● 张 雪[1]，张志强[2,3]，朱冬亮[4]
（1. 西安电子科技大学经济与管理学院，陕西 西安 710126；2. 中国科学院成都文献情报中心，四川 成都 610299；3. 中国科学院大学经济与管理学院信息资源管理系，北京 100190；4. 中国社会科学院图书馆，北京 100732）

基于时间序列分析的潜在学科交叉前沿主题识别研究*

摘 要：［目的/意义］识别学科交叉前沿主题并预测其发展趋势，有助于了解学科内部结构，挖掘领域重点部署方向，为未来创新性、突破性研究提供参考。［方法/过程］以美国国家自然科学基金项目及其产出论文分别作为前端、后端数据，首先，从三个维度测度项目学科交叉度，遴选领域学科交叉项目；其次，从主题关注度、新颖度等方面构建研究前沿主题识别指标体系，对学科交叉主题进行二次遴选，满足阈值的即为学科交叉前沿主题；再次，对比时间序列分析模型ARIMA和LSTM主题拟合效果并选择误差最小模型对学科交叉前沿主题进行趋势预测分析；最后，以生物科学领域为例对方法的有效性和可行性进行实例验证。［结果/结论］生物科学领域在纳米生物学技术、全球变化和海洋环境生物学、生物信息学及壶菌病与两栖动物多样性方面有较好发展前景。经专家咨询和已有研究对比分析，该方法可有效识别领域学科交叉前沿主题，并对其未来研究趋势走向有一定参考借鉴。

关键词：学科交叉前沿主题识别；ARIMA模型；LSTM模型；趋势预测
DOI：10.16353/j.cnki.1000-7490.2024.04.020
引用格式：张雪，张志强，朱冬亮．基于时间序列分析的潜在学科交叉前沿主题识别研究［J］．情报理论与实践，2024，47（4）：152-162.

Identifying Potential Interdisciplinary Front Topics Based on Time Series Analysis

Abstract：［Purpose/significance］Identifying the research fronts of interdisciplinary projects and predicting their development trends will help to understand the internal structure of disciplines, explore the key deployment directions, and provide references for future innovative and breakthrough researches. ［Method/process］Based on the data of NSF projects and its output papers, this study firstly measures the interdisciplinarity of the project from three dimensions and selects interdisciplinary projects. Secondly, the research front identification indicators are constructed according to topic attention, topic creativity, etc., and the topics that meet the threshold are the interdisciplinary front topics. Thirdly, comparing the fitting effect of the time series analysis models such as auto regression integreate moving average（ARIMA）and long short-term memory networks（LSTM）, the least error model is selected for trend prediction. Finally, taking the field of biological sciences as an example to demonstrate the effectiveness and feasibility of the method. ［Result/conclusion］In the field of biological sciences, there are good prospects in nanobiological technology, global change and marine environmental biology, bioinformatics, chytridiomycosis and amphibian biodiversity. Through expert consultation and comparative analysis of existing researches, this method can effectively identify the front topics of interdisciplinary researches, and provide some references for its future research trends.

Keywords：interdisciplinary front topic identification；ARIMA model；LSTM model；trend prediction

0 引言

识别学科交叉前沿主题是探测科学生长点、占领科学研究高地的重要手段。学科交叉是当代科学知识发现的主要方式和重要研究范式，各学科知识在扩散渗透中不断催生和孕育着科学生长点，而交叉主题的识别有助于从微观角度挖掘生长点是什么。研究前沿代表了研究领域的最新动态，以及未来发展的关键走向。已有研究表明，学科交叉是研究前沿主题的关键特征之一[1-2]，即研究前沿主题是学科交叉主题的子集，而学科交叉前沿主题是学科交叉主题和研究前沿主题的交集，同时具备交叉性和前沿性。

* 本文为中央高校基本科研业务费专项资金项目（23年新教师创新基金）“基金视角下学科交叉多元测度方法研究”（项目编号：XJSJ23117）和中国科协战略发展部“科技政策、科技咨询跟踪研究”（项目编号：2023070615CG101615）的成果之一。

因此，若在识别出学科交叉主题的基础上进一步遴选具有交叉性质的前沿主题，将更有利于捕捉当前研究主题中的活跃信号。

学者在“学科交叉主题识别”“研究前沿主题识别”方面已进行了一定探索，并取得相应科研成果，但仍有可探索的空间。在数据来源方面，多以论文文献为数据基础。论文是一系列科学活动的中后端记录，因此，以其为数据源识别的研究主题存在一定滞后性。虽然学者开始利用基金项目数据以提高研究主题前瞻价值，但基金视角下的学科交叉主题识别研究仍较匮乏，有待补充完善。在识别方法方面，更多学者将主题的学科交叉性和前沿性分开探讨。尽管学科交叉主题或研究前沿主题在一定程度上能分别反映特定学科领域的焦点和痛点，但学科交叉增大了学科领域研究轨道非线性变化的可能性，进而使得出现创新性研究的概率增大。因此，若将主题的学科交叉性融入研究前沿识别中，更有助于挖掘高价值的研究前沿主题。然而，目前该方面研究相对较少。在前瞻性方面，已有研究更加注重当下研究主题的识别。研究主题是面向未来的、动态的概念，但基于定量方法对研究主题未来演化的趋势预测研究较少，难以满足对主题潜在价值有更高要求的用户需求。

基于此，本文以美国国家自然科学基金会（National Science Foundation，NSF）资助项目及其产出论文为数据基础，首先，从基金项目产出论文、PEC 代码、文本内容三个维度衡量基金项目学科交叉度，进而遴选具有学科交叉特征的基金项目；其次，采用 LDA 主题模型识别学科交叉主题，并根据基金项目研究前沿主题识别指标体系筛选学科交叉前沿主题；再次，将时间序列分析方法应用于学科交叉前沿主题的趋势预测，以识别潜在学科交叉前沿主题；最后，以生物科学领域为例以论证方法的有效性和可行性。

1 相关研究概述

1.1 学科交叉主题识别

学科交叉研究主要集中在“基础研究”“关键技术研究”“跨学科知识挖掘研究”三个层面，其中学科交叉主题识别是跨学科知识挖掘的重要组成部分，是学科交叉研究的主要目的[3]。学科交叉主题指学科交叉度高于一定阈值的主题群，主要采用共词分析、文本挖掘两种方法识别。其中，共词分析是目前使用最多、发展最为成熟的方法，虽衍生出多项不同研究，但核心思想一致，均以不同学科的共有关键词集合为分析对象，不同之处在于识别关键词集合及挖掘关键词集合方法存在差异。进一步可划分为三点：第一，以两个或多个学科为研究对象，通过对两学科交集关键词的聚类分析确定学科交叉主题[4-5]；第二，以单学科为研究对象，通过对“关键词—学科”二模矩阵的社团划分，确定存在交叉关系的学科及具体的交叉点[6-7]；第三，鉴于视角二以参考文献所属期刊学科类别近似代替该领域的学科分布，存在期刊类别并不能完全代表其刊载文章的学科类别这一局限，如徐庶睿等[8]尝试构建领域术语词典，通过匹配领域术语词与主题词典识别具体的交叉主题。文本挖掘可综合考虑文本语法、语义信息，近年来逐渐受到关注。如黄菡等[9]、陈琼等[10]、张斌[11]根据 PhraseLDA 模型、LDA 模型抽取领域主题集合，然后构建学科交叉主题测度指标体系以遴选学科交叉主题；韩正琪等[12]采用 Rao-Stirling 指数遴选学科交叉文献，并采用 LDA 主题模型识别学科交叉研究主题；王卫军等[13]、魏建香等[14]提出基于文本挖掘的学科交叉文献发现模型与知识挖掘模型。

1.2 研究前沿主题识别

自普赖斯[15]将研究前沿概念引入科技领域，学者就其概念内涵、识别方法等进行了多方面探讨。借鉴卢超等[16-17]研究成果，本文将研究前沿界定为正在兴起的（新颖性）、被学术界高度关注的（关注度）、有一定市场或经济潜力（创新性）和学术传播价值（影响力）的研究主题。研究前沿主题识别方法可从定性、定量两大维度展开，定性方法包括德尔菲法、头脑风暴法、专家咨询法等，这类方法操作方便，但客观性较差[18]。随着科技文献的迅速积累，科学计量方法成为定量识别研究前沿的重要手段。其中，引文分析法是研究前沿识别中发展最早、理论基础最扎实、使用最广泛的方法之一，主要从直接引用[19-20]、共被引[21-22]、文献耦合[23-24]三个视角展开。为克服引文分析难以纳入低被引或零被引文献这一缺陷，部分学者将研究视角聚焦于更细粒度的词簇分析，包括词频（突发词）统计[25-26]、共词分析[27-28]两类。此外，随着各类数据库不断完善、机器学习算法兴起，为进一步丰富研究对象、提高主题可读性，研究前沿识别方法从基于词频统计发展至文本聚类算法[29-30]，从基于主观设定一定阈值的高被引文献集发展至提前预测潜在高被引文献集[31-32]，从主题识别发展至多维指标的前沿主题测度[33-34]，新的研究前沿识别方法使得研究主题语义信息更加丰富、识别粒度更加灵活。

1.3 主题演化趋势预测

主题演化趋势指主题在时间维度的变化趋势，可判别不同主题发展轨迹，从而有针对性地配置科技资源。时间序列分析是以统计指标的历史数据为基础，通过建模分析，预测其未来发展趋势。时间序列分析包含两个前提，

一是统计指标以较小的变化演进，不会突然跳跃；二是历史和当前数据可能表征该指标未来变化趋势。由此可见，时间序列分析对于短期和近期指标预测具有重要价值。因主题内容等虽随时间变化，但短时间内大部分词汇不会突然产生或消失，主题外部属性特征变化也是一个循序渐进的过程，故时间序列分析逐渐应用于主题演化趋势预测。目前主题预测中常用的时间序列模型包括计量经济学模型、基于神经网络的机器学习模型，其中，计量经济学模型包括曲线拟合法[35]、自回归模型（AR）、移动平均模型（MA）、自回归移动平均模型（ARMA）及差分自回归移动平均模型（ARIMA）[36-37]等；基于神经网络的机器学习模型包括支持向量机（SVMs）[38]、递归神经网络（RNN）[39]和长短期记忆人工神经网络（LSTM）[40]等。

由上述分析可知，第一，针对学科交叉及研究前沿主题识别的研究成果丰硕，但对学科交叉前沿主题识别研究关注较少；第二，学者更加注重当下研究主题的识别，采用时间序列分析预测主题演化趋势的研究较少；第三，相比于研究前沿识别，学科交叉主题识别多以论文文献为数据，导致基金项目中的学科交叉前沿主题揭示不足。鉴于在学科日益交叉融合的大环境下，识别具有潜在前瞻价值的学科交叉前沿主题成为科技战略情报工作的重要任务，本文拟针对上述不足进行改进。

2 研究设计

本研究的核心问题是如何识别潜在学科交叉前沿主题，按照“获取数据—遴选学科交叉项目—识别学科交叉主题—识别学科交叉前沿主题—遴选具有潜在价值的学科交叉前沿主题”的分析思路逐层拆解问题，研究方案与流程如图1所示，具体研究步骤如下。

2.1 数据获取与预处理

NSF作为美国的科技管理机构，近年来将资助学科交叉基金项目作为优先事项，并鼓励研究人员在研究前沿领域开展学科交叉相关研究。同时，2008年8月，Web of Science（WoS）开始对论文中的基金信息进行索引标注，使得对基金项目资助产出论文的检索与分析成为可能。因此，本文以NSF及WoS为数据源，检索获得领域基金项目及其产出论文数据，并通过格式转换、关键字段提取、缺失字段剔除等数据预处理分别形成待分析的基金项目及产出论文数据集。

2.2 学科交叉基金项目界定与遴选

学科交叉基金项目的遴选是学科交叉前沿主题识别的基础，而如何界定基金项目是否属于学科交叉研究是回答该问题的要义。根据学科交叉及学科交叉度的定义[3,41]，本文认为基金项目的学科交叉度即以基金项目为测度对

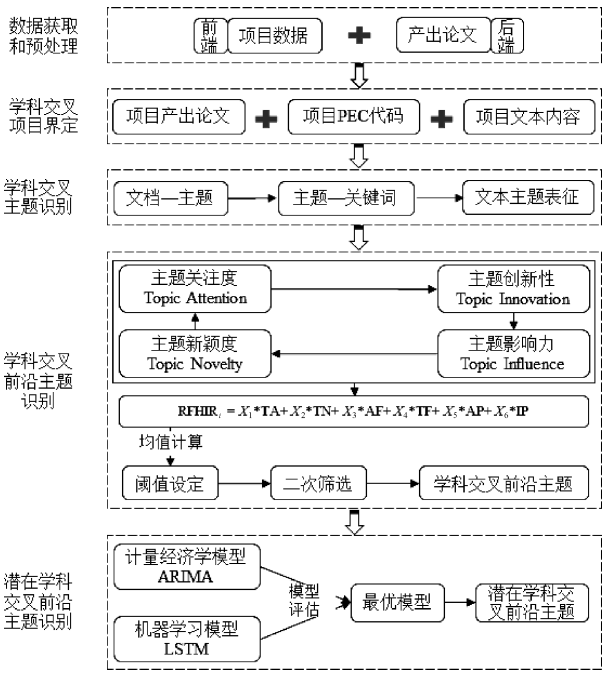


图1 潜在学科交叉前沿主题识别方法框架

Fig. 1 Framework for identifying potential interdisciplinary front topics

象，通过学科交叉测度指标定量分析每个项目参考不同学科的程度。若某项目的学科交叉度高于一定阈值，则为学科交叉基金项目。学科交叉是一个多维、复杂的概念，任何单一维度测度方法都无法全面描述其本质特征。已有研究表明[41-42]，基金项目产出论文作为项目科研成果的主要表现形式，通过论文的引证与被引证关系可以从外在知识融合的角度体现不同学科的理论、方法、技术等输入论文的情况；基金项目PEC代码可以体现每个项目被不同计划合作资助的情况，若这些计划属于不同的研究领域，则可从内在领域合作的角度分析不同学科的计划共同对各个基金项目的资助情况。基于基金项目产出论文及PEC代码的学科交叉测度均是间接、抽象衡量基金项目的学科交叉情况，而基金项目文本内容所传达的信息则体现了不同学科的概念、方法等在基金项目中的具体应用，可以从更加微观的角度捕捉不同学科知识的输入情况。上述三种方法虽均体现了学科交叉是不同专业知识的集成这一概念内涵，但分别从产出论文、研究领域、文本内容三个维度来揭示基金视角下学科交叉的不同或互补特征。为表征不同维度重要性，同时为降低主观赋值法受专家认知影响带来的数据偏差，本文根据不同维度下学科交叉度的离散程度，采用熵值法以不同权值表征不同维度重要性，进而测度基金项目的加权学科交叉度。

具体而言，首先，采用学科丰富性、差异性、多样性等测度指标计算各维度下基金项目的学科交叉度；其次，根据各维度权值，对测度指标进行加权求和；最后，结合

测度结果，遵循被遴选的基金项目学科交叉特征明显，且数据体量满足主题识别的原则，遴选符合条件的基金项目。

2.3 学科交叉前沿主题特征指标体系构建

2.3.1 学科交叉项目文本主题表征 确定学科交叉项目数据集后，需对其进行主题提取、清洗和聚类，然后以主题识别结果为基础，遴选学科交叉前沿主题。鉴于 LDA 主题模型以“项目—主题”“主题—关键词”表征文本主题，便于后续定量计算和分析每个主题内在、外在属性特征，故本文采用 LDA 主题模型提取主题[43]。

LDA 主题模型可增强词间语义关系，但其应用过程中也存在一些不足。其一，首先需确定主题数目，且不同主题数目与最终主题分析效果有直接关系。目前主要采用困惑度（Perplexity）[44]、主题一致性（Topic Coherence）[45]、对数似然性（Log-likelihood）[46]等定量指标确定最优主题数目。其二，传统主题模型识别的主题多由单词表征，可读性、可解释性欠佳。因此，本文基于定量指标，进一步依据专家判读确定主题数目；同时，构建并引入专业词汇字典，整体切分词组，并借助自然语言处理工具包 NLTK 对文本进行去停用词、合并大小写等操作。与仅依据定量指标相比，主题数目更合理；与单个单词相比，词组更能揭示主题内容，可读性较强。

2.3.2 融合内外部特征的前沿主题探测指标 有学者[30,34]根据研究前沿的概念内涵构建不同测度指标，以多维度识别研究前沿。已有指标体系主要包括 4 个关键属性：近期快速增长、激进新颖性、市场或经济潜力及科学性[30]。其中，近期快速增长和激进新颖性用关注度、新颖度来衡量；市场或经济潜力旨在评估主题可能带来的潜在价值，用创新性来衡量；科学性旨在评估主题可能带来的学术传播价值，用影响力来衡量。各指标及其计算方法如下。

1）主题关注度（Topic Attention，TA）。主题受学者重视程度，若某主题被资助基金项目数量逐年递增，即其关注度逐年增长，则表明该主题具有较大的研究潜力。

$$TA_z = \frac{PA_y(z)}{Sum_{PAy}} \tag{1}$$

式中，$PA_y(z)$ 为 y 年与主题 z 相关的被资助项目总数（Project Amount，PA），Sum_{PAy} 为被资助项目总数。

2）主题新颖度（Topic Novelty，TN）。可用某主题下被资助项目的平均资助时间来表示。平均资助时间越晚，说明该主题中近期资助项目占比越大，新颖程度越高，该主题越有可能成为前沿主题。

$$TN_z = \frac{\sum_{P=1}^{Sum_{PA(Z)}} T_P}{Sum_{PA(Z)}} \tag{2}$$

式中，$Sum_{PA(Z)}$ 为与主题 z 相关的被资助项目总数；$\sum_{P=1}^{Sum_{PA(Z)}} T_P$ 为被资助项目资助时间（精确到年）总和。

3）主题创新性（Topic Innovation，TI）。即主题可能带来的潜在价值，如市场或经济潜力，本文使用项目经费及时长表征主题创新性，两个测度指标分别如下：

- 项目经费（Amount of Funding，AF）。某主题下被资助项目经费越多，说明评审专家认为这些项目具有较高的研究价值，有可能产出一批国际领先的原创成果，发展为前沿主题的可能性越高。

$$AF_{y,z,\Delta t} = \frac{\sum_{P=1}^{PA_y(z)} AF_P}{PA_y(z)} \tag{3}$$

式中，$PA_y(z)$ 为 y 年与主题 z 相关的被资助项目总数；$\sum_{P=1}^{PA_y(z)} AF_P$ 为 y 年与主题 z 相关的项目经费总和。

- 项目时长（Time of Funding，TF）。某主题下被资助项目时间越长，说明评审专家认为这些项目具有一定的研究难度，需更长时间探索、攻破，与其他项目相比，发展为前沿主题的可能性越高。

$$TF_{y,z,\Delta t} = \frac{\sum_{P=1}^{PA_y(z)} TF_P}{PA_y(z)} \tag{4}$$

式中，$PA_y(z)$ 为 y 年与主题 z 相关的被资助项目总数；$\sum_{P=1}^{PA_y(z)} TF_P$ 为 y 年与主题 z 相关的项目时长总和。

4）主题影响力（Topic Influence，TI）。即主题可能带来的学术传播价值，基金项目以论文作为其知识形态的主要产出形式，论文数量及其影响力在某种程度上可反映基金项目的资助效果，本文使用项目产出论文数量及其引用影响力来表征主题影响力，两个测度指标分别如下：

- 产出论文数量（Number of Paper，NP）。某主题下被资助项目产出论文数量越多，说明该主题更能被同行专家认可，其前瞻价值也更高。

$$NP_z = \frac{\sum_{P=1}^{Sum_{PA(Z)}} NP_P}{Sum_{PA(Z)}} \tag{5}$$

式中，$Sum_{PA(Z)}$ 为与主题 z 相关的被资助项目总数；$\sum_{P=1}^{Sum_{PA(Z)}} NP_P$ 为被资助项目产出论文总数。

- 产出论文引用影响力（Influence of Paper，IP）。为排除出版年、学科领域与文献类型的影响，科睿唯安推出学科规范化的引文影响力（Category Normalized Citation Impact，CNCI）以应学科学术影响力评价之需，被学术界广泛采用。本文以该指标作为论文学术影响力评价指标。

$$IP_z = \frac{\sum_{P=1}^{Sum_{PA(Z)}} IP_P}{Sum_{PA(Z)}} \tag{6}$$

式中，$Sum_{PA(Z)}$ 为与主题 z 相关的被资助项目总数；$\sum_{P=1}^{Sum_{PA(Z)}} IP_P$ 为被资助项目产出论文的 CNCI 值总和。

基于上述指标，根据分量权重分配方法，构建基于基金项目及其产出论文的学科交叉研究前沿主题识别指标 $RFHIR_i$（Research Front of Highly Interdisciplinary Research），计算公式如下，其中 $X_1 \sim X_6$ 是指标权重。

$$RFHIR_i = X_1 * TA_y + X_2 * TN_i + X_3 * AF_{y,z,\Delta t} + X_4 * TF_{y,z,\Delta t} + X_5 * NP_i + X_6 * IP_i$$

2.4 潜在学科交叉前沿主题识别模型构建

通过上述步骤识别的学科交叉前沿主题是近年来在关注度、新颖度、创新性以及影响力方面均具有明显优势的主题，与一般主题相比，其虽已具备显著的科技政策效力，但根据现有结果仍无法预判这些主题未来是否能延续良好的发展前景，故本文基于各主题历史动态数据采用时间序列分析预测其未来发展趋势。

从方法选取来看，由于 AR、MA、ARMA 模型要求数据符合平稳非白噪声序列，故本文采用适用性更广的 ARIMA 模型。支持向量机等机器学习模型未能考虑输入数值的先后顺序，存在过拟合和优化选择失效等问题，而 RNN 和 LSTM 在计算下个输入特征时，会综合考虑上一步得到的中间特征，可以有效把时间序列关系加入网络结构中。但与 LSTM 相比，RNN 会把很多无关信息考虑进来，使得网络模型整体效果有所下降，而 LSTM 在 RNN 基础上加入控制单元有选择地保留或遗忘部分中间结果，使得模型效果更佳。综上，本研究采用 ARIMA 模型和 LSTM 模型，对主题演化趋势进行预测分析。其中，通过 $RFHIR_i$ 值白噪声检验→平稳性检验→ARIMA (p,d,q) 参数确定→模型拟合效果检验等步骤构建 ARIMA 模型；通过 $RFHIR_i$ 值数据转化→数据集拆分→LSTM 模型训练→$RFHIR_i$ 值预测等步骤构建 LSTM 模型。

因各主题 $RFHIR_i$ 值是以年为单位的连续值，本文使用连续值预测中常见的评估指标平均绝对误差和均方误差对模型拟合效果进行评估，各指标计算公式如下：

平均绝对误差(Mean Absolute Error, MAE) =

$$\frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |$$

均方误差(Mean Square Error, MSE) = $\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

式中，n 表示样本数量；y_i 表示第 i 个样本真实值；$\hat{y}_i$ 表示第 i 个样本预测值。

3 实证分析

3.1 数据获取与处理

目前以“Bio-X”为特色的学科交叉研究实验室、研究团队日益增多，生物科学为许多其他学科的原始研究提供了理论基础，同时其他学科的技术支持推进了生物科学的前沿研究，故分析生物科学领域学科交叉前沿主题及其演化趋势具有实际意义。同时，由于 WoS 的论文基金信息标注开始于 2008 年下半年，因此以 NSF 生物科学学部 2009—2018 年资助项目及其产出论文数据为例进行实证分析，如表 1 所示。

表 1 2009—2018 年 NSF 生物科学领域资助项目及其产出论文

Tab. 1 NSF funding projects and their output papers in the field of biological sciences from 2009 to 2018

年份	资助项目	项目产出论文	去除无参考文献、无摘要的论文
2009	1876	11029	10872
2010	1507	9791	9605
2011	1515	9841	9636
2012	1402	8351	8099
2013	1265	6406	6202
2014	1328	5861	5646
2015	1422	4956	4675
2016	1286	3182	2988
2017	1310	2000	1920
2018	926	821	775

3.2 学科交叉项目识别结果分析

分别计算三个维度下各基金项目的学科交叉度，并根据各维度信息量的不同，通过熵值法确定项目产出论文、PEC 代码、文本内容三个维度的权重分别为：0.2、0.3、0.5，这与已有研究结论相似，即文本内容代表实实在在的观点和内容，可以从微观角度更加详细地剖析不同学科在项目中的具体融合情况；PEC 代码由基金委分配给每个项目，但其动态变化性强，故没有赋予过高权重；对比而言，项目产出论文以参考文献所属期刊学科类别作为学科交叉测度依据，间接、抽象地反映了被引文献的研究主题，故其被赋予权重最低。

根据测度结果，加权学科丰富性、差异性、多样性分布情况如图 2 所示。根据学科交叉项目遴选原则，将加权处理后的学科丰富性、差异性、多样性指标从大到小排序，将三个指标数值均在前 50% 的 3143 个项目作为学科交叉项目。

3.3 学科交叉前沿主题识别结果分析

3.3.1 学科交叉主题识别 对学科交叉项目进行 LDA 主题建模，因主题识别的准确性与主题个数密切相关，故首先计算主题困惑度和一致性指标并绘制趋势图。由图 3 可知，当主题数少于 30 时，曲线较为陡峭，而在 30 之后，困惑度的变化趋于平缓。结合定量指标并综合对比不同主题数目下主题识别结果的可解读性，最终确定主题数目 K

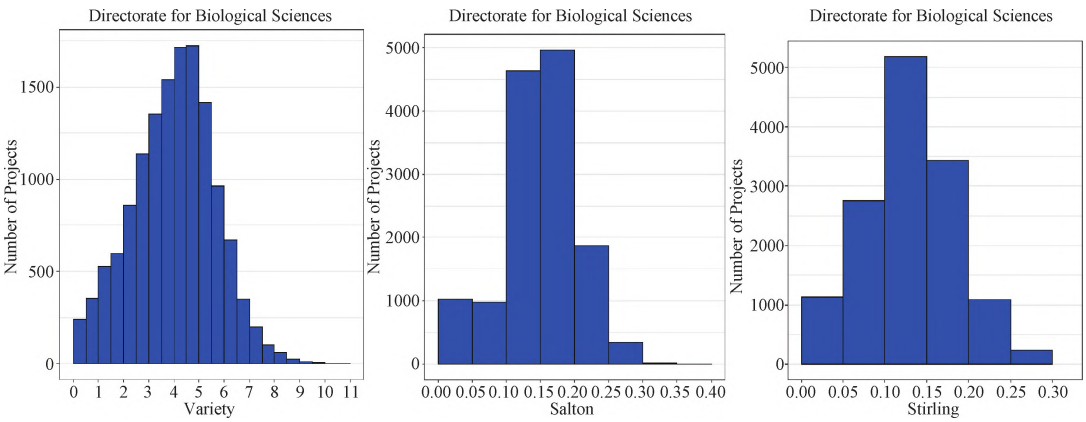


图 2　2009—2018 年 NSF 生物科学领域项目加权学科交叉特征分析

Fig. 2　Analysis of weighted interdisciplinary characteristics of NSF funding projects in the field of biological sciences from 2009 to 2018

为 32。其次，对 sklearn 库中的 Latent Dirichlet Allocation（LDA）函数进行参数设置，如迭代次数为 100，参数 α、η 为默认值 $1/K$，LDA 求解算法 learning_method 为 batch，根据以上参数对预处理后的文本进行主题识别。

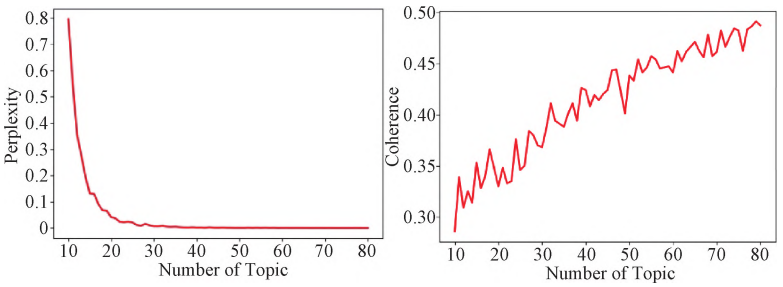


图 3　主题数与困惑度、一致性指标关系曲线

Fig. 3　The relationship curve between the number of topics and perplexity and coherence respectively

3.3.2　学科交叉前沿主题指标计算　基于主题识别结果，首先，获取每个主题下的资助项目及其产出论文；其次，分别计算每个主题关注度、新颖性、创新性、影响力等特征指标，部分计算结果如表 2 所示；最后，分别计算各指标均值，若某主题各指标均高于均值，则其为学科交叉前沿主题，识别出符合条件的主题共 11 个。

3.4　潜在学科交叉前沿主题识别结果分析

3.4.1　预测模型构建与误差检验　因识别出符合条件的主题共 11 个，限于篇幅，无法一一解读，故以 Topic5 的时间序列模型拟合流程为例，对其具体阐述。

以 2009—2018 年逐年 $RFHIR_i$ 值为待分析序列，对于 ARIMA 模型，首先，分别采用 LB 统计量、ADF 单位根检验逐年 $RFHIR_i$ 值是否为白噪声及平稳序列，以保证 Topic5 逐年 $RFHIR_i$ 值符合 ARIMA 模型输入数据特征；其次，根据逐年 $RFHIR_i$ 值，确定 ARIMA(p,d,q) 模型参数，具体根据 ADF 检验，结合 ACF 和 PACF 图，并借助 AIC 值，确定模型最优参数为 ARIMA（0，1，2）；最后，采用残差图、残差正态性检验、自相关检验评估模型拟合效果。结果显示，残差图不存在趋势及异常，Normal Q-Q 图为正态分布，Box. test 残差自相关检验中 $P=0.75>0.05$，说明 ARIMA(0,1,2) 模型对 Topic5 的时间序列拟合较好。

对于 LSTM 模型，本研究采用 TensorFlow 框架和 Keras 模块进行训练。首先，将 2009—2018 年逐年 $RFHIR_i$ 值划分为训练集和验证集。由于数据集样本有限，故设定模型训练的 epochs = 50，batch_size = 1，verbose = 2。同时，为加快收敛速度，对数据进行归一化处理。

重复上述流程，分别对每个主题的 $RFHIR_i$ 值进行时间序列拟合。同时，为评估模型预测性能，采用 MAE 和 MSE 指标分别计算模型预测精度，结果如表 3 所示。根据表 3 可知，ARIMA 模型预测准确率更高，预测误差相对较小。因此，本文选择 ARIMA 模型对主题进行预测和趋

表 2　生物科学领域学科交叉主题前沿指标计算结果（部分）

Tab. 2　Results of research front indicators of interdisciplinary topics in the field of biological sciences (part)

主题		Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
主题关注度		0.01	0.01	0.09	0.08	0.06	0.03	0.02	0.01	0.03
主题新颖度		2011.58	2013.61	2013.38	2012.58	2013.23	2012.98	2012.91	2012.37	2014.22
主题创新性	年均经费	450700.39	401976.29	433045.40	664821.02	469730.39	662834.61	437124.35	522393.33	192262.99
	年均时长	4.01	3.75	3.52	3.72	3.69	3.55	3.25	3.83	2.43
主题影响力	篇均论文	7.83	10.53	8.78	9.91	6.56	11.37	9.24	8.89	6.43
	篇均 CNCI	2.27	1.72	1.91	1.40	1.21	1.65	2.17	1.41	1.66

表3 ARIMA模型和LSTM模型预测误差对比分析
Tab. 3 Comparative analysis of prediction errors between ARIMA model and LSTM model

Topic	ARIMA模型		LSTM模型	
	MAE	MSE	MAE	MSE
#2	1.10	0.17	1.09	0.17
#3	0.96	0.14	0.94	0.15
#4	0.75	0.10	0.98	0.18
#5	0.59	0.08	0.67	0.10
#10	1.23	0.21	1.17	0.22
#12	0.65	0.11	1.29	0.24
#17	0.91	0.13	0.91	0.15
#23	0.93	0.15	1.00	0.20
#25	0.72	0.17	0.85	0.15
#26	0.62	0.06	0.91	0.15
#27	0.88	0.12	1.00	0.14
平均预测误差	0.85	0.13	0.98	0.17

势分析。

3.4.2 潜在学科交叉前沿主题识别与结果解读 采用ARIMA模型对11个主题进行拟合和趋势预测，结果如图4所示。根据主题的变化趋势可知，主要有4种变化趋势。

1）呈明显上升趋势的主题。①Topic#3（纳米生物学技术）：纳米材料因其独特的物理和化学性质在生物科学领域应用广泛，目前研究主要包括纳米微粒细胞分离；纳米粒子细胞染色；纳米微粒药物输送；发展新一代纳米检测技术与设备；基于纳米酶构造纳米机器人；合成生物大分子；细胞结构特性研究等[47-48]。这些研究为肿瘤诊断和治疗、组织修复、器官替换、免疫诊断、药物检测等提供新思路。②Topic#12（全球变化和海洋环境生物学）：海洋酸化和气候变暖等影响海洋生物多样性，研究环境变化如何导致物种灭绝是研究焦点。目前研究以海洋浮游植物、甲壳类节肢动物等为研究对象，综合数学建模、电生理技术等分析海洋生物对环境变化的生理反应（包括听觉等形态学和神经感觉功能）；幼虫发育过程中遗传变异的生理基因组表观表现等[49]。这些研究可预测海洋酸化对生态系统功能影响，为水产养殖和可持续食物资源等提供参考。③Topic#26（生物信息学）：对基因组信息的管理和分析催生了生物信息学发展，目前研究主要包括建立生物数据库，如利用无人机和计算机视觉等技术，获得稀有物种在群体中的运动轨迹；对数百万个基因序列进行测序；膜和膜蛋白建模；采用机器学习方法对建模数据进行预测分析等[50-51]。这些分析有助于识别与特定疾病相关的基因，也为科学研究、监测和管理动物提供参考。④Topic#27（壶菌病与两栖动物多样性）：壶菌病已经并将继续造成两栖动物生物多样性巨大损失，目前研究主要包括壶菌病致病菌是什么；其如何在宿主动物中传播；宿主动物免疫遗传学研究；如何与蛙壶菌抗衡（主要组织

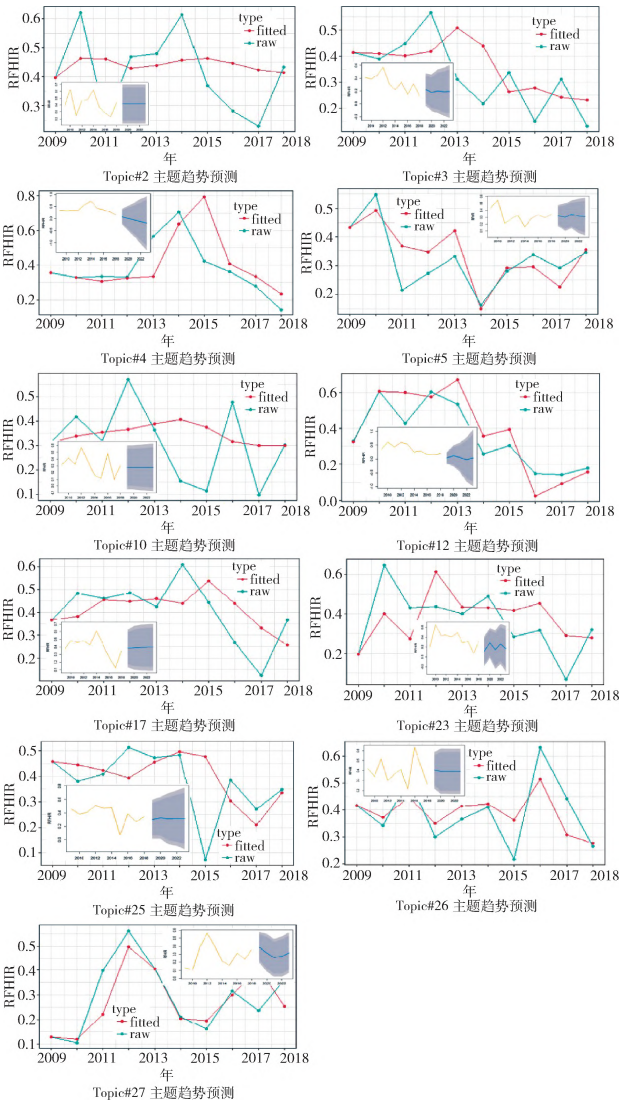


图4 学科交叉前沿主题趋势预测结果
Fig. 4 Prediction results of interdisciplinary front topics

相容性复合体（MHC）变异性）；不同个体病菌感染差异等[52]。这些研究为保护两栖生物多样性提供参考，同时也有助于预测尚未接触这类疾病的其他种群或物种流行病学情况。

2）基本保持平稳趋势的主题。①Topic#2（微生物多样性与生态系统功能）：微生物控制大部分生物化学循环，且决定生态系统对环境变化的响应，目前研究主要包括微生物多样性对环境变化、植物生长力、植物凋零等影响；全球气候变化对生物多样性和生态过程影响；建立气候变化、生物多样性和生态系统功能长期数据库；整合生物多样性信息来源，开发新计算方法描绘生物多样性组成结构等[53]。量化微生物功能多样性为预测生态系统对全球变化响应提供参考。②Topic#10（古生物学）：古生物学研究目的之一为探索生命起源及其与环境的协同演化过程，

目前该方面研究的重点是综合生态学、水文学和社会学观点和技术，主要研究内容包括根据近百年来的森林中化石、沉积物、树木年轮、土壤等记录数据预测森林对环境变化的响应轨迹；生物灭绝规模和速度与环境变化之间关系等[54]。这些研究为协调生物变迁、灭绝、复苏等与生态系统平衡提供参考。③Topic#17（生物进化谱系结构）：分析微生物等物种进化过程并测量其基因和生态功能变化率对了解细胞可塑性具有重要意义。目前研究以化石物种、活物种的蛋白质编码核基因为数据源，综合数学建模、统计推断等方法分析其谱系结构；使用下一代测序方法预测物种亲缘关系；了解基因组多样化及新陈代谢模式和过程等[55]。以此解开生命早期进化和谱系秘密，为保护和管理生物多样性提供数据。④Topic#25（根瘤菌共生固氮机制）：氮是植物所需的最重要营养物质之一，在农业系统中通常用作肥料。目前研究主要包括分析豆科植物结瘤固氮原因及参与其中的细胞分子模块或基因；共生固氮菌与宿主植物间相互作用的遗传学关系；宿主植物细胞调控根瘤菌侵染机制；根瘤菌在根瘤细胞中分化过程等[56]。研究结果可为丰富土壤与最佳根瘤菌种类、减少肥料需求提供指导。

3）呈明显下降趋势的主题。①Topic#4（神经元动力学机制）：动物在不同环境中产生不同行为是行为科学长期谜团之一，目前研究主要采用数学建模、功能成像等方法，研究内容包括探究神经元细胞如何产生呼吸、运动等节律性运动行为；揭示特定神经元在获得特殊功能方面作用；改进记录大脑神经元信息的材料仪器等[57]。这些研究有利于新材料开发和理解动物社会性，从而可用于特殊疾病治疗等。②Topic#5（病原微生物鉴定与溯源）：传染病被列为导致物种灭绝第六大原因，研究宿主与寄生虫相互作用及其对宿主健康和疾病传播的影响至关重要。目前研究融合生态学、化学和植物生理学等学科，建立包括宿主运动、传播途径、病原体毒力和宿主抗性在内的流行病学模型，分析寄生虫在宿主内传播过程；预测疾病在生态系统中传播动态；分析影响寄生虫传播因素等[58-59]。这些研究为制定应对流行病相关对策提供参考。

4）没有形成稳定的上升或下降趋势的主题。Topic#23（影响种群数量变化因素）：生物及非生物因素对植物物种、群落和植被类型分布的影响是生态学长期关注话题。目前研究采用卫星图像、生物标记物等方法分析生物体对气候变化如何做出反应、为什么反应以及对哪些环境压力做出反应；测试微生物共生体是否能调控气候变化对生态系统功能影响；预测成功入侵者的空间生长模式等[60-61]。这些研究为了解全球变暖对现代生物群影响进而促进物种保护提供信息。

通过对交叉前沿主题在未来 1～5 年内的预测分析，可以看到：在遴选出来的 11 个交叉前沿主题中，Topic#3、Topic#12、Topic#26、Topic#27 的热度将继续保持持续上升状态，与 Topic#4、Topic#5 相比，更加具有潜在前瞻价值，更有可能成为引领生物科学领域发展的研究方向。这些主题也更值得科研工作者和领域决策者的关注，以寻求重大科学问题和新的科学突破点。

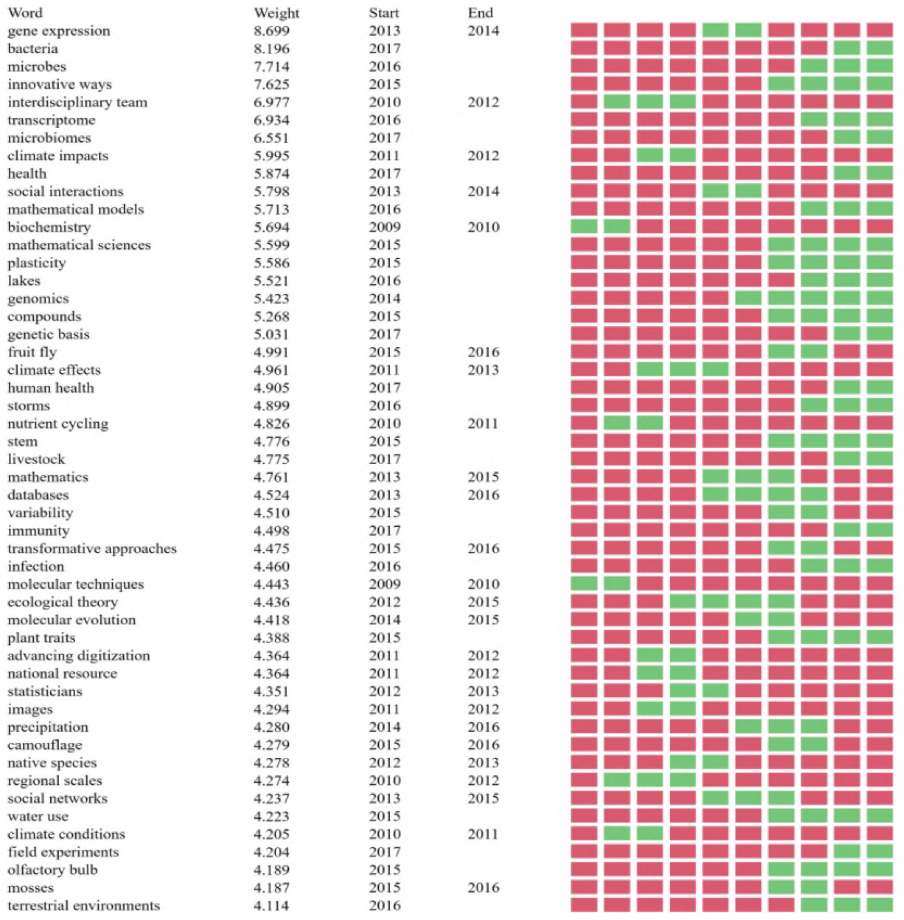
3.4.3　结果验证　采用专家咨询、已有研究及方法对比评估并验证本方法的有效性。

1）领域专家咨询。向吉林大学、华中科技大学生物科学领域相关专家咨询本研究识别结果的可靠性，专家认为本研究结果可有效揭示生物科学领域研究前沿主题，且相关主题均具有学科交叉性；趋势预测结果也为前沿主题未来发展方向提供一定参考和借鉴。

2）已有相关研究。静发冲等[62]指出生物科学新兴前沿主题包括气候变化、海洋酸化、微生物、复杂生物系统等 9 个方面，与本研究识别结果 Topic#2、Topic#5、Topic#12、Topic#23 相吻合；喻亚静[63]指出微生物和免疫学是热点中的热点，与本研究识别结果 Topic#2、Topic#5、Topic#25、Topic#27 相吻合；周群等[64]将生物科学研究前沿分为艾滋病及免疫系统、神经退行性疾病、流行病、技术应用与更新、药物检测和作用机理 5 大方向，与本研究识别结果 Topic#3、Topic#4、Topic#25、Topic#26、Topic#27 相吻合。通过对比分析以往研究，其结果在一定程度上验证了本文识别结果的可信性。进一步地，对主题未来演化趋势进行预测分析，筛选出更具前瞻价值的主题，是对已有结果的补充和完善。

3）基于突发词监测的研究前沿识别。为更全面评估本方法的可信性和有效性，本文选择研究前沿识别中的代表性方法“突发词监测”进行对比分析。突发词指一定时段内频次突发性增长的词或词组，可反映领域动态变化，从而捕捉潜在的研究前沿主题。本文采用 SCI2 中嵌入的“Burst Detection”，图 5 展示了突现强度最高的 Top 50 关键词。其中，Start 和 End 分别指突现词出现的最早和最晚时间；Weight 指突现强度，强度越高，可信度越高。

如图 5 所示，识别结果中如“climate impacts”“microbes”“genomics”“databases”等词与本研究识别结果一致，均反映了生物科学领域在气候变化、生物信息学等方面的新发展。但识别结果为孤立的单个词或词组，无法捕捉到词与词之间的关联关系，使得结果的可读性、可解释性较差。与之相比，本文提出的方法不仅直观计算出主题的 $RFHIR_i$ 值，还囊括各主题包含的主要关键词及其关联关系。进一步地，对主题未来演化趋势进行预测分析，



Word	Weight	Start	End
gene expression	8.699	2013	2014
bacteria	8.196	2017	
microbes	7.714	2016	
innovative ways	7.625	2015	
interdisciplinary team	6.977	2010	2012
transcriptome	6.934	2016	
microbiomes	6.551	2017	
climate impacts	5.995	2011	2012
health	5.874	2017	
social interactions	5.798	2013	2014
mathematical models	5.713	2016	
biochemistry	5.694	2009	2010
mathematical sciences	5.599	2015	
plasticity	5.586	2015	
lakes	5.521	2016	
genomics	5.423	2014	
compounds	5.268	2015	
genetic basis	5.031	2017	
fruit fly	4.991	2015	2016
climate effects	4.961	2011	2013
human health	4.905	2017	
storms	4.899	2016	
nutrient cycling	4.826	2010	2011
stem	4.776	2015	
livestock	4.775	2017	
mathematics	4.761	2013	2015
databases	4.524	2013	2016
variability	4.510	2015	
immunity	4.498	2017	
transformative approaches	4.475	2015	2016
infection	4.460	2016	
molecular techniques	4.443	2009	2010
ecological theory	4.436	2012	2015
molecular evolution	4.418	2014	2015
plant traits	4.388	2015	
advancing digitization	4.364	2011	2012
national resource	4.364	2011	2012
statisticians	4.351	2012	2013
images	4.294	2011	2012
precipitation	4.280	2014	2016
camouflage	4.279	2015	2016
native species	4.278	2012	2013
regional scales	4.274	2010	2012
social networks	4.237	2013	2015
water use	4.223	2015	
climate conditions	4.205	2010	2011
field experiments	4.204	2017	
olfactory bulb	4.189	2015	
mosses	4.187	2015	2016
terrestrial environments	4.114	2016	

图5 基于SCI2的研究前沿识别结果

Fig. 5 Research fronts identification results based on SCI2

将前沿主题划分为不同类型，有助于遴选前瞻主题，体现了本方法的优越和创新之处。

4 结论与展望

本文以NSF的2009—2018年生物科学领域项目及其产出论文为研究对象，首先，根据加权学科交叉测度结果遴选学科交叉项目；其次，根据研究前沿探测指标识别学科交叉前沿主题；最后，采用ARIMA模型识别潜在学科交叉前沿主题。主要研究结论如下。

1）相较于目前研究中常用的学科交叉单维测度方法，加权学科交叉度综合基金项目产出论文、PEC代码、文本内容，能多维度地捕捉基金项目的学科交叉度，是遴选学科交叉项目的有益尝试。

2）学科交叉前沿主题是学科交叉主题与研究前沿主题的交集，在主题交叉性、关注度、新颖度、创新性、影响力等方面具有显著特征。与一般主题相比，更有利于推动学科的创新发展，促进综合性复杂性问题的解决。因此，对该类主题的识别具有重要的现实意义。

3）主题是面向未来的、动态的概念，时间序列分析方法能直观、有效地预测主题未来演化趋势，以遴选具有潜在前瞻价值的学科交叉前沿主题。而此类研究主题不仅能满足对主题潜在价值有更高要求的用户情报需求，对研究人员实时把握学科发展动态，决策者合理配置科技资源也具有重要意义。

本研究也存在一定不足，首先，LDA主题模型识别出的主题中掺杂单个词汇，对主题的解读造成困难；其次，研究前沿最常见的共同特征是它有可能改变并为人们对某一问题的认知注入新的理解，从某种程度来说，研究前沿并不是完全可计量的，本文构建的指标体系未能有效涵盖和表征学科交叉前沿主题的所有特征；最后，时间序列分析更适用于长周期序列数据的拟合，由于数据量有限，识别结果易受噪声数据影响，其准确性有待进一步提高。因此，研究团队会在后续研究中对以上问题进行进一步探讨。□

参考文献

[1] European Research Council. Measuring frontier research in the ERC research proposals: effect on selection outcome [EB/OL]. [2022-06-30]. https://erc.europa.eu/sites/default/files/content/events/P5_DBF_Measuring_frontier_research_in_ERC_proposals-Effect_on_selection_outcome.pdf.

[2] 吴菲菲，杨梓，黄鲁成. 基于创新性和学科交叉性的研究前沿探测模型——以智能材料领域研究前沿探测为例[J]. 科学学研究，2015，33（1）：11-20.

[3] 章成志，吴小兰. 跨学科研究综述[J]. 情报学报，2017，36（5）：523-535.

[4] 闵超，孙建军. 基于关键词交集的学科交叉研究热点分析——以图书情报学和新闻传播学为例[J]. 情报杂志，2014（5）：76-82.

[5] 李长玲，高峰，牌艳欣. 试论跨学科潜在知识生长点及其识别方法[J]. 科学学研究，2021，39（6）：1007-1014.

[6] 岳增慧，许海云，郭婷，等. “情报学”与“计算机跨学科应用”的学科交叉对比研究[J]. 情报资料工作，2016（2）：16-22.

[7] DONG Kun，XU Haiyun，LUO Rui，et al. An integrated method for interdisciplinary topic identification and prediction：a case study on information science and library science [J]. Scientometrics，2018，115（2）：849-868.

[8] 徐庶睿，卢超，章成志．术语引用视角下的学科交叉测度——以 PLOS ONE 上六个学科为例 [J]．情报学报，2017，36（8）：809-820.

[9] 黄菡，王晓光，王依蒙．复杂网络视角下的研究主题学科交叉测度研究 [J]．图书情报工作，2022，66（19）：99-109.

[10] 陈琼，朱庆华，闵华，等．基于领域主题的学科交叉特征识别方法研究——以医学信息学为例 [J]．现代情报，2022，42（4）：11-24.

[11] 张斌．交叉学科主题探究：从主题聚类视角 [J]．情报科学，2020，38（10）：49-55.

[12] 韩正琪．基于主题模型的学科交叉主题识别 [D]．北京：中国科学院大学，2018.

[13] 王卫军，姚畅，乔子越，等．基于词嵌入的国家自然科学基金学科交叉知识发现方法——以“人工智能”与“信息管理”为例 [J]．情报学报，2021，40（8）：831-845.

[14] 魏建香，孙越泓，苏新宁．学科交叉知识挖掘模型研究 [J]．情报理论与实践，2012，35（4）：80-84.

[15] PRICE D J D. Networks of scientific papers [J]. Science，1965，149（3683）：510-515.

[16] 卢超，侯海燕，DING Ying，等．国外新兴研究话题发现研究综述 [J]．情报学报，2019，38（1）：97-110.

[17] 张雪，张志强，曹玲静，等．学科领域研究前沿识别方法研究进展 [J]．图书情报工作，2022，66（12）：139-151.

[18] 沙振江，张蓉，刘桂锋．国内技术预见方法研究述评 [J]．情报理论与实践，2015，38（6）：140-144，120.

[19] GARFIELD E. Citation indexes in sociological and historical research [J]. American Documentation，1963，14（4）：289-291.

[20] KLAVANS R，BOYACK K W. Identifying a better measure of relatedness for mapping science [J]. Journal of the American Society for Information Science and Technology，2006，57（2）：251-263.

[21] SMALL H. Co-citation in the scientific literature：a new measure of the relationship between two documents [J]. Journal of the American Society for Information Science，1973，24（4）：265-269.

[22] SMALL H，GRIFFITH B C. The structure of scientific literatures I：identifying and graphing specialties [J]. Science Studies，1974，4（1）：17-40.

[23] KESSLER M M. Bibliographic coupling between scientific papers [J]. American Documentation，1963，14（1）：10-25.

[24] GLÄNZEL W，CZERWON H. A new methodological approach to bibliographic coupling and its application to the national，regional and institutional level [J]. Scientometrics，1996，37（2）：195-221.

[25] KLEINBERG J. Bursty and hierarchical structure in streams [J]. Data Mining and Knowledge Discovery，2003，7（4）：373-397.

[26] CHEN C. CiteSpace II：detecting and visualizing emerging trends and transient patterns in scientific literature [J]. Journal of the American Society for Information Science and Technology，2006，57（3）：359-377.

[27] CALLON M，COURTIAL J P，TURNER W A，et al. From translations to problematic networks：an introduction to co-word analysis [J]. Information（International Social Science Council），1983，22（2）：191-235.

[28] KATSURAI M，ONO S. TrendNets：mapping emerging research trends from dynamic co-word networks via sparse representation [J]. Scientometrics，2019，121（3）：1583-1598.

[29] KONTOSTATHIS A，GALITSKY L M，POTTENGER W M，et al. A survey of emerging trend detection in textual data mining [M] //Survey of text mining. New York：Springer，2004：185-224.

[30] XU Shuo，HAO Liyuan，YANG Guancan，et al. A topic models based framework for detecting and forecasting emerging technologies [J]. Technological Forecasting and Social Change，2021，162：120366.

[31] LEE C，KWON O，KIM M，et al. Early identification of emerging technologies：a machine learning approach using multiple patent indicators [J]. Technological Forecasting and Social Change，2018，127：291-303.

[32] 李欣，温阳，黄鲁成，等．一种基于机器学习的研究前沿识别方法研究 [J]．科研管理，2021，42（1）：20-32.

[33] 白如江，刘博文，冷伏海．基于多维指标的未来新兴科学研究前沿识别研究 [J]．情报学报，2020，39（7）：747-760.

[34] WANG Qi. A bibliometric model for identifying emerging research topics [J]. Journal of the Association for Information Science and Technology，2018，69（2）：290-304.

[35] 白敬毅，颜端武，陈琼．基于主题模型和曲线拟合的新兴主题趋势预测研究 [J]．情报理论与实践，2020，43（7）：130-136，193.

[36] 岳丽欣，周晓英，陈旖旎．基于 ARIMA 模型的信息构建研究主题趋势预测研究 [J]．图书情报知识，2019（5）：54-63，72.

[37] 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.

[38] 李静，徐路路，赵素君．基于时间序列分析和 SVM 模型的

基金项目新兴主题趋势预测与可视化研究［J］. 情报理论与实践，2019，42（1）：118-123，152.

［39］ 朱光，刘蕾，李凤景. 基于LDA和LSTM模型的研究主题关联与预测研究——以隐私研究为例［J］. 现代情报，2020，40（8）：38-50.

［40］ 霍朝光，霍帆帆，董克. 基于LSTM神经网络的学科主题热度预测模型［J］. 图书情报知识，2021（2）：25-34.

［41］ 张雪，张志强. 美国科学基金会资助项目的学科交叉度演化规律及影响研究［J］. 情报理论与实践，2021，44（12）：122-132.

［42］ 杨洁，王曰芬，陈必坤，等. 基金项目学部分部的交叉网络分析——以美国NSF数据中AI领域为例［J］. 情报学报，2022，41（9）：945-955.

［43］ BLEI D M，NG A Y，JORDAN M I. Latent dirichlet allocation［J］. Journal of Machine Learning Research，2003，3（1）：993-1022.

［44］ PIETQUIN，O，HELEN H. A survey on metrics for the evaluation of user simulations［J］. The Knowledge Engineering Review. 2013，28（1）：59-73.

［45］ STEVENS K，KEGELMEYER P，ANDRZEJEWSKI D，et al. Exploring topic coherence over many models and many topics［C］//Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Jeju Island：Association for Computational Linguistics，2012：952-961.

［46］ SBALCHIERO S，EDER M. Topic modeling，long texts and the best number of topics. Some Problems and solutions［J］. Quality & Quantity，2020，54：1095-1108.

［47］ SMITH D M，SIMON J K，BAKER JR J R. Applications of nanotechnology for immunology［J］. Nature Reviews Immunology，2013，13（8）：592-605.

［48］ PINHEIRO A V，HAN D，SHIH W M，et al. Challenges and opportunities for structural DNA nanotechnology［J］. Nature Nanotechnology，2011，6（12）：763-772.

［49］ HONG Haizheng，LI Dongmei，LIN Wenfang，et al. Nitrogen nutritional condition affects the response of energy metabolism in diatoms to elevated carbon dioxide［J］. Marine Ecology Progress Series，2017，567：41-56.

［50］ MIN S，LEE B，YOON S. Deep learning in bioinformatics［J］. Briefings in Bioinformatics，2017，18（5）：851-869.

［51］ PICKETT B E，SADAT E L，ZHANG Yun，et al. ViPR：an open bioinformatics database and analysis resource for virology research［J］. Nucleic Acids Research，2012，40（1）：593-598.

［52］ MARTEL A，BLOOI M，ADRIAENSEN C，et al. Recent introduction of a chytrid fungus endangers Western Palearctic salamanders［J］. Science，2014，346（6209）：630-631.

［53］ DELGADO-BAQUERIZO M，MAESTRE F T，REICH P B，et al. Microbial diversity drives multifunctionality in terrestrial ecosystems［J］. Nature Communications，2016，7（1）：1-8.

［54］ WILLIAMS J W，GRIMM E C，BLOIS J L，et al. The Neotoma Paleoecology Database，a multiproxy，international，community-curated data resource［J］. Quaternary Research，2018，89（1）：156-177.

［55］ PELLETIER F，GARANT D，HENDRY A P. Eco-evolutionary dynamics［J］. Philosophical Transactions of the Royal Society B：Biological Sciences，2009，364（1523）：1483-1489.

［56］ REN Bo，WANG Xutong，DUAN Jingbo，et al. Rhizobial tRNA-derived small RNAs are signal molecules regulating plant nodulation［J］. Science，2019，365（6456）：919-922.

［57］ URWYLER O，IZADIFAR A，VANDENBOGAERDE S，et al. Branch-restricted localization of phosphatase Prl-1 specifies axonal synaptogenesis domains［J］. Science，2019，364（6439）：1-10.

［58］ KAWAI T，AKIRA S. The roles of TLRs，RLRs and NLRs in pathogen recognition［J］. International Immunology，2009，21（4）：317-337.

［59］ HADFIELD J，MEGILL C，BELL S M，et al. Next strain：real-time tracking of pathogen evolution［J］. Bioinformatics，2018，34（23）：4121-4123.

［60］ OZGUL A，CHILDS D Z，OLI M K，et al. Coupled dynamics of body mass and population growth in response to environmental change［J］. Nature，2010，466（7305）：482-485.

［61］ HU Qihou，SUN Liguang，XIE Zhouqing，et al. Increase in penguin populations during the Little Ice Age in the Ross Sea，Antarctica［J］. Scientific Reports，2013，3（1）：1-6.

［62］ 静发冲，李晨英，韩明杰，等. 基于文本挖掘的美国NSF生物科学部新兴前沿项目主题分析［J］. 现代情报，2014，34（12）：107-112.

［63］ 喻亚静. 《2020研究前沿》发布，微生物和免疫学成为热点中的热点［J］. 微生物学报，2020，60（12）：2872-2875.

［64］ 周群，周秋菊，冷伏海. 生物科学研究前沿演进时序分析［J］. 中国科学院院刊，2017，32（4）：405-412.

作者简介：张雪（ORCID：0000-0002-8908-310X），女，1994年生，博士生，讲师。研究方向：科学计量与科技评价。**张志强**（ORCID：0000-0001-7323-501X，通信作者，Email：zhangzq@clas.ac.cn），男，1964年生，研究员，博士生导师。研究方向：科技战略规划，科技政策与管理，科学学，科学计量学与科技评价等。**朱冬亮**，男，1995年生，助理馆员。研究方向：数据挖掘与机器学习。

作者贡献声明：张雪，研究资料和数据的收集、整理与分析，起草与修订论文。**张志强**，提出论文研究思路，参与论文修订。**朱冬亮**，数据处理，参与论文修订。

录用日期：2023－11－13