语言可以更好地应用于中文文

本分类任务中。赵璐偲等^[8]将基于 BERT 特征的神经网络方法应用在了中文电子病历输入推荐中。

因此本文在研究美国 NSF 的项目部署及关键词统计时,尝试将 LDA 模型与 BERT 模型相结合,首先采用基于 Bert 的模型将文本向量化,然后基于 LDA (Latent Dirichlet Allocation) 算法对文本聚类,以解决提高聚类准确度的文本自动分类任务。

(1) BERT 步骤

BERT 的网络架构使用的是多层 Transformer 结构,其最大的特点是抛弃了传统的 RNN 和 CNN,通过 Attention 机制将任意位置的两个单词的距离转换成 1,有效的解决了 NLP 中棘手的长期依赖问题。Transformer 的结构在 NLP 领域中已经得到了广泛应用。

由于 NSF 数据源中的文本数据都是短文本数据,因此无论是摘要或者是关键词,都非常适合对 Bert 进行训练。与此同时,我们认为 NSF 项目名称、项目单位、摘要、关键字等都是高浓缩的关键信息,且文本长度可控,同时噪音比较低(不规则表示、错别字等),文本信息浓缩度高,噪音低也适合采用 Bert 架构^[9]。在 BERT 步骤的主要工作包括利用 NSF 数据源中的 NSF 项目名称、项目单位、摘要、关键字等短文本训练出最优模型。并基于此模型对文本进行嵌入,并将嵌入结果最为 LDA 步骤的输入。

(2) LDA 步骤

贝叶斯模型,是一种具有良好的统计性质的生成模型。LDA 模型的出现解决了传统分析方法中“维数灾难”的问题,为文本数据挖掘

提供了“主题”思路。此外,该方法具有完整的针对参数估计的算法,因而在文本分类、文本聚类、情感分析、垃圾邮件过滤、搜索引擎等方面得到广泛应用,并不断得到发展和完善。基于 Bert 对文本信息的嵌入,将文本信息作为特征输入到 LDA 中。

(3) BERT 对 LDA 的互补

LDA 在词和文档之间加入主题的概念,虽然可以较好的解决一词多义和多次一义的问题。但由于基于传统统计与贝叶斯的数学方式,因此在理解词的上下文含义和相关性还存在差距。而基于 Bert 的词嵌入策略可以对此缺点进行弥补。

LDA 用于单文档往往效果不明显,Bert 则恰恰适合短文本。因此我们采用 Bert 提取短文本之间的关联信息。而采用短文档“链接”成长文档的机制使得词被分配给某个主题的次数和一个主题包括的词数目在运算层面进行收敛。我们的做法是把每个 NSF 每个项目属于不同特征下的所有文本进行合并,“链接”成一个较长的文本作为 LDA 的聚类数据中的一个样本。

2 BertPlusLDA 方法在 NSF 项目资助情况的应用

2.1 NSF一级学科资助情况

NSF 于 1950 年由美国国会创立,是独立的联邦机构,其宗旨在于促进全美科学进步,促进国家向健康、繁荣和富强方向不断发展,确保国防安全等。在过去的几十年里,NSF 支持的研究人员已经获得了大约 223 项诺贝尔奖

以及其他不胜枚举的荣誉。可见 NSF 的成立对美国各项基础研究、学术人才培养发挥了重要

作用。分析其总体情况,对我国基础研究的发展及人才培养也会有一定的启示作用^[10]。

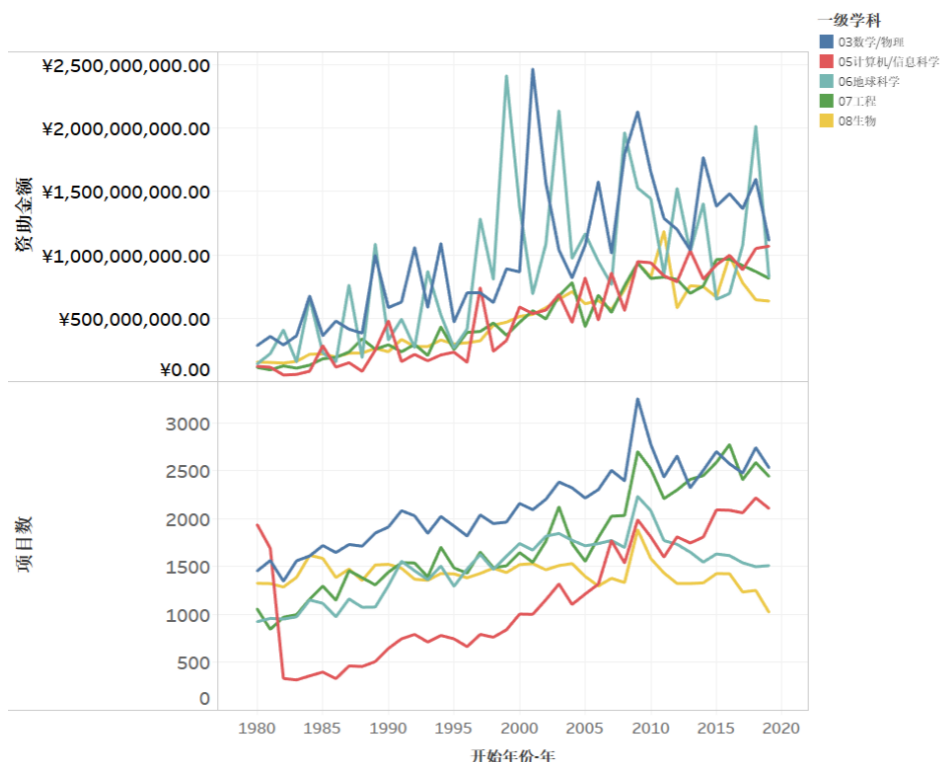


图1 各学科40年来资助项目及数量对比

在各学科总体资助项目和数量上,通过图1可以看出,数学/物理学科的数值最高,震荡最为激烈,工程学科的数值最低,震荡幅度最为平缓,即越是偏基础研究,数值震荡越大,越是偏工程应用,数值震荡越小。数学/物理是 NSF 的重点资助领域,项目资助数量及项目

资助金额都优于其他学科。工程学科的资助项目数量较多,但就单独项目的支持经费来说,地球科学、数学/物理、计算机/信息科学排名前三,工程排名较靠后。可见越偏基础类的研究,资助金额越多,越偏工程应用类的研究,资助金额越少。

表1 各学科受资助金额情况

序号	一级学科	总金额	项目数量	项目平均金额
1	数学/物理	401亿6054万	85289	47.09万
2	地球科学	358亿124万	59592	60.08万
3	计算机/信息科学	200亿4237万	47004	42.64万
4	工程	198亿3247万	70386	28.18万
5	生物	195亿1718万	56753	34.39万

从上表 1 可以看出,受资助金额和项目数量最多的均是数理学科,数理学科多为基础研究项目,这也与 NSF 重视资助基础学科项目相吻合。项目平均金额最多的是地球科学,地球科学受资助项目研究周期普遍较长,资助金额较高。

2.2 NSF 二级学科受资助时间特点分析

鉴于一级学科范围较大,项目具体数据量庞杂,对 NSF 的二级学科数据进行统计,对应的各学科项目数和项目资助金额如下图 2。

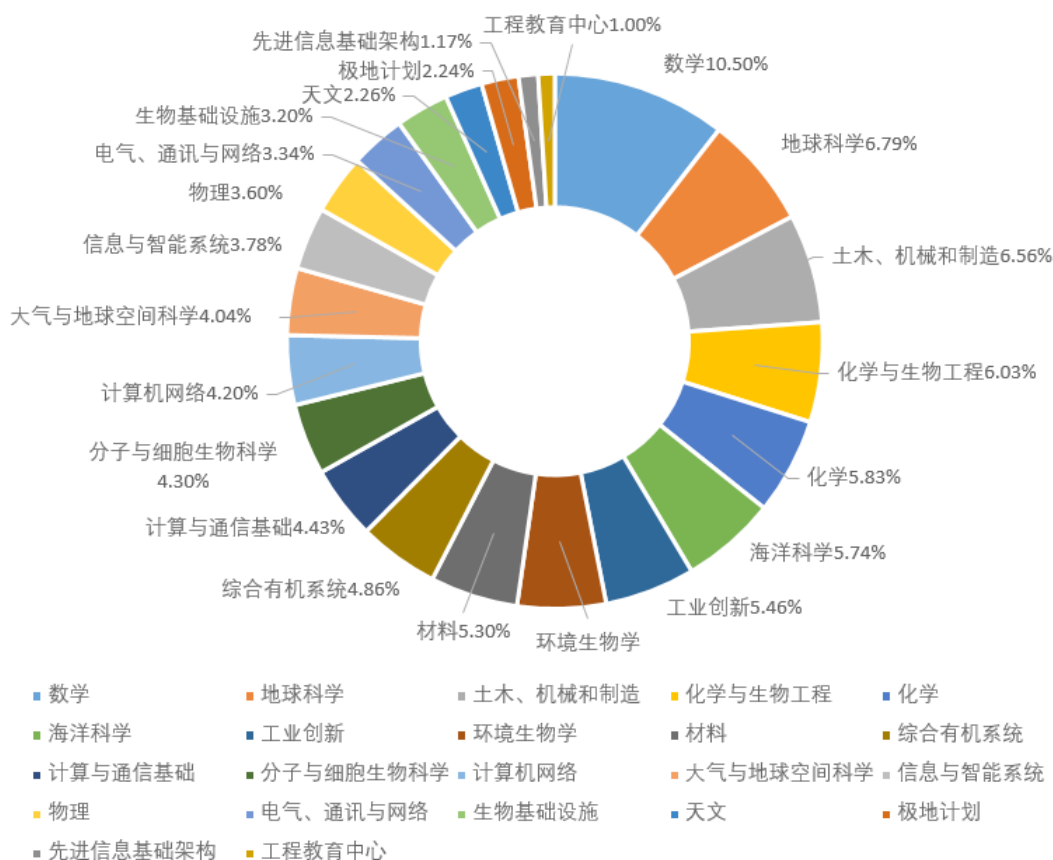


图 2 项目数量的二级学科分布

从项目总数上可以看出,近四十年来,NSF 对二级学科数学资助的项目最多,占所有学科的 10.5%。地球科学受资助项目数量第二,占所有学科的 6.79%,土木、机械和制造受资助项目数量第三,占所有学科的 6.56%。

从图 3 可以看出,近四十年来,NSF 对二级学科海洋科学资助的经费最多,占总经费的

9.41%,其次是天文,受资助经费占总经费的 7.16%,材料学受资助经费占总经费的 6.84%,位列第三。

从表 2 可以看出,二级学科先进信息基础架构平均资助金额最高,第二为天文,第三为极地计划。本文对资助经费排名第一的海洋科学领域进行实证研究,试图有效识别出该领域的前沿科学问题^[11]。

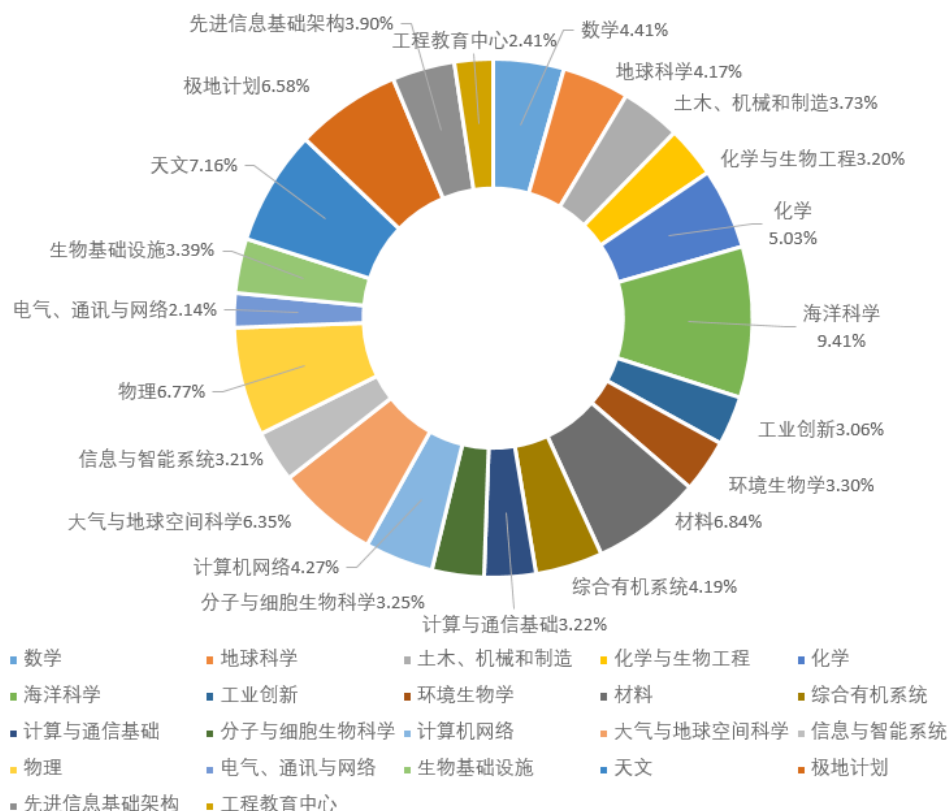


图3 项目经费的学科分布

表2 项目平均资助金额

二级学科	项目 总数	项目总额 (亿)	平均金额 (万)
海洋科学	17806	124.91	70.15
天文	6992	95.05	135.94
材料	16425	90.86	55.32
物理	11161	89.84	80.49
极地计划	6945	87.39	125.83
大气与地球空间科学	12532	84.37	67.33
化学	18067	66.78	36.96
数学	32567	58.49	17.96
计算机网络	13030	56.740	43.55
综合有机系统	15077	55.69	36.93
地球科学	21057	55.38	26.29
先进信息基础架构	3642	51.74	142.06
土木、机械和制造	17806	49.58	24.36
生物基础设施	6992	45.02	45.36
环境生物学	16425	43.87	26.48

2.3 算法实例研究

如上述分析可看出，NSF 对二级学科海洋科学资助的经费最多，可见美国对海洋领域的投入及重视。相较美国，我国在党的十八大作出了建设海洋强国的重大部署，党的十九大进一步明确，坚持陆海统筹，加快建设海洋强国。都说明我国对海洋领域有了越来越多的投入与布局，分析 NSF 对于海洋学科的资助方向及热点布局，对于我国海洋领域发展具有指导意义^[12,13]。因此，本文以海洋学科研究领域为例，选取海洋学科作为范本进行进一步分析，对其未来的前沿技术方向及热点开展识别研究。

(1) 确定数据集

从 NSF 项目数据库以 1980–2019 年为时间

范围,检索范围限定在海洋学科,获取资助项目的题目、摘要、资助金额、资助单位等数据,共计 495018 条数据。对基金项目数据进行标点符号、非相关高频词、无用词、无意词等数据预处理工作,最终对 342610 条数据进行分析。

(2) BERT 算法

数据集共计有九大类,涵盖海洋二级学科。对应的样本数依次为 23251、46859、39782、12809、61248、23890、47612、54378、32781。训练的过程记录如表 3,以 5 轮循环为例,得出训练过程的主要指标,具体分析如下。

(3) LDA 困惑度计算结果

根据困惑度计算结果,设定 NSF 基金项目数据数据集,对每个单词计算困惑度。

表 3 BertPlus 训练过程统计表

Epoch	Loss	F1	Precision	Recall	Accuracy
1	0.1254	0.351	0.210	0.326	0.221
2	0.1102	0.415	0.321	0.584	0.385
3	0.0842	0.512	0.498	0.662	0.421
4	0.0921	0.625	0.581	0.681	0.558
5	0.0512	0.771	0.625	0.745	0.612

$$p(\tilde{w}|M)=\prod_{m=1}^M p(\tilde{w}_m|M)^{-\frac{1}{N}}=\exp-\frac{\sum_{m=1}^M \log p(\tilde{w}_m|M)}{\sum_{m=1}^M N_m} \quad (1)$$

图 4 以 NSF 中资助的题为“海洋鱼类远程跟踪声学数据存储标签的研制”的项目为例进行计算。

根据公式(1)计算困惑度,以主题数 7 和 9 进行实验,结果如图 5 所示。

1 These animals migrate over great distances in the course of their life cycles however their specific migration patterns are difficult to ascertain.
2 Specifically the PIs propose to fabricate a DST that will listen for and record the exact arrival times of signals emitted from moored sound sources.
3 To obtain additional habitat information the tag will also periodically record pressure and temperature.
4 The devices having a design lifetime of two years would be attached to or implanted in fish using techniques developed during studies using conventional DSTs.
5 The data will be retrieved from the fish upon capture.
6 In case of an implant an external ribbon or marker would alert fishermen to the presence of the tag.
7 Acoustic detection ranges on the order of 100 to 1000 km could be obtained with a resolution of a few kilometers depending upon the tracking conditions for a given species.
8 The PIs envision the tags to consist of a small hydrophone that will also serve as the housing for the miniature circuit board timing crystal pressure sensor and battery.
9 Prior to deployment the tag will be programmed from a host PC via a serial communication interface.
10 This work evolves from an earlier development of a temperature-only DST.
11 The PIs have already contacted external vendors to ensure that there is an eventual transition from a prototype tag to a mass produced device.

图 4 NSF 资助项目示例数据

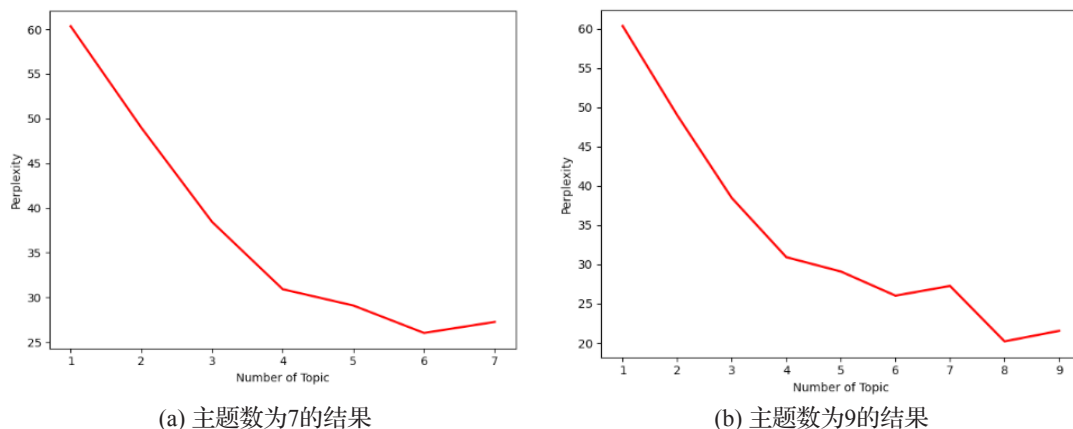


图 5 不同主题数的结果

对数据集进行 LDA 主题识别后, NSF 基金项目数据得到 9 个主题, 依次命名为 F01、

F02、F03、F04、F05、F06、F07、F08、F09, 研究主题识别结果如表 4 所示。

表 4 NSF 基金项目数据主题识别结果

序号	主题词
F01	Climate change, multi-sphere coupling, air-sea interaction, air-sea interaction, ocean wind field, sea surface temperature, atmospheric circulation, marine life process, marine life, deep-sea microorganisms, marine mammals, marine plankton, origin of marine life, marine dynamics Science, ocean circulation, marine sediments, deep-sea sediments, cold-water corals, ocean multi-layer energy and material cycles, seabed geological processesyucitong
F02	Manned submersible, underwater propulsion system, underwater optical imaging, trajectory control, titanium alloy crew cabin, non-metallic crew cabin, transparent crew cabin, watertight connector, deep-sea hydraulic system, underwater navigation and positioning, underwater sound Communication, low-density buoyancy materials, high-density energy, underwater lighting photography, acoustic measurement, sonar imaging, sonar collision avoidance, life support system, underwater path planning, underwater environment perception and map construction, submersible floating and emergency throwing Onboard systems, unmanned submersibles, autonomous submersibles
F03	Underwater detection, ocean detection, acoustic detection, acoustic sensor, optical fiber sensor, electromagnetic sensor, spectral sensor, conductivity sensor, optical sensor, temperature sensor, pressure sensor, in-situ detection, in-situ online sensing, underwater sensor network , Thermosalt, current meter, wave meter, tide gauge, underwater sound velocity meter, profiler
F04	Ocean remote sensing, ocean satellite, ocean radar, synthetic aperture radar, biological remote sensing, ocean dynamic remote sensing, acoustic inversion imaging, long-range ground wave radar, SAR marine target monitoring, spaceborne lidar, satellite inversion, underwater imaging, ocean water color Remote sensing, microwave radiometer, laser altimeter, Haiyan aqua scanner, infrared radiometer
F05	Ocean surveying and mapping, undersea surveying, undersea topography survey, undersea terrain exploration, depth gauge, altimeter, undersea terrain exploration, underwater acoustic positioning, deep towing system, multi-velocity sounding system, side scan sonar, undersea stratigraphic profiler, undersea seismograph , Ocean gravimeter, ocean proton magnetometer, ocean proton magnetic gradiometer, insulation and pressure core remover
F06	Ocean observation, ocean monitoring, sea-air interface, seamless prediction, multi-scale coupling, data assimilation/ data assimilation, fusion assimilation technology, assimilation algorithm, marine meteorology, ocean weather, coupled assimilation, earth system model data assimilation, Sea-ice-land-atmosphere coupling
F07	Wireless power transmission, underwater signal transmission, underwater information transmission, synthetic aperture sonar, ocean Internet of things, ocean observation network, ocean quantum communication, underwater-surface structure and gateway, unmanned fleet, underwater submarine, underwater target detection , Underwater networking, distributed underwater acoustic network, underwater wireless sensor network, mobile platform networking, underwater navigation and positioning, underwater connection, connection base station (underwater connection system), non-contact connection, acoustic Communication
F08	Ocean information, multi-source data fusion, ocean big data, remote sensing big data, big data forecasting, big data assimilation, ocean big data management, ocean big data mining, ocean big data visualization, ocean environment big data space-time unified modeling, ocean big data Intelligent data analysis, supercomputing, data standardization, data storage, coupled computing, data atlas, deep sea artificial intelligence, deep learning
F09	Marine environment, marine ecology, marine pollution, marine ecological protection, marine ecological environment, marine pollution prevention, marine environmental protection, disaster prevention, disaster reduction, tsunami disaster reduction, marine disasters, coastal hazards, natural disasters, marine eutrophication, microplastic pollution , Ocean acidification, land-sea gas pollution, algae blooms, harmful algal blooms, red tides

(3) NSF 资助数据主题关联

通过识别后, 我们对项目数据进行主题关联, 构建结果如下图所示。我们可以发现,

NSF 在海洋科学的资助主要集中在 9 个领域, 分别是包含海洋物理化学生物等在内的基础研究领域、全海深潜水器、水下传感技术、海洋

卫星与遥感、海洋卫星与遥感、海洋环境预测预报、水下通信传输与组网技术、海洋大数据与人工智能、海洋生态修复及灾害预警^[14]。

其中 F01 的关键词汇聚出的关键技术主要集中在海洋物理、海洋化学、海洋生物、海洋地质等，旨在海洋基础性研究。F02 的关键词汇聚出的关键技术主要集中在全海深潜水器，主要是用于深海探测，属于深海关键技术及装备。F03 的关键词汇聚出的关键技术为水下传感技术，如图 6 所示，可以看出在水下传感技术领域，主要集中在水下探测、水下预警、目标检测、海洋水文环境要素监测等。F04 的关键词汇聚出的关键技术为海洋卫星与遥感，主要用于海洋环境监测等。F05 的关键词汇聚出的关键技术为海底地形探测与测绘技术，主

要用于海洋测绘、海底地形探测、勘察等。F06 的关键词汇聚出的关键技术为海洋环境预测预报，主要是通过资料同化，集合预报与适应性观测方面的研究对海洋环境进行预报。F07 的关键词汇聚出的关键技术如图 7 所示，为水下通信传输与组网技术，旨在通过水下信息传输与目标探测识别进行海洋观测。F08 的关键词汇聚出的关键技术为海洋大数据与人工智能，旨在通过大数据的深度学习与分析，建立多层次、多功能的海洋综合信息平台。F09 的关键词汇聚出的关键技术为海洋生态修复及灾害预警，旨在通过污染的跟踪监测技术，污染物控制技术和数值模拟技术等，对我国的沿海赤潮、绿潮、水母等海洋生态灾害进行预警与修复。

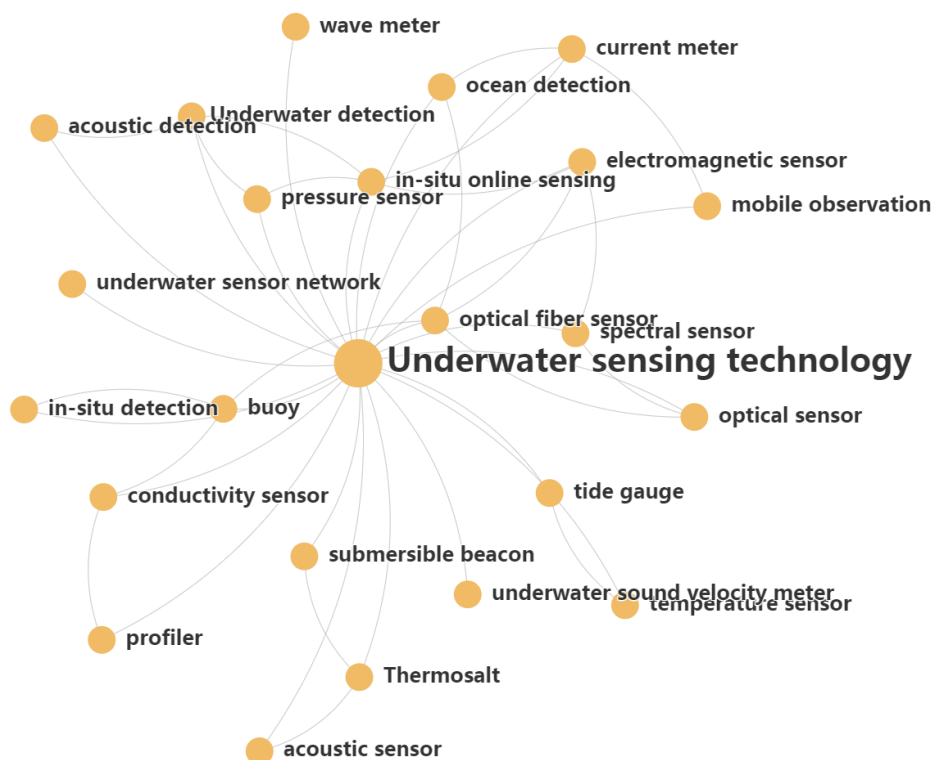


图 6 F03

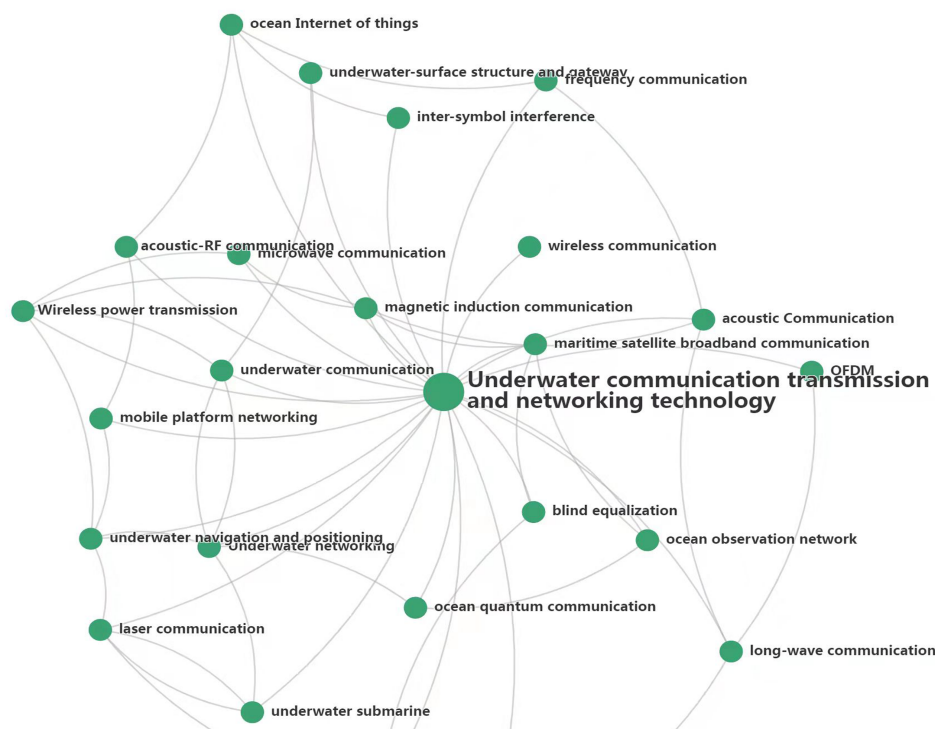


图 7 F07

(4) 总结归纳为四个方向

对九个主题进行总结和归纳,可以发现 NSF 近四十年在海洋科学的资助主要集中在海洋基础科学研究、海洋监测观测、海洋信息化以及海洋环境保护与防灾减灾领域。

海洋物理、海洋化学、海洋生物、海洋地质等基础研究属于海洋基础科学研究方向,海洋基础科学研究既包括微观层面对海水、溶解和悬浮于海水中的物质、生活于海洋中的生物、海底沉积和海底岩石圈,以及海面上的大气边界层和河口海岸带等具体问题的研究,也包括宏观层面对海洋多圈层(水圈、大气圈、生物圈、岩石圈)相互作用机制和过程的研究。

深海潜水器、水下传感技术、海洋卫星与遥感、海底地形探测与测绘技术、海洋环境预测预报五个关键技术属于海洋监测观测研究方

向,海洋监测观测包括卫星、岸基、智能浮/潜标和海床基等固定移动的自动观测平台建设,海上跨域通信、多平台组网、区域水下导航、海底测绘等前沿观测通信技术。

水下通信传输与组网技术、海洋大数据与人工智能两个关键技术属于海洋信息化研究方向,海洋信息化研究主要是推进空天-海面-水下-海底探测监测装备的体系化发展和数据融合处理与应用,包括新一代三维高分“卫星观澜”系统,实现海洋遥感从中尺度到亚中尺度、从海面到上混合层剖面探测的拓展;“海气交互”系统,实现海洋-大气界面关键要素的高时空分辨率、信息的高精度感知和对水下水面目标智能探测;“深海星空”系统,实现深海观测水下移动节点和固定节点协同;“海底透视”系统,建立重/磁/电/震/声的背景和异常地球物理

场探测重大前沿技术体系;“深蓝大脑”系统,基于海洋大数据与人工智能技术建立多层次、多功能的海洋综合信息平台,重点突破海洋复杂环境感知与跨介质一体化通讯组网技术。

海洋生态修复及灾害预警关键技术属于海洋环境保护与防灾减灾方向,研究主要包括发展海洋观测/监测新型传感器技术,研发“两洋一海”自主海洋动力环境数值预报模式;提高我国近海典型生态灾害的预防与早期预测预警能力;研发国家海洋环境安全保障原型平台等。

4 对照验证

美国国家科学技术委员会(NSTC)发布了题为《美国国家海洋科技发展:未来十年愿景》(*Science and Technology for America's Oceans: A Decadal Vision*)的报告,旨在推广过去10年美国取得的成就和进步成果以及确定未来10年海洋科技发展的五大战略目标^[15]。

我们可以发现,五大目标第一是深入认识海洋,主要为现代化的基础设施研发、利用海洋大数据、开发海洋系统模型和应用海洋研究,包括潜水器、飞机、卫星、陆基雷达、系泊和电缆浮标,以及各种无人水下、水面和空中航行器等。第二是发展海洋经济,主要包括海产品生产与加工、探索 and 开发海洋能源、勘探海洋矿物资源、保护海洋生态系统。第三是确保海上安全,包括提高海洋事务感知能力、维护和加强海上运输、了解北极的变化(ocean-ice-Atmosphere),尤其在海洋观测、数据管理、建模和预测方面保持军方在海洋领域的战略、行动和战术优势。第四针对保障人类健康,主

要是预防和减少塑料污染、海洋污染物和病原体、有害藻华防控、海洋药物与天然产物。第五则是沿海地区的可持续发展,包括研究海岸带灾害应对(自然灾害和极端天气)和海岸带防灾减灾能力建设。

通过对十年规划的研究,可以验证我们的研究方法是基本准确的,对比上述方向和识别结果,愿景中提到的发展重点领域大部分是吻合的,缺少的部分主要是海洋油气勘探以及极地研究的部分,而在NSF资助的情况中,二级学科除海洋科学外,NSF还有单独对极地计划的资助。

5 研究结论与启示

根据对我国海洋科学领域的总体研究,虽然我国海洋科技发展起步晚,但发展势头强劲,在海洋科技创新体系建设、海洋科技研发投入、重大基础科学问题研究、关键技术与重点装备研发等方面取得了显著成就,已经成为世界海洋科技网络中的重要力量,海洋科技创新能力基本处于全球海洋科技竞争格局中的第二梯队。呈现“三跑并存、跟跑为主、追赶迅速、局部领先”的局面。

相对于美国、英国、日本等传统海洋强国,目前我国仍有三方面问题比较突出。一是海洋技术与工程建设水平较为落后。二是面向海洋安全保障的观测探测支撑能力较为薄弱。三是水下通信和组网技术等关键核心技术仍受制于人,海洋通用技术尤其是材料和工艺落后。

本文以1980–2019年NSF资助的海洋科学共计342610条项目为数据源,采用一种基于

BERT 和 LDA 主题建模技术的科学研究前沿识别方法,利用深度学习模型 BERT 中词嵌入技术抽取语义近似或同语义概率分配词汇,结合 LDA 主题建模技术进行语义主题建模,对海洋科学领域的语义主题进行建模及前沿方向的推测,揭示了美国海洋领域在 NSF 资助的研究热点信息,希望为我国海洋战略部署重点和未来发展提供信息支撑。

参考文献

- [1] 郭子俊,杨爱华.美国国家科学基金会的发展特点及经验启示[J].自然辩证法研究,2019,35(12):65-70.
- [2] 樊春良,李东阳,樊天.美国国家科学基金会对融合研究的资助及启示[J].中国科学院院刊,2020,35(1):19-26.
- [3] Greengrove C, Lichtenwalner S, Palevsky H., et al. Using Authentic Data from NSF's Ocean Observatories Initiative in Undergraduate Teaching: An Invitation[J]. Oceanography (Washington D.C.), 2020, 33(1):62-73.
- [4] Zhang Y, Kong W, Dong Z Y, et al. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network[J]. IEEE Transactions on Smart Grid, 2019,10(1):841-851.
- [5] 王瑞,龙华,邵玉斌,等.基于 Labeled-LDA 模型的文本特征提取方法[J].电子测量技术,2020,43(1):146-151.
- [6] 张卫卫,胡亚琦,翟广宇,等.基于 LDA 模型和 Doc2vec 的学术摘要聚类方法[J].计算机工程与应用,2020,56(6):186-191.
- [7] 严佩敏,唐婉琪.基于改进 BERT 的中文文本分类[J].工业控制计算机,2020(7):108-110.
- [8] 赵璐偲,岁波,罗海琼,等.基于 BERT 特征的双向 LSTM 神经网络在中文电子病历输入推荐中的应用[J].中国数字医学,2020,15(4):55-57,51.
- [9] Xiao Q, Qin M, Guo P., et al. Multimodal Fusion Based on LSTM and a Couple Conditional Hidden Markov Model for Chinese Sign Language Recognition[J]. IEEE Access, 2019(6):112258-112268.
- [10] 刘铭,咎红英,原慧斌.基于 SVM 与 RNN 的文本情感关键句判定与抽取[J].山东大学学报(理学版),2014(49):68-73.
- [11] 万晨,胡小飞.基于文献计量和复杂网络的环境/生态学高被引论文分析[J].情报工程,2021,7(1):115-125.
- [12] 汪雪锋,于苗苗,韦华楠,等.中国学者在顶级期刊发文的历史变迁与特征演化[J].情报工程,2021,7(3):3-17.
- [13] 李新龙,陈雅丹,胡烨胤,等.1997-2019 年国家自然科学基金中医内科学领域研究前沿的可视化分析[J].中医杂志,2021,62(13):1120-1125.
- [14] 张世童,张宏伟,王延辉,等.自主水下航行器导航技术发展现状与分析[J].导航定位学报,2020(2):1-7.
- [15] Meldrum D. Marine Technology: challenges of the last frontier[J]. Porcupine Marine Natural History Society Newsletter, 2002:2.