

2024

Hotspot perspectives and future directions in the research field of strategic scientists based on LDA-SARIMA model

Tao HUANG

College of Law and Economics, Wuhan University of Science and Technology, China

Zhecan ZOU

College of Law and Economics, Wuhan University of Science and Technology, China

Follow this and additional works at: <https://jstm.researchcommons.org/journal>



Part of the [Public Affairs, Public Policy and Public Administration Commons](#), [Science and Technology Studies Commons](#), and the [Technology and Innovation Commons](#)

Recommended Citation

HUANG, Tao and ZOU, Zhecan (2024) "Hotspot perspectives and future directions in the research field of strategic scientists based on LDA-SARIMA model," *Journal of Science and Technology Management*: Vol. 26: Iss. 3, Article 3.

DOI: 10.16315/j.stm.2024.03.008

Available at: <https://jstm.researchcommons.org/journal/vol26/iss3/3>

This Article is brought to you for free and open access by Journal of Science and Technology Management. It has been accepted for inclusion in Journal of Science and Technology Management by an authorized editor of Journal of Science and Technology Management.

Creative Commons License



This work is licensed under a [Creative Commons Attribution-NonCommercial-No Derivative Works 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

我国战略科学家研究的热点透视与未来路向： 基于 LDA - SARIMA 模型的研究

黄涛, 邹浙灿

(武汉科技大学 法学与经济学院,湖北 武汉 430065)

摘要:战略科学家是科技人才中的帅才,是支撑我国高水平科技自立自强的重要力量。探索战略科学家研究领域主题存在的取向和呈现的趋向,可为相关科技政策研究和制定工作提供借鉴。本文采用 LDA 主题建模和 SARIMA 预测分析方法对战略科学家这一重要研究领域进行主题挖掘与实证分析,抽取与战略科学家相关的6个核心主题,对该领域的研究现状与发展动态进行探讨。研究表明:战略科学家的研究具有显著的周期性特征;研究集中在建设国家战略人才力量层面,强调战略科学家对于承担国家战略任务,推动战略人才体系建设的作用;通过模型的定量预测功能得出培育和使用战略科学家、借助新媒体技术开展科学普及和教育工作是未来研究的趋向。建议在探讨如何完善战略科学家的培养体系时,紧跟时代发展的步伐,培养符合新时代需求的战略科学家。

关键词:战略科学家;实证分析;主题挖掘;LDA;主题模型

DOI:10.16315/j.stm.2024.03.008

中图分类号:G 316 **文献标志码:**A

战略科学家是国家战略人才力量中的“关键少数”^[1]。国际环境的动态变化与我国目前发展阶段的迫切现实需求,促成了战略科学家成为国家科技安全与发展的“刚需”与学界研究的焦点^[2]。战略科学家既指国家战略科学家,也包括学科领域或行业领域的战略科学家。国外尚无针对战略科学家的专门研究,诺贝尔奖获得者是学科领域战略科学家的典型代表,国外学者对诺贝尔奖获得者进行了深入研究。Zucherman^[3]对美国诺贝尔奖获得者的研究表明,获奖者的师承关系、家庭影响、工作单位、科研产出等对人才的成长起到促进作用。Berry^[4]对诺贝尔奖获得者的研究显示,个体水平上所获得的成就与特定的阶层和家庭背景有关。Hillebrand^[5]通过对1901—2000年间诺贝尔物理学奖获得者的传记资料进行研究,发现获奖者的年龄结构、团队合作程度、人员流动情况、文化背景、高等教育经历以及工作调动等因素能够影响诺贝尔获奖者的成长成才;Stephan等^[6]、Karazija等^[7]、Costas等^[8]运用定量

统计分析,探究年龄对科学家科研能力的影响,概括出年龄相关规律。

相较于国外学界,我国有不少学者围绕战略科学家的概念与特质、成长路径和培养策略进行了学术探讨。在战略科学家的概念与特质方面,杨玉琴等^[9]利用科学家角色模型,指出战略科学家是以爱国主义为核心,兼具深厚科研能力、实践经验和宏观战略规划能力的高层次人才。芮绍炜等^[10]通过对钱学森、钱伟长、黄大年等战略科学家的人物事例进行分析,认为战略科学家应具备遵从国家使命、有爱国情怀、奋斗在科学研究第一线、具备深厚科学素养、跨学科能力和组织领导能力等特征;在战略科学家的成长路径方面,孙昌璞^[11]基于科技发展战略和新型举国体制的内涵,提出国家的信任和老一辈科学家的放手培养是战略科学家获取成功的重要因素。瞿群臻等^[12]基于生存分析视角,对战略科学家成长路径展开研究,得出工作单位性质、海外留学经历、最高学历授予单位等是影响战略科学家成长的重要因素。在战略科学家的培育策略方面,潘俊强^[13]通过访谈老干部、研读有关老科学家回忆文章,得到培养战略科学家的重要启示,其中包括养成正确的三观、磨练心性、聚焦拔尖创新人才培养等。冯燊等^[14]选取中外代表性战略科学家为研究对象,运用

收稿日期:2024-03-11

基金项目:教育部人文社会科学研究规划基金项目(23YJAZH027);湖北省高校哲学社会科学研究重大项目(23ZD153);湖北省教育厅人文社科重点项目(22D024);中国高等教育研究会重点项目(22YZ0204)

作者简介:黄涛(1971—),男,教授,博士生导师;
邹浙灿(2000—),男,硕士研究生。

CV 分析法,从社会背景、科技贡献、科学精神等维度归纳战略科学家特征,进而提出改善战略科学家成长的社会生态、优化战略科学家的发现和评价机制等培养战略科学家的相关建议。

由此可见,针对战略科学家领域的研究,大部分以定性描述为主,如以访谈法、履历分析或是理论模型对战略科学家的概念内涵、成长经历及培育方法进行研究。为进一步探讨战略科学家领域的研究进展、特点和发展趋势,本文拟运用潜在狄利克雷分配(LDA)模型和回归线性模型、季节性差分自回归移动平均模型(SARIMA),对战略科学家研究主题进行当下发展状况的分析以及未来发展趋势的预测。

1 研究设计

1.1 数据来源及处理

本文所采用的数据资料主要源自中国知网(CNKI)、万方、维普等数据库。由于各大数据库网站均有相应的版权保护,防止未经授权的数据爬取行为,因此本文并未使用 python 的数据爬虫功能,而是使用各大数据库网站的文献导出功能,以“自定义”模式,选取了包括文献来源、标题、摘要、发表日期和关键词等信息指标。为了保证数据的全面性和连贯性,选取的文献类型涵盖了全部期刊论文、硕博学位论文、会议论文及报纸。因此,设置检索条件为“主题 = 战略科学家 OR 标题、摘要或关键字 = 战略科学家 AND 年份:1997—2024 AND 搜索范围:总库”,聚焦的具体数据时段为 1997 年 1 月 1 日起至 2024 年 4 月 2 日止。

在导出上述文本数据的基础上,使用 python 的数据清洗功能并对数据进行预处理。本研究构筑了一个特定的停用词表,将 github 上包括哈工大停用词、四川大学机器智能实验室停用词、百度停用词、最全中文停用词进行整合得到新的中文停用词表,合计停用词 2 462 个。随后,针对导出文献的具体特征进行人为调整,提升模型精确度。该流程涵盖了去除文本格式标识、非文本元素,以及在分析过程中较为常见但并非关键的词语,从而对无关文献、文献中的无关内容、错误内容和冗余内容进行清理,令文档中的文本格式和文字内容符合模型分析的标准^[15]。

最后,利用 Python 中的 jieba 分词库对文本进行再处理。此阶段的操作包括创建专属的词汇表,移除停用词以及融合意义相近的词汇等。为了提升分词结果的质量,研究中扩展了停用词表和同义词表的覆盖范围,并针对每篇文献,将相关联文档转换为词袋(bag-of-words)模型并进行训练,为接下来的主

题分析工作奠定坚实基础^[16]。最终,成功提取了 514 篇战略科学家研究主题的文献资料,且以 Excel 的文件形式进行整合并呈现。

1.2 研究方法

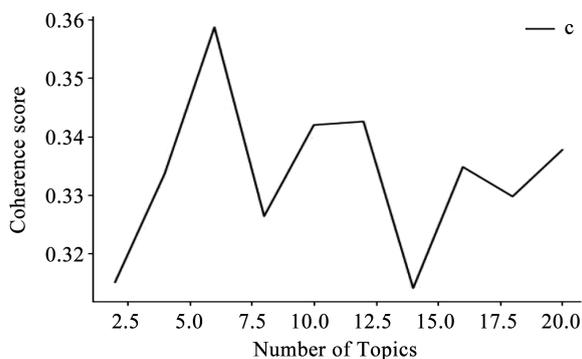
1.2.1 文档 - 主题生成模型 LDA

利用 LDA 进行主题建模,研究者能够便捷地对所需研究领域进行主题归类或是对其相关的文本数据进行分类处理。各个主题和各篇文档均拥有独特的概率分布,分别用符号 ϕ_k 和 θ_d 加以表示^[17]。

此外,在主题建模领域,选定理想主题数的策略主要涉及应用模型困惑指标、主题一致性检验以及结合 pyLDAvis 的图形化分析。模型的困惑指标越低,便昭示着模型的品质更卓越^[18]。主题一致性评级愈高意味着主题的连贯性愈加强烈,同样对于主题的阐释性和建模成效亦有益处。

然而,鉴于本研究涉及的数据体量相对较小,运用困惑指标所得出的拟合成效可能会不够理想^[19]。基于此,本研究采取了主题一致性检验,并借助 pyLDAvis 的图像结果来确认最合适的主题数额。最终的分析结果由 matplotlib 库生成主题一致性的曲线图予以图形化呈现,从而更形象地展现出主题数量和模型效能之间的相关性。

在具体应用过程中,通常会挑选出一致性评级最佳的模型作为最理想之选,如图 1 所示。由图 1 折线的峰值位置可知,当模型设定的主题数为 6 个时,所呈现的主题一致性获得了最佳表现。



Optimal number of topics:6
<Figure size 640x480 with 0 Axes>

图 1 一致性曲线变化图

Fig. 1 Diagram illustrating the variations in consensus curves

pyLDAvis 是针对展现主题模型分析结果而设计的 Python 工具包,适用于呈现 LDA 模型的分析成果。在 pyLDAvis 的视觉呈现界面里,每个圆形图案代表 1 个主题,其中圆形位置及其尺寸展示了该主题在整体语料库中的重要性及分布特性。若 2 个主

题的圆形图案重合,则表明这 2 个主题可能具有较高相似度。在最理想的场景下,期望每个主题能够与其他主题有明显区分。因此,2 个圆形图案之间

存在较少的重叠通常被看作是结果较佳的展现,下面展示了不同主题间的分布及其相互间的关系,如图 2 所示。

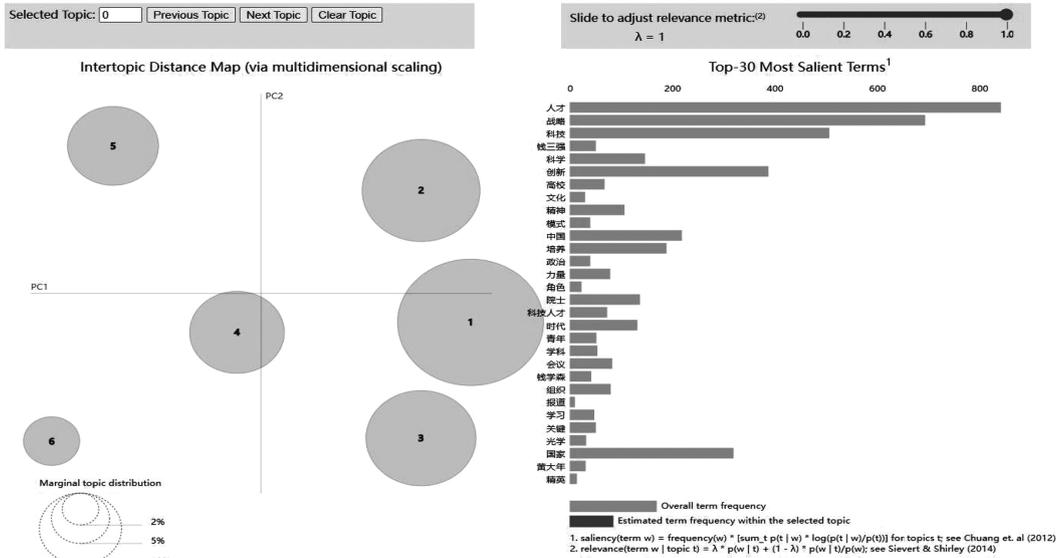


图 2 pyLDAvis 可视化示意图

Fig. 2 Schematic visualization of pyLDAvis topic model analysis

借助于 pyLDAvis 工具的图形展示,观察到在设置主题数目为 6 个的状况下,各主题间保持了清晰的界限,没有出现交叉重叠现象,这进一步证明了分类的效果十分显著。基于以上观察和分析,本研究决定采用 6 个主题作为最理想的主题数量选择。

随后,利用 LDA 模型的主题建模技术,对与战略科学家研究领域相关的文献资料进行主题挖掘,提炼出 6 个显著主题,并构建展现主题与词汇项概

率关系的分布图表,并罗列与每个主题相关的 10 个核心词汇,如表 1 所示。例如,在 Topic 1 中,借助 pyLDAvis 可视化示意图可观察到若干个词项显示出高概率值,另外可以发现,该主题前 10 位的高频关键词为“人才、战略、科学家、培养、科技、发展、创新、国家、党、科技人才”。因此,该主题针对的是战略科学家的特质与培养策略的相关议题。上述即为主题识别过程。

表 1 文献主题 - 词项概率分布

Tab. 1 Distribution of theme-term probabilities in the literature

主题类别	主题名称	主题内前 10 位高频关键词	比重/%
Topic 1	战略科学家的特质与培养策略 ^[20]	人才、战略、科学家、培养、科技、发展、创新、国家、党、科技人才	31.5
Topic 2	关键核心技术攻关的组织机制	发展、力量、关键技术、核心技术、基础、自立自强、组织、现代化、科研、机制	20.5
Topic 3	新时代科技强国建设与科技领军人才 ^[21] 的使命	院士、时代、会议、学习、中国、强国、健康、世界、重要讲话	17.9
Topic 4	加快培育壮大国家战略人才力量 ^[22] 与服务支撑国家战略	改革、国家、学科、高校、企业、发展、工程、培育、材料、人才体系	13.2
Topic 5	战略科学家钱学森与战略科技人才培养机理	科学、钱学森、中国、文化、政治、角色、钱三强、模式、精英、学院	12.2
Topic 6	科学家精神传播与科普教育的媒体角色	报道、基地、模式、精神、时代、典型人物、新闻、生产、楷模、科普	4.6

接着,基于主题识别和时间序列分析建立强度模型。使用 LDA 模型挖掘的战略科学家研究主题,显示其强度如何随时间变迁而波动,如图 3 所示。这些图表由 Matplotlib 库绘制,它们以直观的方式描绘了不同主题的关注度演变趋势^[23]。

最后,筛选拥有潜在重要意义的热门主题。本过程采用了 Python 的 Scikit-learn 库以实施线性回

归模型(LR)的构建,从而辨别各个主题的热度排名。其中,运行代码中所使用的线性回归模型基于最小二乘法,通过 linregress 函数计算出回归线的斜率(slope)和截距(intercept),以下是如何从一组给定的数据点中通过最小二乘法计算斜率和截距的具体公式:

使用最小二乘法来找到每个主题强度随时间变

换的最佳拟合线的过程,涉及到利用1组线性方程来最小化预测值和实际数据点之间的差异平方和。所谓的“最佳拟合线”是指能够最好地代表数据点分布趋势的直线。

对于每个主题(i),假设有1组时间点(T_1, T_2, \dots, T_n)和相应的主题强度($y_{i1}, y_{i2}, \dots, y_{in}$),则期望斜率(m_i)和截距(b_i)被找到来最小化残差平方和,其数学表达式如下:

$$\min_{m_i, b_i} \sum_{j=1}^n (y_{ij} - (m_i t_j + b_i))^2. \quad (1)$$

要解决这一最小化问题,通常要计算斜率(m_i)和截距(b_i)的公式,这些公式可以通过下面2个正规等式(normal equations)计算得出:

$$m_i = \frac{n \sum (t_j y_{ij}) - \sum t_j \sum y_{ij}}{n \sum (t_j^2) - (\sum t_j)^2}, \quad (2)$$

$$b_i = \frac{\sum y_{ij} - m_i \sum t_j}{n}. \quad (3)$$

其中: n 是数据点的数量, \sum 是对所有观测到的数据点求和的步骤,根据上式求得每个主题(i)的斜率(m_i)和截距(b_i),就可以应用于主题强度随时间变化的最佳拟合线方程:

$$y_{ij} = m_i t_j + b_i. \quad (4)$$

根据式(4),本研究最终将定义热门主题为那些强度超出平均水平且总体展现正增长态势的主题,冷门主题则相反。筛选热门主题的核心过程包括以下几个步骤:第一,测算出每个研究主题强度时间序列的平均值,并淘汰低于此均值的主题;第二,应用线性回归模型对各主题强度时间序列进行趋势拟合,并使用建立的斜率进行对比,最终通过排名结果,识别出热、冷门主题。

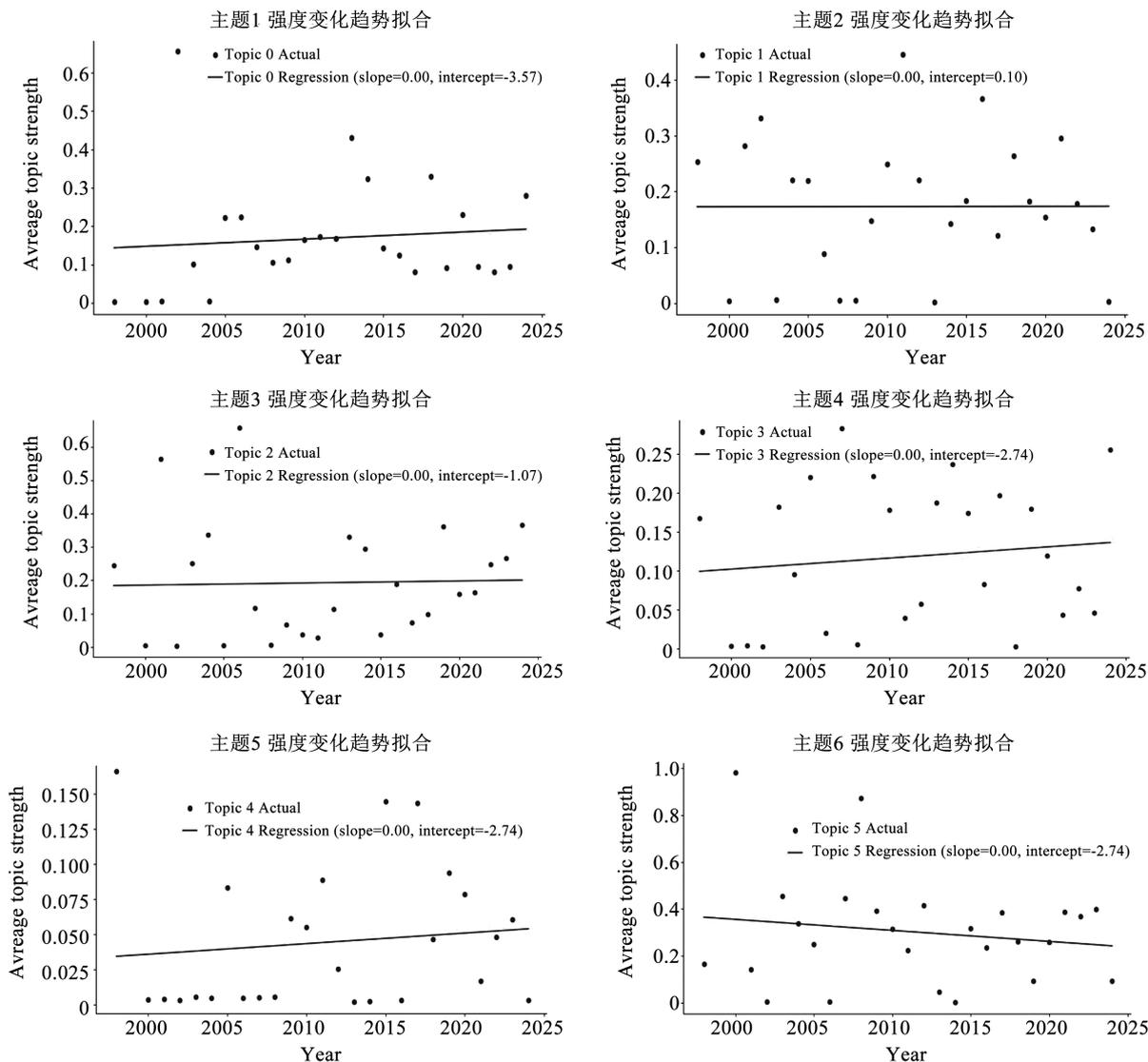


图3 主题强度趋势变化图(1997—2024年)

Fig. 3 Graph illustrating trends in topic intensity over time(1997—2024)

1.2.2 SARIMA 模型

SARIMA 是一种时间序列预测模型,在该模型中,共包含非季节性的 3 个参数(p, d, q),同时它还带有 4 个季节性参数(P, D, Q, s)。非季节性部分的 3 个参数可处理大多数时间序列数据的趋势和非周期性变动,而季节性部分的 4 个参数在数据中存在明显的周期时非常有用,例如每年、每季、每月或每周等周期性变化。SARIMA 模型在继承 ARIMA 模型对趋势变动敏感捕捉的特性基础上,额外纳入了季节性变化的分析参数,以适用于具有季节周期性波动的时间序列数据^[24]。

本文拟运用“模型融合”的方法^[25],在 LDA 模型完成对我国战略科学家研究的热门研究主题识别工作的基础上,结合 SARIMA 模型对各大热门主题的未来路向展开研究,研究的具体过程如下:

1)使用 SARIMA 模型对主题数据进行预测。其中,用四步检验方法排除 ARIMA 模型,确定 SARIMA 模型的可行性,并生成主题强度变化趋势预测图,它能模拟 6 大主题未来 5 年(2025—2030)的发展趋势。

2)结合 LDA 和 SARIMA 2 种模型的建模结果,对各主题的热门主题、主题的发展前景分别进行分析探讨,并依序提出自己的见解和建议。

2 结果与讨论

2.1 主题内容解析

综合分析战略科学家研究 6 大主题间的关键词交集、内容联动性和各主题所占权重,可将 6 个主题进一步聚焦为 3 个相关联的群组。这些群组的内容不仅紧密连接,且彼此间也无排斥性,它们集中体现了关于战略科学家研究的多元主题框架。其中,将 Topic 1、Topic 2、Topic 3 主题并入“国家科技发展战略”群组;Topic 4、Topic 5 主题归类于“国家战略人才力量”群组;Topic 6 被单独归类为“科技传播与社会影响”群组。

2.1.1 国家科技发展战略

“战略科学家的特质与培养策略”“关键核心技术攻关的组织机制”和“新时代科技强国建设与科技领军人才的使命”,三者间的共通点在于强调战略科学家参与新时代科技强国建设和推进关键核心技术协同攻关的重要性。在关键产业技术领域识别、培养一批战略科学家^[26]是新时代科技强国建设、依靠人才引领提升科技攻关体系化能力、不断提升国家创新体系整体效能的重中之重。“十四五”规划、“中国制造 2025”以“国家中长期科学和技术

发展规划纲要”等一系列政策文件的颁布体现了对战略科学家的重视,其中也包含了上述 3 个主题间的联系。

2.1.2 国家战略人才力量

“加快培育壮大国家战略人才力量与服务支撑国家战略”“战略科学家钱学森与战略科技人才培养机理”强调了建设国家战略人才力量的重要性。国家战略人才力量包括战略科学家、科技领军人才和创新团队、青年科技人才队伍和卓越工程师等^[27]。要加快建设国家战略人才力量,需要坚持长远眼光,有意识地发现和培养更多具有战略科学家潜质的高层次复合型人才,形成战略科学家成长梯队。

建设和完善战略科学家发现、培养、激励机制受国内政策调整、国家重大人才工程实施、高等教育改革等因素的影响。例如,通过实施如“科技人才激励政策十五条”“长江学者奖励计划”等,我国加大了对顶尖科学家的培养和引进力度。钱学森计划的启动、钱学森实验班的创办等,体现了我国在教育改革方面,尤其是高校教育层面为培养战略科学家所做的努力。综上,我国通过实施科技人才激励政策、提供发展平台、强化教育资源供给等,不仅激发了年轻一代学者的科学兴趣和创新潜能,还将国家战略需求内化、嵌入到了科技人才培养的体系中。

2.1.3 科技传播与社会影响

媒体作为新时代传播公众信息的重要渠道,肩负着一定的社会责任。借助媒体,我国科技传播内容创新既要夯实基础,又要聚焦于前沿性科技创新内容的传播。科技传播取得的良好社会效果能为战略科学家成长营造优良的外部环境。通过打造“科普中国”平台,开展“全国科普日”“五四科技传播日”活动等,一批战略科学家的主要事迹经由媒体传播,使他们逐渐成为公众心目中的科学偶像和楷模,这为战略科学家的培育创造了良好的舆论环境。

2.2 主题趋势分析

2.2.1 主题强度分析

通过应用 LDA 模型,辨识出了 2 类重点主题:一部分随着时间递进逐步吸引了学界广泛的注意(即热门主题),而另一部分则逐渐减少了其被关注的程度(即冷门主题)。在研究过程中,关键的一环是追踪并分析这些主题的变迁轨迹。为了判断主题受欢迎程度的变化,本研究检视了各主题随时间流逝的权重波动。同样,研究还探究了主题的影响力,并评估它们如何与时间相互作用。

经过计算和可视化筛选,最后得出斜率排名“Topic 4 > Topic 5 > Topic 1 > Topic 3 > Topic 2 >

Topic 6”,如图3所示。可见在6个主题之中,“主题4 加快培育壮大国家战略人才力量与服务支撑国家战略”是其中的热门主题,而“主题6 科学家精神传播与科普教育的媒体角色”是其中的冷门主题。

针对上述结果,将分别对热门、冷门主题如此筛选的原因,以及其研究价值展开论述:

第一,大力培养使用战略科学家对国家未来发展以及构建具有全球竞争力的科技创新体系至关重要。在新时代背景下,党和国家事业发展对教育的需要、对科学知识和优秀人才的需要比以往任何时候都更迫切。因此,加快建设国家战略人才力量,培养战略科学家,以高质量人才服务支撑国家战略是当前国之所需。教育部、财政部等在《关于深入推进世界一流大学和一流学科建设的若干意见》中就提出了高层次创新人才供给能力不足的问题,要求强化立德树人,造就一流自立自强人才方阵,即大力培养、引进一大批具有国际水平的战略科学家、一流科技领军人才、青年科技人才和创新团队。“海外高层次人才引进计划”“青年英才计划”等均是為了识别、吸引和培养能在科技领域参与战略决策的顶尖人才而推出的重大举措。

第二,战略科学家的成功事例往往能激励政府形成人才发展项目和改善人才培育机制,进一步推动相关主题的研究工作。例如,由于钱学森在近程导弹、中程导弹研制时做出的贡献,袁隆平在杂交水稻技术层面的成就以及黄大年在移动平台探测技术等高新技术领域取得的重大贡献,国家更愿把资金以项目形式投入到这些领域的人才培养和科研工作中。

第三,科学家精神传播和科普教育在提升公民基本科学素质和塑造未来战略科学家方面发挥着基础性作用,但目前并未得到足够的研究。一方面,科学家精神传播和科普教育在提升公众科学素质,以及在普及科学知识方面扮演着重要角色,并在培养创新大军和领军人才、培育创新氛围和战略科技力量层面有着愈发显著的支撑作用^[28]。另一方面,由于科普教育与科学家精神传播的研究工作在成效上不够直接和明显,其政策优先级弱于包括战略人才、强国建设、技术攻关等内容研究主题,这间接导致学界对其关注度的降低。然而长远来看,科学家精神传播和科普教育对于培育具有探索精神和创新能力的战略科学家仍不可或缺。

此外,结合关键词的内容来看,主题4中排名前5的关键词分别为改革、国家、学科、高校和企业,体现了战略科学家研究涉及国家政策实施、高校教育、

企业改革等多个层面,侧面说明了科技创新“四链”融合的重要性。科技创新的四链融合强调引领产业技术变革的创新链、教育链、人才链、产业链深度融合对于锻造战略人才力量方阵的作用。这就需要以人才链为总牵引,优化教育链、激活创新链、服务产业链,形成以战略科学家为引领的“雁阵格局”,并依托高水平实验室和大科学装置等创新平台,协同科技领军企业牵头打造创新基地,加强基础研究和关键核心技术攻关,推动原创技术策源地的建设。后5位的关键词分别为发展、工程、培育、材料、人才体系,可以看出材料科学家、关键基础材料的研发及其工程应用与战略科学家存在一定联系。

主题6的关键词部分主要以报道、精神、新闻、典型人物、科普等带有宣传性质的内容进行呈现。我国目前十分重视科普工作,其中,开展人物科普工作是为广大群众树立良好的科学家形象,传播科学家精神的有效途径。相较于主题4对于人才识别、培养的重视,主题6根本上与主题4研究内容的目标一致,都是为了尽快建立战略科技人才图谱,提升战略科学家培育使用成效所做的努力。

综上所述,加快培育壮大国家战略人才力量与服务支撑国家战略这一研究主题相对于其他主题来说,在政策支持、社会需求、实践应用以及公众关注等方面占据了优势,故而成为战略科学家研究领域中最热门的研究主题。科普教育和科学家精神传播虽然由于上述提及的原因没能获得广泛关注,但同样是战略科学家研究领域不可忽视、具备重要价值的研究内容。

2.2.2 SARIMA 模型有效性分析

为了对主题未来的发展状况进行预测分析,对上述6个研究主题执行SARIMA模型搭建。由于在研究过程中比较ARIMA模型和SARIMA模型后,对于数据的拟合情况判断出战略科学家主题数据在加入季节性因素后的SARIMA模型的拟合效果和预测效果普遍优于普通ARIMA模型,因此认为其数据可能具有周期性特性,并选择SARIMA模型对该主题进行可视化预测分析。

1) 国家战略需求的调整。国家战略需求的调整会导致战略科学家的研究工作呈现周期性变化。例如,当我国政府强调要在某一科技领域实现突破时(如“中国制造2025”计划中的高端制造业),就会推动学者和研究机构在这些领域展开深入研究。加之近些年我国频繁受到美国发起的贸易战、科技战的威胁,国家对培养战略科学家的需求也在有关芯片断供、技术封锁、贸易限制的事件中被反复提

及,这将一定程度上使战略科学家的研究工作在时间上呈现某种周期性特征。

2) 重大科研项目和基金的审批与结题。我国重要科研项目往往有固定的申请和结题周期,诸如国家高新技术研究发展计划(863 计划)、国家自然科学基金等。战略科学家研究领域的学者们会在一些项目周期内加快研究进度,并在结题或检查阶段及时发布研究成果,这会导致特定时间内文献数量的增加。

3) 学术工作与政府议程在时间上的同步。有关战略科学家的研究工作可能受政府议程的影响。如在政府工作报告发布、全国两会等政治活动开展的前后,涉及国家重大战略需求、战略科学家培育、战略人才培养等内容的相关研究领域的发文量可能会增多,以反映或应对政府会议中所聚焦的战略议题。

4) 国际环境与合作周期。有关战略科学家的研究工作也可能受国际合作项目和全球研究趋势变化的影响。例如,中外合作研究项目的启动和成果通报往往有特定的时间框架,这可能导致该领域的学术文献资料呈现出一定的周期性特征。

鉴于此,由于战略科学家的研究领域与国家战略需求和政策调整存在一定的关联性,其周期性特征将在数据中体现。因此在进行主题建模和时间序

列分析时,使用能捕捉这些模式的 SARIMA 模型就比较合适。

建构 SARIMA 模型时,参数设定与模型验证显得十分关键。在详细的实验操作中,采用 BIC 指数来锁定模型参数,即通过遍历各种潜在参数来计算模型的 BIC 值,挑选出最小值作为最佳参数设置,以此来对模型优化^[29]。最终,对 6 个主题的最优模型参数进行筛选,最终结果都指向“Topic 1-6: Best BIC: -inf, Best Order: (0, 1, 1), Best Seasonal Order: (0, 1, 1, 12)”,可见,在 SARIMA 模型中,6 个主题的最佳模型参数是一致的,都显示为非季节性部分 (0, 1, 1), 周期为一年的季节性部分 (0, 1, 1, 12)。

由于在进行 SARIMA 模型诊断时,需要对模型的残差进行分析,确保它们表现出良好的统计属性。因此,在设定好最优模型参数之后,对所搭建的模型进行 4 个阶段的验证。如果四步检验通过,即残差类似于白噪声、呈正态分布、通过 QQ 图检验显示残差分布接近正态分布,且 Ljung-Box 检验显示残差没有显著的自相关性,则可以认为该 SARIMA 模型是有效的。

综上,为展示四步检验的效果,以 Topic 1 为例,生成了可视化的四步检验图,如图 4 所示。

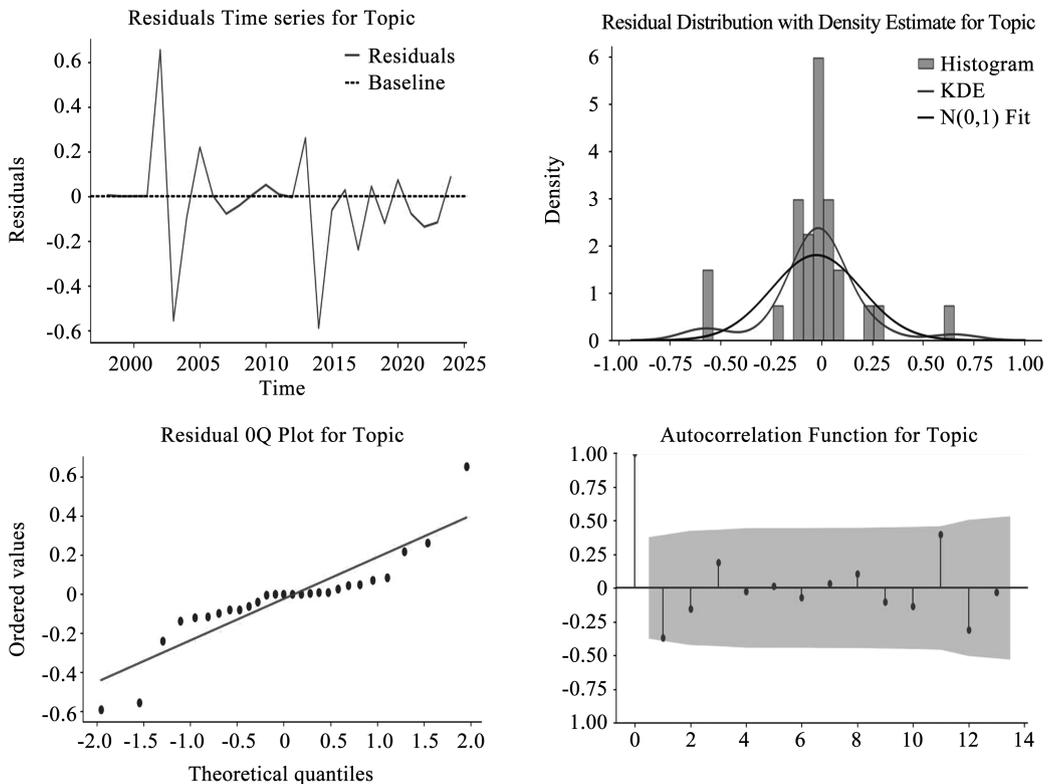


图 4 SARIMA 四步检验图

Fig. 4 Diagram depicting the four-Step SARIMA model diagnostic checks

由图 4 可知,建立的 SARIMA 模型顺利通过了四阶段检验,表明它适用于预测主题强度的变化趋势。其他主题与 Topic 1 类似,皆通过了四阶段的检验,而以同样条件建立的 ARIMA 模型在实验过程中并未通过四步检验,因此选用 SARIMA 模型并进入下一步的预测工作。

2.2.3 主题发展趋势预测

通过应用 SARIMA 模型预测 6 个主题的趋势,据此绘制趋势预测图,针对性地展示热门主题的时间序列预测。通过对 1997—2024 年期间主题强度时间序列的原始数据拟合,进一步证实模型的有效性和精确度。随后,基于构建的 SARIMA 模型,对 2025 年至 2030 年间主题的趋势变化进行预测,具体结果如图 5 所示。

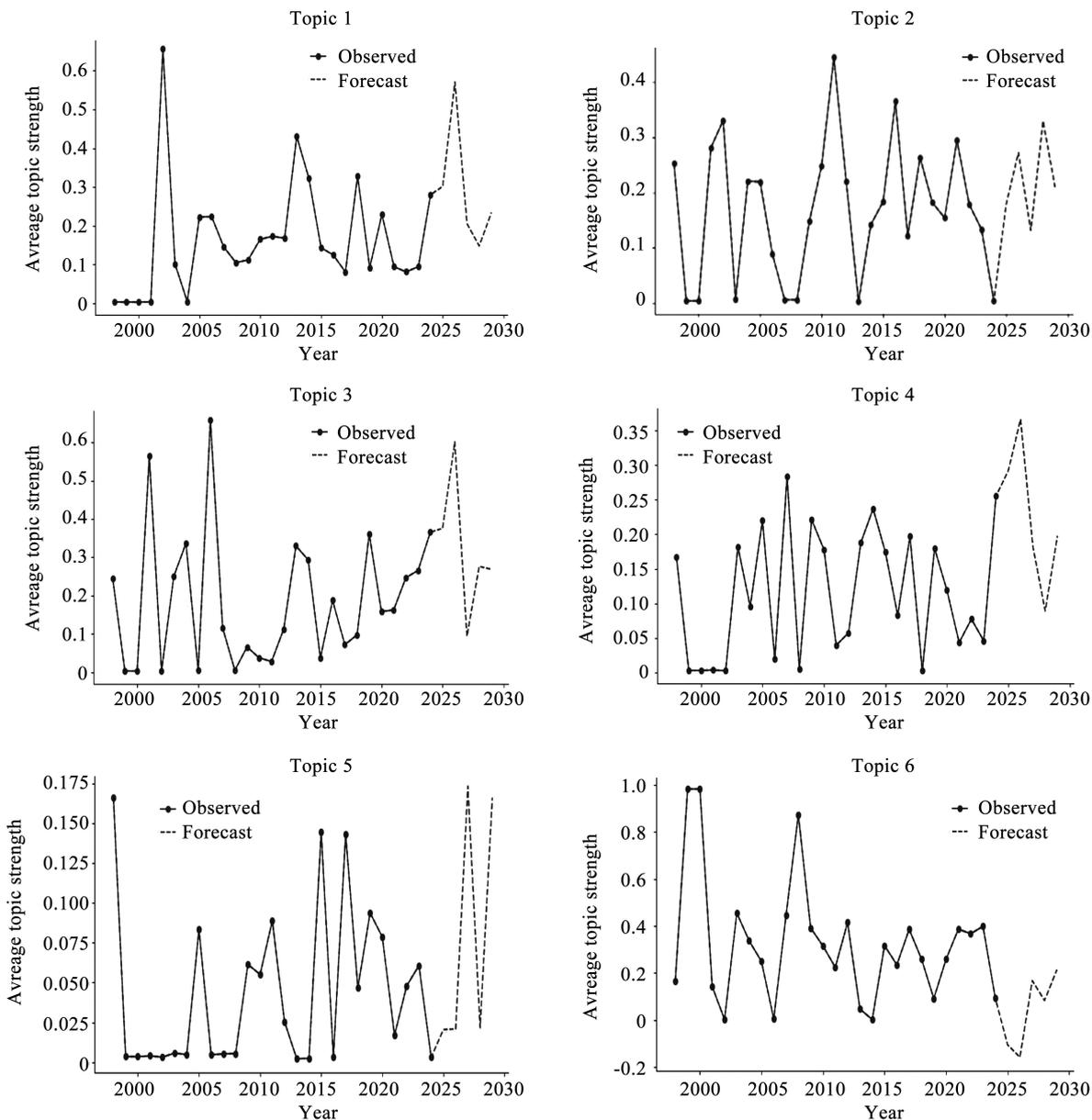


图 5 主题强度变化趋势预测

Fig. 5 Forecast of Trends in Topic Intensity Variation

由此可见,尽管根据先前的评估,六大焦点主题的整体发展倾斜度排名呈现“Topic 4 > Topic 5 > Topic 1 > Topic 3 > Topic 2 > Topic 6”。不过,根据 SARIMA 模型的预估结果观察,预计在接下来的 5 年内,Topic 5 的成长态势将最为显著,Topic 6 则

总体上呈稳步上升趋势。此外,其余的主题增长趋势显示出一些波动现象,而 Topic 3 的发展速度预测可能将是最缓慢的。根据上述的预测结果,可以对相关主题进行以下推断和分析。

1) Topic 3: 新时代科技强国建设与科技领军

人才的使命。由图5可知,该主题强度的增速最慢,这反映了新时代科技强国建设与科技领军人才使命相关政策、理论研究和实践活动的深度和复杂性。尽管其进展缓慢,由于新一轮科技革命和产业变革加速演进,以科技为核心的国际竞争日趋激烈,如何推进科技强国建设并发挥科技领军人才作用依然是目前国内学者研究的重点。

目前,我国正朝着创新型国家的目标努力,科技进步和创新逐渐成为推进我国新时代科技强国建设的主要途径,科技领军人才的培养也愈发占据重要地位。其中,科技领军人才作为战略科学家的“后备军”^[30],不仅要学习培养战略科学家的战略眼光,更要学习其能团结大批科技创新人才共同奋斗的崇高道德风尚和人格魅力。

综上所述,新时代科技强国建设和强化科技领军人才使命感是在当前战略科学家研究领域中值得探讨的内容,反映了我国学者对于国家发展趋势的把握和对战略科技人才培养的深刻理解。因此,尽管SARIMA模型预测上述探讨的主题在未来5年的发展速度最缓慢,但该主题在我国学界依旧具备一定的研究价值。

2) Topic 5: 战略科学家钱学森与中国战略人才培养机理。由图5可知,从2025—2030年,该主题强度的成长趋势最为显著。钱学森作为杰出的战略科学家,成长中开放的社会环境、新中国提供的机会与平台、善于处理科学与政治的关系^[31]等因素造就了这位战略科学家。钱学森能够成为战略科学家,是多种因素综合而成的结果,探究这些影响因素的作用机理,能提供培养使用战略科学家的启示。

钱学森认为科学人才需要注重系统科学的思维方式和解决问题能力的培养。钱学森本人更是战略科学家的重要典范之一^[32],这使得部分学者更倾向在战略科学家研究中以钱学森为典型的示范对象,通过探讨分析其参与过的具体事件及发表的言论,在研究工作中挖掘其科技战略思想。SARIMA模型预测该主题在接下来5年内发展势头最为强劲,这意味着通过研究钱学森等战略科学家的成长规律,探讨如何培育战略科学家将会成为未来我国学界的主要研究方向之一。

3) Topic 6: 科学家精神传播与科普教育的媒体角色。该主题强度呈现稳步上升趋势,反映了通过媒体传播科学家精神以及实现科普教育的相关内容将逐渐受到学界的重视。在我国科技事业不断发展壮大的过程中,在一代代科技工作者勇攀高峰的接续奋斗中,科学家精神孕育而生^[33]。而战略科学

家作为众多科技工作者中的帅才,其精神更值得学习和借鉴。深入探讨战略科学家的精神特质,长远来看将有益于科学家精神的传播以及战略科学家的培育。

随着数智时代的到来,媒体功能的不断升级,媒体平台在传播科学家精神和科普教育方面将具有不可替代的作用。虽然该主题目前被识别为冷门主题,但随着人工智能技术与媒体的结合,基层数字化平台的普及等,该主题的热度预计将在未来五年不断升高,科普教育和科学家精神传播的重要性将更加凸显,成为社会和学界关注的热点。

4) 其他。其他主题如战略科学家的特质与培养策略(Topic 1)、关键核心技术攻关的组织机制(Topic 2)和高教改革与国家人才培养机制(Topic 4),均围绕提升国家的科技创新能力和人才培养质量展开。这些主题的研究内容包含推动教育体制改革、加强科技领军人才与创新团队的合作,以及形成有利于战略科学家成长的政策环境等,提出了系统推进战略科学家培养和使用的对策,这些是战略科学家研究领域需关注的重点。在厘定战略科学家的概念内涵、特质和培养策略的基础上,借助教育改革和国家激励政策培育和使用战略科学家,是发挥其在关键核心技术攻关引领作用的重要手段。但值得注意的是,无论是教育改革,政策实施还是技术攻关,都需要循序渐进的实现过程,且与时代背景息息相关。我国在不同历史阶段对战略科学家的定义和作用有一定的调整,因为“战略科学家是一个从实践中提炼出来的概念”^[34]。

结合上述对主题5和主题6的分析可以得知,目前国内学者侧重对老一辈战略科学家如钱学森、钱三强、邓稼先等人的成才路径和科学家精神进行研究。国内学者应立足于我国当前科技发展战略的需求,能够充分发挥我国新型举国体制优势,来探讨战略科学家的发现、培养、使用问题。

综上,通过对上述4个分析框架的分析与探讨,可以看出我国学界目前在战略科学家研究领域的侧重点和未来的研究关注点。学者们正逐渐将研究重心转移到培养战略科学家的策略上,强调科学家精神的传播,重视科普教育的作用。最后,模型预测部分主题的研究热度虽然将有所降低,但鉴于其研究价值和研究过程的复杂性、时代性和迫切性,这些主题仍值得继续深入研究。

3 结语

本文采用LDA、SARIMA 2种模型,对战略科学

家这一研究领域进行主题挖掘与实证分析,在利用LDA模型进行热门主题识别的基础上,用预测模型SARIMA展开定量预测,能为单纯通过研读文献进行研究热点和研究发展趋势的分析工作起到补充作用。对LDA模型生成的结果进行分析可知,战略科学家研究领域的主题呈现出6个显著的发展方向并存在周期性。其中,加快培育壮大国家战略人才力量与服务支撑国家战略是目前战略科学家研究领域中受关注度最高的研究主题。战略科学家作为国家战略科技力量的“关键少数”,对承担国家战略科技任务、发挥人才链在四链中的牵引作用、推动战略人才的培育使用等有着至关重要的作用,因而成为研究的热点。此外,研究主题显著的周期性特征推测与国家战略需求的调整、重大科研项目审批与结题或国际环境与合作周期等因素有关,这为战略科学家研究提供了新的研究视角。

对SARIMA模型生成的结果进行分析可知,战略科学家钱学森与中国战略人才培养机理的研究主题最有发展潜力,科学家精神传播与科普教育的媒体角色这一主题的发展前景也十分可观。随着大国科技竞争愈发激烈以及数智新时代的开启,我国开始致力于探索先进技术、培育战略科技人才,这使得发展和利用数字媒体技术进行科教宣传,借鉴老一辈战略科学家的经验推动战略科学家的培育和使用,将成为战略科学家研究的热点。不过,战略科学家的概念内涵和培育目标因时而新,培养使用适应新时代发展要求的战略科学家,不能局限于探讨老一辈战略科学家的成长路径。

本文对战略科学家这一研究范畴进行了相关文献资料的梳理和分析,但其研究过程并非足够完善且存在一定的限制,具体包括:研究资料限于中国知网、万方、维普等文献数据库的收录范围,可能忽略了其他研究型数据库如中国研究数据服务平台(CNRDS)、国泰安CSMAR数据库等的相关资料,这或许会对研究结论产生一定影响。此外,尽管本研究探讨了关于中国战略科学家研究的热门主题及其发展趋势的问题,但尚未揭示未来全球范围内战略科学家或科技领军人才等研究领域可能呈现的关注点和研究趋势。因此,未来的研究工作将不仅仅扩展资料库的种类,并且要涉及国际层面的相关研究,从而为战略科学家的培养使用提供更加全面和系统的参考借鉴。

参考文献:

- [1] 李婧姝,董贵成. 习近平关于战略科学家重要论述的精髓要义[J]. 科学社会主义,2022(3):48-52.
- [2] LI J Z,DONG G C. The essence and crux of Xi Jinping's significant discourses on strategic scientists[J]. Scientific Socialism, 2022(3):48-52.
- [2] 张赞,王小凡. 我们为什么需要战略科学家[J]. 科技导报, 2022,40(16):14-17.
- [2] ZHANG Y,WANG X F. Why we need strategic scientists[J]. Science & Technology Review,2022,40(16):14-17.
- [3] ZUCKERMAN H. Scientific elite: Nobel laureates in the United States[J]. Isis,1979,70(2):283-284.
- [4] BERRY C. The Nobel scientists and the origins of scientific achievement[J]. The British Journal of Sociology,1981,32(3):381-391.
- [5] HILLEBRAND C D. Nobel century: A biographical analysis of physics laureates[J]. Interdisciplinary Science Reviews,2002,27(2):87-93.
- [6] STEPHAN P,LEVIN S. Age and the Nobel Prize revisited[J]. Scientometrics,1993,28(3):387-399.
- [7] KARAZIJA R,MOMKAUSKAITÉ A. The Nobel prize in physics-regularities and tendencies[J]. Scientometrics,2004(61):191-205.
- [8] COSTAS R,VAN LEEUWEN T N,BORDONS M. A bibliometric classificatory approach for the study and assessment of research performance at the individual level:The effects of age on productivity and impact[J]. Journal of the American society for information science and technology,2010,61(8):1564-1581.
- [9] 杨玉琴,程曦,王国燕. 科学家参与公共决策的角色模型及应用探析[J]. 科普研究,2023,18(5):40-48.
- [9] YANG Y Q,CHENG X,WANG G Y. Analysis of role models for scientists' participation in public policy decision-making and their applications[J]. Science Communication Research,2023,18(5):40-48.
- [10] 芮绍炜,康琪,操友根. 科技自立自强背景下加强战略科技人才培养与梯队建设研究:基于上海实践[J]. 中国科技论坛,2023(9):28-37.
- [10] RUI S W,KANG Q,CAO Y G. Research on strengthening the cultivation and succession building of strategic sci-tech talent under the background of independent and strong technology[J]. China Science and Technology Forum,2023(9):28-37.
- [11] 孙昌璞. 战略科学家培育之我见[J]. 科技导报,2022,40(16):18-26.
- [11] SUN C P. My perspective on cultivating strategic scientists[J]. Science & Technology Review,2022,40(16):18-26.
- [12] 瞿群臻,高思玉,汪鹏飞,等. 基于生存分析视角的战略科学家成长路径研究[J]. 中国科技论坛,2022(7):157-166.
- [12] QU Q Z,GAO S Y,WANG P F,et al. A study on the development path of strategic scientists from a survival analysis perspective[J]. China Science and Technology Forum,2022(7):157-166.
- [13] 潘俊强. 从老干部成才路径看战略人才培养[J]. 国家教育行政学院学报,2023(6):6-9.
- [13] PAN J Q. Cultivating strategic talents from the perspective of the career paths of veteran cadres[J]. Journal of National Academy of Education Administration,2023(6):6-9.

- [14] 冯繁,童杨,闫金定. 关于培养使用战略科学家的思考:基于中外100位战略科学家的履历分析[J]. 科技导报, 2022, 40(16):38-45.
FENG C, TONG Y, YAN J D. Reflections on cultivating and employing strategic scientists: Based on the resume analysis of 100 Chinese and foreign strategic scientists[J]. Science & Technology Review, 2022, 40(16):38-45.
- [15] BYSTROV V, VIKTORIIA N K, ANNA S B, et al. Choosing the number of topics in LDA models: A Monte Carlo comparison of selection criteria[J]. Journal of Machine Learning Research, 2024, 25(79):1-30.
- [16] 陈琦,张君冬,郑婉婷,等. 基于 LDA 模型的中医药人工智能领域主题演化分析[J]. 世界科学技术-中医药现代化, 2022, 24(9):3315-3324.
CHEN Q, ZHANG J D, ZHENG W T, et al. Topic evolution analysis in the field of traditional Chinese medicine artificial intelligence based on LDA model[J]. World Journal of Science and Technology-Modernization of Traditional Chinese Medicine, 2022, 24(9):3315-3324.
- [17] 王鹏,高铖,陈晓美. 基于 LDA 模型的文本聚类研究[J]. 情报科学, 2015, 33(1):63-68.
WANG P, GAO C, CHEN X M. Text clustering research based on LDA model[J]. Information Science, 2015, 33(1):63-68.
- [18] FERNÁNDEZ P M, PRADA C M, LOSADA D E, et al. An unsupervised perplexity-based method for boilerplate removal[J]. Natural Language Engineering, 2024, 30(1):132-149.
- [19] 冯艳铭,郝志梅,董春栎. 基于 LDA 模型的老年人生活满意度主题挖掘与文本实证分析[J]. 华北理工大学学报(社会科学版), 2024, 24(2):19-25.
FENG Y M, HAO Z M, DONG C L. Topic mining and textual empirical analysis of elderly's life satisfaction based on LDA model[J]. Journal of North China University of Science and Technology (Social Sciences Edition), 2024, 24(2):19-25.
- [20] 吕科伟,王国强,姜彩楼. 高水平科技自立自强视角下战略科学家的特质与培养策略[J]. 科技导报, 2023, 41(17):32-39.
LV K W, WANG G Q, JIANG C L. Characteristics and cultivation strategies of strategic scientists from the perspective of high-level self-reliance and self-strengthening in science and technology[J]. Science & Technology Review, 2023, 41(17):32-39.
- [21] 王昉,申金升,武虹,等. 战略科学家典型特征量化评估:探索与实证研究[J]. 中国科学院院刊, 2023, 38(10):1475-1489.
WANG F, SHEN J S, WU H, et al. Quantitative assessment of typical features of strategic scientists: An exploratory and empirical study[J]. Bulletin of the Chinese Academy of Sciences, 2023, 38(10):1475-1489.
- [22] 苏中兴. 全面提高科技人才自主培养质量加快建设国家战略人才力量[J]. 中国高等教育, 2022(23):10-12.
SU Z X. Comprehensively improving the quality of autonomous cultivation of science and technology talent to accelerate the construction of national strategic talent force[J]. China Higher Education, 2022(23):10-12.
- [23] 袁哲,唐恒. 基于社会感知和专利分析的技术发展趋势研究:以新能源汽车为例[J]. 科技管理学报, 2023, 25(5):10-21.
YUAN Z, TANG H. Research on technological development trends based on social perception and patent analysis: A case study of new energy vehicles[J]. Journal of Science and Technology Management, 2023, 25(5):10-21.
- [24] 李少雄,李本光. 基于 SARIMA 模型和 X-12-ARIMA 季节调整方法预测的比较[J]. 统计与决策, 2018, 34(18):39-42.
LI S X, LI B G. Comparison of predictions based on SARIMA model and X-12-ARIMA seasonal adjustment method[J]. Statistics and Decision, 2018, 34(18):39-42.
- [25] ZHAO X L, KAI Y, XING L, et al. Research on ultrasonic defect imaging based on a neural network with Gaussian weight function fusion model[J]. Construction and Building Materials, 2024, 411:134229.
- [26] 尹西明,陈劲,吴善超,等. 加强国家战略科技力量体系化发展打造国家科技先导能力[J]. 科技中国, 2024(2):6-11.
YIN X M, CHEN J, WU S C, et al. Promoting systematic development of national strategic sci-tech forces to build pioneering national sci-tech capabilities[J]. Science & Technology of China, 2024(2):6-11.
- [27] 习近平. 深入实施新时代人才强国战略 加快建设世界重要人才中心和创新高地[J]. 求是, 2021(24):4-15.
XI J P. Deeply implementing the talent strong country strategy in the new era to accelerate the construction of world's important centers for talent and innovation[J]. Qiushi, 2021(24):4-15.
- [28] 付文婷. 推进科普智库建设助力新时代科普高质量发展[J]. 科普研究, 2022, 17(5):22-25.
FU W T. Promoting the construction of popular science think tanks to facilitate high-quality development of popular science in the new era[J]. Research on Science Popularization, 2022, 17(5):22-25.
- [29] 岳丽欣,刘自强,胡正银. 面向趋势预测的热点主题演化分析方法研究[J]. 数据分析与知识发现, 2020, 4(6):22-34.
YUE L X, LIU Z Q, HU Z Y. Research on hot topic evolution analysis method oriented to trend forecasting[J]. Data Analysis and Knowledge Discovery, 2020, 4(6):22-34.
- [30] 程龙,姜春. 心理需要满足感对科技领军人才公共服务动机的影响机制:社会支持的调节作用[J]. 科技进步与对策, 2024, 41(5):119-128.
CHENG L, JIANG C. The impact mechanism of psychological need satisfaction on public service motivation of science and technology leaders: The moderating role of social support[J]. Science and Technology Progress and Countermeasures, 2024, 41(5):119-128.
- [31] 姜玉平. 钱学森:杰出的战略科学家[J]. 科技导报, 2022, 40(16):72-81.
JIANG Y P. Qian Xuesen: An outstanding strategic scientist[J]. Science & Technology Review, 2022, 40(16):72-81.
- [32] 黄涛. 战略科学家是如何炼成的:以钱学森为例[J]. 中国科学基金, 2010, 24(2):87-90.
HUANG T. How strategic scientists are made: A case study of Qian Xuesen[J]. China Science Foundation, 2010, 24(2):87-90.
- [33] 何林,陈朝晖. 大力弘扬科学家精神实现高水平科技自立自强[J]. 党建, 2024(1):39-41.
HE L, CHEN C H. Vigorously promoting the spirit of scientists to

achieve high-level self-reliance and strength in science and technology[J]. Party Building, 2024(1):39-41.

[34] 张丹丹, 王小明. 战略科学家研究述评及未来研究议题展望[J]. 中国科技论坛, 2024(3):142-149.

ZHANG D D, WANG X L. A review of studies on strategic scien-

tists and prospects for future research agenda[J]. China Science and Technology Forum, 2024(3):142-149.

[编辑:刘素菊]

Hotspot perspectives and future directions in the research field of strategic scientists based on LDA-SARIMA model

HUANG Tao, ZOU Zhecan

(College of Law and Economics, Wuhan University of Science and Technology, Wuhan 430065, China)

Abstract: In the contemporary epoch, where technological prowess is increasingly intertwined with national competitiveness, strategic scientists emerge as pivotal figures in the realm of science and technology. These ‘scientific and technological talents with leadership’, a term that denotes their extraordinary intellectual capital and innovative capabilities, constitute a cornerstone in China’s quest for high-caliber technological autonomy and progressive enhancement. To further strengthen this foundation, there is a pressing need to systematically investigate the thematic foci and evolutionary trends characterizing research on strategic scientists, thereby informing and enriching the design and implementation of scientific and technological policies. This study embarks on a rigorous analytical journey by harnessing the combined power of Latent Dirichlet Allocation (LDA) topic modeling and Seasonal Autoregressive Integrated Moving Average (SARIMA) prediction analysis. LDA, celebrated for its efficacy in unearthing latent thematic structures from vast corpora, is synergistically paired with SARIMA, a statistical forecasting tool adept at identifying and predicting cyclical patterns in time-series data. This multidimensional approach enables a comprehensive exploration and empirical analysis within the dynamic landscape of strategic scientist research. The application of these methodologies uncovers six cardinal themes that encapsulate the essence of inquiries related to strategic scientists. These themes span across realms such as the formation of a national strategic talent cadre, emphasizing not merely the execution of strategic national assignments by these scientists but also their pivotal role in architecting a robust strategic talent infrastructure. This underscores the recognition of strategic scientists as linchpins in the strategic planning and execution of national objectives, catalyzing advancements in technology and fostering an ecosystem conducive to sustained innovation. A pivotal and helpful finding from the SARIMA analysis reveals a discernible cyclical pattern in the research trends surrounding strategic scientists, indicating fluctuations in scholarly focus and resource allocation over time. This insight accentuates the necessity for anticipatory strategies that can maintain a steady momentum in research and development efforts, averting potential downturns during periods of waning interest. Prospective avenues of research, as illuminated by our analysis, point towards the innovative application of predictive models to identify, nurture, and optimally deploy strategic scientists. This forward-thinking approach seeks to leverage the quantitative forecasting capabilities of such models to anticipate talent requirements, thereby enabling a proactive and targeted approach to talent cultivation. Furthermore, the integration of new media technologies in disseminating scientific knowledge and fostering public education is posited as a transformative trend, capable of amplifying the societal impact of scientific breakthroughs and nurturing a culture of scientific literacy and innovation. In the quest for refining training systems tailored for strategic scientists, the paper concludes by advocating for a contemporaneous approach that aligns with the exigencies of the new era. This entails not only adapting educational curricula to reflect the latest advancements and challenges in the global scientific landscape but also fostering an environment that encourages interdisciplinary collaboration, promotes international engagement, and nurtures a spirit of entrepreneurship among strategic scientists. By doing so, China can ensure a steady pipeline of versatile and forward-thinking experts capable of steering the nation’s scientific and technological agenda towards unprecedented heights.

Keywords: strategic scientists; empirical analysis; theme mining; LDA; theme model