

基于ELO模型的技术主题热度评估框架研究

陈洪侃, 刘金昌, 步 一

(北京大学信息管理系, 北京 100871)

摘 要 技术主题热度的评估对决策者了解市场动向和技术发展趋势具有重要意义。然而, 传统的热点评估指标或早期技术的弱信号识别方法尚存在热点技术主题识别欠缺前瞻视角、时间间隔阈值调整主观且较困难、学科框架以及粒度需要预先确定、输出结果不易直接辅助情报工作决策等问题。为此, 本文引入了基于ELO (Elo rating system) 模型的期望热度概念, 构建了一种新的技术主题热度评估方法。本文从理论上阐释了基于ELO模型的方法能够较为准确地反映技术主题的热度指标, 并以碳纤维领域的技术主题热度识别为例, 验证了本文方法的可行性与有效性。相比于传统的热度评估方法, 本文提出的技术主题热度评估方法丰富了主题热度的评估内容, 并为决策者提供了更多的情报支持。

关键词 技术主题识别; ELO模型; 信号识别; 主题热度识别; 弱信号识别

Toward a Technical Topic Popularity Evaluation Framework Based on the ELO Model

Chen Hongkan, Liu Jinchang and Bu Yi

(Department of Information Management, Peking University, Beijing 100871)

Abstract: Evaluating the popularity of technical topics is of great significance for decision-makers to understand market and technical development trends. However, extant indicators for evaluating popularity or recognizing weak signals still suffer from four serious issues: a lack of forward-looking perspectives, subjective and challenging adjustment of time interval thresholds, predetermined disciplinary frameworks and granularity, and difficulty in assisting intelligent decision-making with output results directly. To this end, this study introduces the concept of “expected popularity” based on the Elo rating system (ELO) model and constructs a new method for evaluating the popularity of technical topics. First, we theoretically discuss the feasibility of applying ELO methods to these tasks and take the recognition of the popularity of technical topics in the field of carbon fiber as an example to showcase its effectiveness. Compared to existing methods, the method proposed in this paper enriches the evaluation framework and provides decision-makers with more intelligence-level support.

Keywords: technical topic identification; ELO model; signal recognition; topic popularity identification; weak signal recognition

0 引 言

在信息爆炸的时代, 科学研究和技术创新的速

度空前加快, 及时捕捉和理解最新科学技术发展动态是学术界和工业界共同关注的议题。技术热点主题识别^[1-4]和新兴主题识别^[5-7]作为数据驱动的知识

收稿日期: 2024-11-06; 修回日期: 2025-01-27

基金项目: 国家自然科学基金项目“知识单元视角下基于超图表示学习的原创性科技成果识别研究”(72474009); 国家自然科学基金青年科学基金项目“基于因果推断的高学术影响力跨学科团队早期识别研究”(72104007)。

作者简介: 陈洪侃, 男, 2001年生, 硕士研究生, 主要研究领域为科学学、文献计量学、复杂网络; 刘金昌, 男, 2004年生, 本科生, 主要研究领域为文献计量学; 步一, 通信作者, 男, 1994年生, 博士, 助理教授, 研究员, 博士生导师, 主要研究领域为信息计量、科学评价、复杂网络、大数据应用, E-mail: buyi@pku.edu.cn。

发现方法，利用数据分析和机器学习模型，通过挖掘和分析来自新闻、社交媒体、学术论文等多样化数据源的信息，能有效识别和预测社会、经济和技术领域中的关键趋势和事件。技术主题的有效识别不仅加强了决策者对市场动向和技术发展的预见性，而且有助于促进风险管理和策略调整，提升决策的质量和时效性。技术热点主题识别和新兴主题识别工作也是情报工作部门承接的重要任务，旨在为各级决策者提供准确、及时的情报支持。相较于具体领域的专家，情报工作者并不一定专精于任务的具体科学领域，故能在没有“知识负担”（knowledge burden）^[8]的情况下，提供更为客观的情报信息。此外，正因情报工作者没有丰富的领域知识，他们需要鲁棒性更强的技术方法来确保任务的完成。

对于热点主题这一概念，Tseng等^[9]认为，热点主题与上升趋势、新兴趋势基本同义，都代表一段时期内受关注的主题，主要涵盖两类情形：一是本身具有较高热度的主题，二是短期内热度显著提升的主题。新兴主题识别则是在主题识别的基础上，使用与“新颖性”相关的指标量化主题的新兴程度^[10]。如此看来，无论热点主题识别还是新兴主题识别，其核心任务均为对特定主题的信号强度及其动态变化进行量化评估。以热点主题识别为例，一个主题在每个时间段中发表了若干篇相关文献，则意味着其表现出相应的信号强度（如文献数量）^[11]，信号强度较大的主题被视为“较高热度主题”^[12]；通过计算主题在相邻时间段的信号强度变化程度，筛选出信号变化率较大的主题，视其为“近期热度迅速上升的主题”；若该主题的初始信号强度较小，则可视其为“新兴主题”^[13]。

传统的热点主题识别方法主要包括专家调查法^[14]、文献计量法^[15-16]、社会网络法^[17-18]和文本挖掘法^[19-20]等。专家调查法通过征询领域内权威专家的意见收集深度见解，虽然能够获得高质量的观点，但容易受限于专家的主观判断和个人视角，可能缺乏代表性和时效性。文献计量法通过分析主题论文的引用和发表频率等指标，能够客观地识别热点主题，但该方法可能受到发表偏差^[21]的影响，对新兴主题的响应存在明显的滞后性。社会网络法利用网络分析技术揭示领域内的热点主题及主题间的相关关系，但同样对新兴主题的识别存在一定的滞后现象。文本挖掘法使用自然语言处理技术，从大量文献中提取主题并识别热点主题，尽管高效，但

需要较高的技术知识和依赖技术方法的参数调整，技术复杂且对低热度主题的敏感度不足。总的来看，传统的热点主题识别方法主要用于识别具有显著热度的主题，在识别工作上表现出滞后性，那么基于这些方法提供的情报事实就无法满足即时性要求，不利于决策者实现带有预见性的决策。

为了解决传统热点主题识别方法中存在的滞后性问题，学者们借助“弱信号”（weak signal）的思想捕捉主题热度的早期变化规律，从而预测热点主题，即弱信号角度下的新兴主题识别。目前，采用弱信号进行热点主题识别的方法主要有三种^[22]：组合图法、文本挖掘法和机器学习法。组合图法通过计算主题的可见度和扩散度，使用特征值及其变化形成对应的坐标系，根据主题在坐标系中的位置识别弱信号主题。文本挖掘法通过发现词汇出现的规律，从非结构化文本信息中抽取潜在信息。机器学习法则是通过模拟人脑或采用人类的方法，使计算机获得自动学习的功能，进而检测和识别弱信号。总的来说，弱信号识别方法主要是在传统的热点主题识别方法基础之上，设置阈值用于筛选主题以发现“弱信号”，从而实现前瞻性的热点主题识别。

不论使用传统的热点主题方法还是弱信号识别方法来识别热点技术主题，都存在以下四个方面的问题。第一，热点技术主题识别欠缺前瞻视角。传统的热点主题识别方法往往识别出学者们已经意识到的技术主题，无法帮助学者扩大前瞻性的视野。在使用弱信号识别方法解决滞后性问题时，需要注意避免使用被引次数等加重滞后性问题的指标。第二，时间间隔阈值调整主观且较困难。弱信号识别方法中最难设置的参数不是定义“弱信号”的阈值，而是用于计算指标增长幅度的时间间隔阈值设定。若设定时间间隔过小，则增长幅度容易出现剧烈波动；若设定时间间隔过大，则容易忽视真正需要关注的新兴主题。第三，学科框架以及粒度需要预先确定。这两类方法都需要预先确定领域才能找到相应的文献集合，对确定文献集合的检索式精度要求很高，该步骤的偏差将极大地影响最终情报决策。此外，这一问题也不利于跨学科研究热点技术主题的识别。第四，输出结果不易直接辅助情报工作决策。情报决策者需要根据各类指标自行确定阈值，指标阈值的不统一不仅增加了决策者的负担，还可能导致结果的主观性和不一致性。

针对传统热点识别方法在时效性、跨学科可比性上的局限以及弱信号方法对阈值设定的敏感性，

本文提出基于ELO (Elo rating system) 模型的技术主题热度评估框架。该框架通过两项核心设计实现目标: 其一, 将主题的绝对频次转化为动态更新的“期望热度”指标, 捕捉其相对竞争力变化; 其二, 结合历史ELO值与实时数据差异, 识别早期新兴主题与潜在“睡美人”主题。相较于既有方法, ELO方法避免了学科边界预设对结果的影响, 而且通过参数自适应机制 (如滑动窗口平滑、Logistic函数缩放) 降低了时间阈值设定的主观性, 最终实现前瞻性情报挖掘与跨领域热度比较的双重突破。

1 基于ELO模型的技术热度评估方法

为打破传统方法的局限, 本文通过ELO模型思想引入期望热度概念, 提出基于ELO模型的技术主题热度评估框架 (以下简称“ELO方法”), 以度量任意时间段前后技术主题期望热度及其变化情况, 开展热点主题识别工作。ELO模型于20世纪60年代由阿尔帕德·埃洛 (Arpad Elo) 提出, 最早被用于评估棋类双代理对抗中代理的竞技水平, 其核心是动态更新ELO值反映实际能力, 依对局实际与期望结果差异及比赛权重调整, 具有灵活性与广泛适用性^[23]。ELO模型能推广至多代理场景, 借期望与实际排名差异调整ELO值, 为技术主题热度动态评估提供理论支撑; 动态适应性使其成为构建主题热度评估框架的首选, 助力捕捉热度变化, 支持情报决策。ELO值代表的期望热度能动态反映主题热度变化, 有效应对数据动态更新挑战, 而且无须大量历史数据积累即可识别新兴主题, 为技术主题热度评估带来新契机。

在技术主题热度评估时, 主题可分为单个关键词的术语主题和多个关键词的研究主题^[24], 鉴于低频关键词易被忽视的问题, 本文侧重于以单个关键词代表研究主题进行热度评估。热度方面采用主题出现的绝对频次或相对频率作为传统热度指标^[25], 并借助ELO模型转化为期望热度指标。本节将详述ELO模型双代理、多代理的方法细节、扩展, 以及与传统方法的对比, 并总结其优势。

1.1 ELO方法中主题热度的动态更新机制

ELO方法的核心思路是将传统热度指标 (如文献频次) 作为输入数据, 将各主题定义为代理, 按固定时间间隔 (如每月) 进行基于热度的竞争。每

次竞争后, 通过实际热度与期望值的差异动态更新ELO值, 该值即表征主题的期望热度。实际应用中, 根据主题数量选择双代理 (两两对比) 或多代理 (全域竞争) 模型。以双代理模型为例, 两个主题 (A和B) 的ELO值的更新值取决于以下三个部分。

(1) 主题原先的ELO值 R_i 。该值综合表征主题的历史热度水平, 同时表征其期望热度, 并反映两者在静态竞争下的预期胜率关系^①, 具体的期望胜率为

$$E_A = \frac{1}{1 + 10^{(R_B - R_A)/400}} \quad (1)$$

$$E_B = \frac{1}{1 + 10^{(R_A - R_B)/400}} \quad (2)$$

其中, E_A 、 E_B 分别表示主题A、B的期望胜率; R_A 、 R_B 分别表示主题A、B的ELO值。

(2) 实际对局结果 S_i 。两个主题的比赛结果可以根据该主题热度指标相对于总体热度的比值来确定, 即

$$S_i = \frac{\text{Hot}_i}{\text{Hot}_A + \text{Hot}_B}, i \in \{A, B\} \quad (3)$$

其中, Hot_i 表示主题 i 的热度指标; S_i 表示主题 i 的实际胜率。鉴于ELO值具有加权汇聚历史数据的能力, 可选择更大时间窗口下的热度指标, 以此增强方法的鲁棒性。

(3) 对局权重 K 。 K 值越大, 表示越侧重近期热度信号; 反之, 近期热度信号对ELO值的影响越小。此外, 还可用于区分高质量与低质量期刊发文数量对ELO值的影响。

因此, 各主题的期望热度可通过公式

$$R'_i = R_i + K \cdot (S_i - E_i), i \in \{A, B\} \quad (4)$$

计算得出。其中, R_i 表示主题 i 当前的ELO值; E_i 则为主题 i 的期望排名。ELO双代理模型经过推广后, 也可以被应用于多主题竞技的场景, 其核心思想是根据每个主题的期望排名和实际排名的差异来调整其ELO值, 即

$$E_i = \sum_{j=1}^N \frac{1}{1 + 10^{(R_j - R_i)/400}} \quad (5)$$

$$R'_i = R_i + K \cdot (S_i - E_i) \quad (6)$$

其中, N 表示总主题数量。根据实际排名 S_i 与期望排名 E_i 的差异, 结合由比赛重要性决定的参数 K , 计算出主题 i 的新ELO值 R'_i 。

S形曲线是识别技术生命周期阶段的经典模型,

① 最原始期望概率公式由微积分公式计算, 但在实际操作中, 使用微积分公式计算较为复杂, 经统计简化为文中表达式, 其中的任意数值均不可随意调整。

其中Logistic曲线通常被用于模拟语言的变化（如边缘关键词随时间变化的传播速度、传播范围模拟），适用于快速且显著增长的技术生命周期预测^[26]。Logistic函数中的较低值表示该主题和关键词具有稀缺性，研究这些函数值较低的主题和关键词，可以检测到处于技术生命周期早期阶段、尚未显现出明显快速发展迹象的技术的早期微弱信号^[27]。因此，在实证过程中，可以利用S形曲线将主题热度的实际排名进行适当转化。

ELO双代理模型通过历史数据动态生成主题的热度值，为决策者提供兼具直观性与深度的热度评估工具。该模型以权重参数 K 调节历史数据的衰减速率，结合相对热度（也就是实际比赛结果 S_i ）与期望热度（也就是期望胜率 E_i ）的差异（ $S_i - E_i$ ）计算ELO值变化量（ ΔR_i ），既可以量化单一主题的热度演化趋势（如爆发增长或渐进衰退），又能实

现跨领域主题的直接对比。在此基础上扩展的多代理模型通过全域主题竞争排名（式(5)）突破学科边界限制，构建宏观研究热度图谱。主题ELO值的时序曲线通过整合历史与实时数据，揭示技术生命周期的阶段性特征（如潜伏期、爆发期、衰退期），相较于传统频次统计方法，其优势在于消除领域依赖性并提供统一标尺。进一步地，ELO差值分析（ ΔR_i^t 表示主题 i 当前ELO值与 t 时间前的差异）可触发战略预警机制。例如， $\Delta R_i^t > 200$ ，标记为高潜力新兴主题； $\Delta R_i^t < -100$ ，则提示衰退风险。这种融合动态竞争机制与标准化输出的框架，为情报决策提供了兼具时效性、可比性与可解释性的量化支撑。ELO方法的优势将在1.3节详述。

1.2 ELO方法与传统热点主题识别方法对比

图1展示了热点主题识别的全流程，该流程主

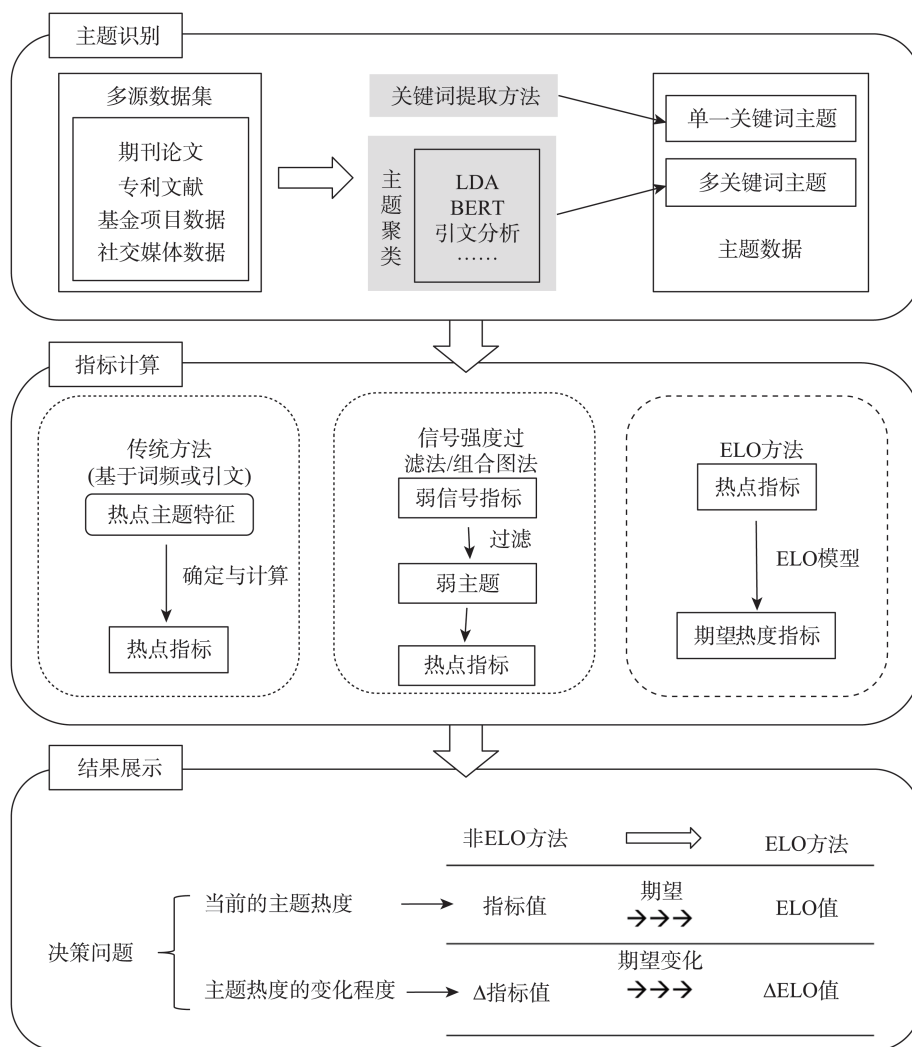


图1 ELO方法与其他方法流程对比

要由主题识别、指标计算和结果展示三个部分组成。在主题识别部分,裘惠麟等^[28]证实了多源数据集对于热点主题识别的有效性;更多的现有研究则是围绕主题识别的方式展开,既有基于关键词出现频率^[29]或结合关键词的多种引文数据分析热点主题^[30]的研究,也有结合LDA(latent Dirichlet allocation)和word2vec^[19]或BERTopic主题建模^[31]等主题聚类形成多个关键词来代表研究主题的研究。鉴于传统弱信号识别方法将单一关键词作为识别对象,在实证部分,本文选取BERTopic主题建模后具有典型性的单一关键词作为候选弱信号主题,以此展开研究。

ELO方法与传统方法在指标计算和结果展示方面差异显著。信号强度过滤法和组合图法基于传统方法用弱信号指标筛选后计算热度,ELO方法则借助ELO模型将指标加工为期望数据,其期望热度能精准描述主题热度强度,突破主题当前表现局限,更优地回应决策问题。传统方法下,决策者须先熟知热度指标基准值才能理解主题当前热度,不同领域基准差异致其负担加重;ELO方法根据成熟评级系统采用期望方式,决策者了解一次ELO数值分布即可多场景快速决策,而且ELO值的期望属性在体现当前热度的同时能更好地反映主题热度变化。传统方法需设时间段参数的指标值变化程度(Δ 指标值),ELO方法仅对前后时间节点ELO值做差就能计算出主题热度变化程度。

1.3 ELO方法优势

相较于传统的热点主题或弱信号识别方法,ELO方法的独特之处在于引入“期望”的概念,用期望信号强度来衡量某个时刻研究主题的热度。*Intelligence Analysis for Tomorrow*^[32]一书中的信号检测理论指出了期望在预测中的重要性:“如果某个事件发生的可能性很小,则不应对其进行预测,除非有非常强烈的信号或非常有必要进行分析。”ELO方法在主题热度评估任务中的优势主要有以下四点。

(1) 时效性强。相较于传统方法,ELO方法能够处理大规模的动态数据,可以及时更新主题热度,并且能够在给定的小时间间隔下,实时更新此时预测出的弱信号技术主题。此外,ELO方法避免了传统识别方法中的时间段长度和数量的参数设定问题(热度变化幅度受时间段长短设定影响较大),

因此,采用ELO方法能够有效提高热点主题识别的时效性和鲁棒性。

(2) 弱信号评估完整。ELO方法不仅能识别出新兴的主题,还能发现主题再度兴起或逐渐衰败的信号,即具有能够识别任意变化趋势的早期预警的功能。在弱信号识别方法中,变化频率的平均值是一个非常重要的指标,然而该指标存在两个方面的局限性。第一,倘若在计算的时间段后半程出现新兴主题,其变化率平均值往往被定义为实际变化率的一半,这低估了弱信号的信号强度。即使考虑出现日期,(早期已出现的)“睡美人”型主题依旧会被低估。第二,倘若一个主题在时间段的后半程已经不再呈现高热度趋势,则该指标无法体现时间对热度评估的影响。这两大问题均可通过结合历史信号强度、代表期望热度的ELO值解决。若特定主题绝对频次出现迅速上升或下降的情况,其ELO值不会同步大幅度上升或下降,而是考虑特定主题与其他所有主题的热度相对大小关系,并且相对大小关系不能简单地通过特定主题的相对频率来衡量^①。ELO方法在评估主题热度时,不仅考虑主题本身的热度,而且考虑了其相对于其他主题的热度情况。通过ELO值的变化程度,还可以评估出主题热度的兴起和衰败程度,从而呈现更完整的主题热度分析。

(3) 具备领域可比性。传统方法预先界定领域范畴的工作较为困难,尤其是研究目标领域为跨学科领域的情况。若领域范畴太大,则容易涵盖一些较大的子领域,从而掩盖了部分热点主题的信号强度;若领域范畴太小,则容易出现主题识别结果不够完整的情况。即使领域范畴被精准且完整地界定,跨学科领域可能是由一个主导性强的子领域与多个较弱的子领域构成,前者的研究主题容易在传统热度指标的评估下占据优势地位,而后者的研究主题被准确识别的概率将下降。此外,传统方法不仅受限于领域界定的准确性,还无法进行跨领域的主题热度比较,特别是当某一领域的研究主题相对集中而另一领域的研究主题较为分散时,两个领域的研究主题的传统热度指标就不具备可比性。相比之下,ELO方法不仅无须依赖领域范畴的界定,还能够比较两个跨学科研究主题的ELO值,从而更适用于跨学科的应用场景。相较于以往依赖于特定研究领域定义的方法,ELO方法具有更广泛的适用

① 假定特定主题和所有主题的总绝对频次保持不变,考虑两种情况:第一,主题数量增多,同时其他主题绝对频次减少;第二,主题数量减少,同时其他主题绝对频次增多。特定主题的相对频率无法区分或判断出相应情况。

性。它不依赖于领域范畴的界定,有助于减少情报决策者在接受预定假设时的被动性;同时,其具备的领域可比性使情报决策者在决策过程中能够更主动地发挥作用,发现并挖掘交叉领域的研究热点。

(4) 可解释性强。传统方法需要决策者对领域有较深的了解,当情报工作者展示某个研究主题热度呈上升趋势,并说明其每月的文献数量从10篇增长到1000篇时,决策者无法直观地感受990篇到底意味着多大的增幅。因为不同领域下文献数量存在差异,而且通过领域的整体文献数量进行归一化将会受到不同领域主题数量差异的影响,更不能适应跨学科领域。传统方法主要呈现的绝对数值、相对数值及其变化率都不能非常直观地向决策者展现数据,需要决策者使用更多的心智进行交叉比较与思考。此外,传统方法中的变化幅度指标非常依赖时间段划分的设定,不同的时间段划分可能会有不同的解读结果,决策者想要了解目标主题在给定时间段前后的热度变化将无从下手。ELO方法通过最新的热度指标信号强度结合历史的预期热度更新,得到该主题最新的预期热度。ELO值更新的过程本质上借鉴“零模型”概念,通过事实与期望的对比,为结果的非随机性提供事实证据,能够助力情报决策者区分偶然性与必然性的结果。代表主题期望热度的ELO值将有助于决策者用更少的心智做出更为准确的决策。决策者既可以看某个主题的历史热点变化趋势,也可以用两个数值展示某个主题在给定时间段前后的热度信号强度,相关信号强度也允许跨领域比较。

2 弱信号视角下的技术主题热度评估验证:以碳纤维领域为例

为验证基于ELO模型的主题热度评估效果,本文以碳纤维领域为对象、以弱信号识别为应用场景,对比ELO方法与两种传统弱信号方法(信号强度过滤法和DoD/DoV组合图法)在弱信号识别方面的效果,用于验证ELO方法的可行性与有效性。

2.1 数据集

本文利用OpenAlex数据集^[33](2022年4月版本)获取碳纤维领域的相关论文数据,对论文的标题和摘要进行匹配,匹配条件如下:①出现与碳纤维相关的基本术语;②出现碳纤维领域特定技术术

语缩写,并同时出现与碳纤维领域相关的其他关键词(由初步文献调研确定)。具体的匹配表达式为(carbon fiber OR carbon fibre OR carbon composite*) OR (CFRP AND (manufactur* OR product* OR process* OR fabric* OR reinforce*))。最终共检索出79881篇相关论文。在文本预处理方面,本文使用NLTK工具移除检索出的论文文本中的标点符号和停用词,并对文本中的单词进行词干提取,以减少文本中词语的冗余,帮助聚类算法更准确地识别和分类文本中的研究主题。本文通过分析历年碳纤维领域的发文量分布,确定2018年为弱信号识别的应用年份。为了更好地覆盖2018年的弱信号,本文从2017年和2018年发表的论文文本中提取可能的技术主题,使用三种方法筛选出技术主题中的弱信号关键词进行分析。

本文使用BERTopic模型对碳纤维领域的科学文献进行技术主题热度识别,BERTopic是Grootendorst^[34]于2022年提出的一种使用预训练模型的主题模型,结合了多种先进技术,包括UMAP(uniform manifold approximation and projection)降维、HDBSCAN(hierarchical density-based spatial clustering of applications with noise)聚类、SentenceTransformer嵌入以及KeyBERTInspired表示等,主题识别效果较好。具体来说,本文使用SentenceTransformer模型(all-MiniLM-L6-v2版本)将文本转化为高维的语义嵌入,使用UMAP模型将文本嵌入降至5维(参数设置:n_neighbors=10,min_dist=0.0),使用HDBSCAN算法进行聚类分析(参数设置:min_cluster_size=3,cluster_selection_method='eom'),使用KeyBERTInspired模型进行主题表征,每个主题由50个关键词进行表征,以平衡主题表征效果和后续的计算复杂度。最终,BERTopic模型从碳纤维领域识别出486个主题。

2.2 弱信号识别方法

(1) 基于ELO方法的弱信号识别方法(本文方法)。受限于计算资源,本文无法将所有主题涉及的关键词纳入ELO方法。因此,本文从表征主题的所有关键词中筛选出词频大于1的关键词作为研究对象(即技术主题),共计2369个。在ELO值的计算中,这些关键词的ELO值每月都会更新一次^①,每次比赛结果是每个关键词在该时间段内涉及的文档频次。

① 为避免早期文献较少的影响,1999年12月及之前的数据,以及2021年1月及以后的数据分别被合并为一次计算。

本文结合S形曲线对ELO值的更新过程进行以下处理：①采用6个月的平均滑动窗口，以减少弱信号关键词因发表频率较低受到的影响；②对文档频次取对数，以避免极端值的影响；③进行最大最小值归一化，以避免每月发文数量不同带来的影响；④使用Logistic函数（ $k=0.5$ ），以更好地识别出快速且显著增长的关键词；⑤将Logistic函数结果缩放至热度获得的ELO值范围[1500,3000]，并据此计算出关键词*i*的实际排名 S_i 。ELO模型中的各关键词的初始ELO值为1200，参数*K*为128。

为了找到在2018年增长幅度较大的关键词，本文采用两种筛选策略。其一，筛选在2018年内，每个月的ELO值均呈现增长（即环比增长）的关键词（month over month, MoM），共117个；其二，筛选在2018年每个月的ELO值相较于2017年同月都有所增长（即同比增长）的关键词（year over year, YoY），共436个。

图2a和图2b分别展示了MoM和YoY中关键词的ELO值分布。可以发现，关键词的ELO值分布集中在两端，即较低的ELO值（约1000）和较高的ELO值（约2000），其中大多数关键词的ELO值较低。由于多个单词组成的关键词（如“fracture tough”）通常包含更多信息，本文在图2c和图2d中单独展示了这些关键词的ELO值分布。结果显示，这些关键词的ELO值分布与包含单个单词的关键词（如“preload”）结果类似，绝大部分的ELO

值较低。此外，观察到ELO值在1600左右存在一个断点，推测ELO值低于1600的为低频词，高于1600的为高频词。基于这一阈值，本文筛选出了MoM和YoY关键词集合中的弱信号关键词，两个基于ELO方法识别出的弱信号关键词集合分别命名为MoM_weak（93个）和YoY_weak（406个）。

（2）信号强度过滤法^[35-36]（基线方法）。该方法分别计算主题和主题内关键词的信号强度，通过设置合适的筛选阈值获取弱信号关键词。主题信号强度基于Logistic函数，由接近中心性、主题权重和自相关系数等要素衡量。其中，主题的接近中心性使用具体主题与其他所有主题之间海林格距离的平均值表示，主题权重通过对具体主题的连贯性与所有主题连贯性的总和求比值来计算，自相关性则是由具体主题在24个月（2017年和2018年）中相关文献数量序列在第12个月的协方差确定。具体公式为

$$CC(t) = \frac{1}{\sum_i d(t, t_i)} \quad (7)$$

$$W(t) = \frac{Coh(t)}{\sum Coh} \quad (8)$$

$$AC(t) = \frac{Cov(t)_{12}}{Var(t)} \quad (9)$$

其中，*t*代表具体主题； $d(t, t_i)$ 表示主题*t*与其他主题 t_i 之间的海林格距离； $Coh(t)$ 是主题*t*的连贯性； $Cov(t)_{12}$ 表示主题*t*的相关文献数量序列在滞后12

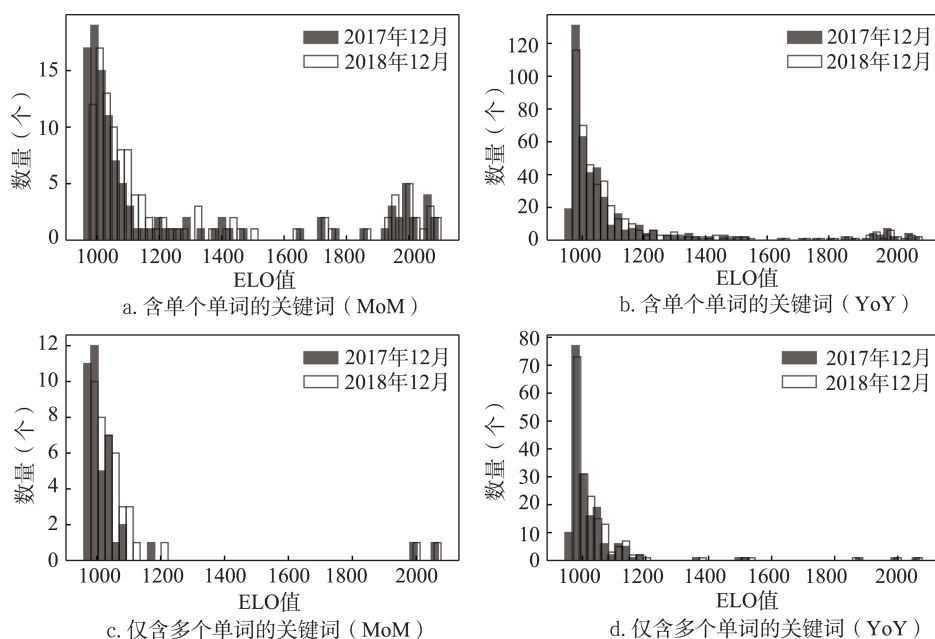


图2 关键词ELO值分布

个月处的协方差; $\text{Var}(t)$ 表示主题 t 的相关文献数量序列的方差。

综合以上三个指标, 定义主题的弱信号函数 $\text{WK}(t)$, 其计算公式为

$$\text{WK}(t) = \frac{W(t) \times \text{CC}(t)}{1 + \exp(-\text{AC}(t))} \quad (10)$$

关键词信号强度的计算依然以 Logistic 函数为理论基础, 通过计算关键词在主题中的归一化频率和 BERTopic 模型获得的主题-关键词概率关系, 建立关键词信号强度的预测函数 $\text{PW}(w)$, 具体计算方法为

$$\text{PW}(w) = \frac{\text{NF}(w)}{1 + \exp(-\phi(w) \times \log(\phi(w)))} \quad (11)$$

其中, w 表示关键词; $\text{NF}(w)$ 是该关键词在对应主题中的归一化频率; $\phi(w)$ 是从 BERTopic 模型中获得的该关键词在主题中的概率。

在对比验证中, 本文设置信号强度过滤法的主题阈值为[5%,10%]^[36], 并从弱信号主题中过滤出信号强度介于[5%,30%]的关键词, 最终获得 766 个弱信号关键词, 后续对比验证中称该集合为 weak signal。

(3) DoD/DoV 组合图法 (基线方法)。这是由韩国学者 Yoon^[37]于 2012 年提出的一种弱信号识别方法, 经过后续学者的大量应用与优化, 现已成为弱信号识别研究中的重要方法之一^[36,38]。在这一方法中, Yoon 将弱信号视为与关键词相关的新兴话题, 提出弱信号关键词应具备的特征: 出现频率低、增长速度快。基于这些特征, Yoon 定义了两个指标——可见度 (degree of visibility, DoV) 和扩散度 (degree of diffusion, DoD), 用于衡量关键词在特定时间段内的可见程度和扩展程度。

DoV 指标是指在特定时间段内某关键词出现频次占所有关键词频次的比例, 其计算公式为

$$\text{DoV}_{ij} = \frac{\text{TF}_{ij}}{N_j} \times \{1 - t_{\omega} \times (n - j)\} \quad (12)$$

其中, TF_{ij} 表示关键词 i 在时间段 j 内出现的总频次; N_j 表示时间段 j 内文档的总数量; t_{ω} 为时间权重, 通常由专家组确定, 本文设定为 0.05^[37]。

DoD 指标是指在特定时间段内包含该关键词的文档数量占总文档数量的比值, 其计算公式为

$$\text{DoD}_{ij} = \frac{\text{DF}_{ij}}{N_j} \times \{1 - t_{\omega} \times (n - j)\} \quad (13)$$

其中, DF_{ij} 为时间段 j 内包含关键词 i 的文档数量。

在组合图弱信号识别方法的传统研究流程中, 研究者通过计算关键词历年的 DoV 和 DoD, 对各年份 DoV 和 DoD 的平均值进行可视化, 并结合 DoD

和 DoV 年平均变化率, 构建关键词显现图 (keyword emergence map, KEM) 和关键词发布图 (keyword issue map, KIM), 结合关键词的出现频率和增长速度这两个弱信号特征, 从 KEM 图和 KIM 图中选取符合条件的弱信号关键词。

具体地, 本文从 BERTopic 模型的识别结果中筛选出词频大于 1 的关键词, 利用 2009—2018 年的数据^[36], 计算每个关键词历年的 DoV 和 DoD, 筛选出 DoV 和 DoD 时间加权增长率排名前 30% 且文档频率低于平均文档频率的关键词。最终, 识别出 600 个弱信号关键词, 后续对比验证中称该集合为 DoDV。

2.3 结果评价

2.3.1 量化评估

本节从准确率和召回率两个量化指标来分析 ELO 方法的有效性。表 1 展示了不同关键词集合中的关键词数量及其与其他集合的重叠比例。例如, 表中的 20.51% 表示在包含单个单词的关键词集合中, MoM 与 DoDV 的重合关键词占 MoM 的 20.51%。为直观展示 ELO 方法与其他两种方法所得关键词的重合程度, 图 3 使用维恩图展示这三种方法的关键词重合情况, 其中图 3a~图 3d 展示了包含单个单词的关键词重合情况, 图 3e~图 3h 展示了仅包含多个单词的关键词重合情况。结合表 1 中数据的含义可以认为, 表 1 倒数第二、第三列的百分数分别表示以 DoDV 和 weak signal 为基准时, 纵表头所示的集合对应识别方法的准确率; 相应地, 表 1 中每种类型最后两行的百分数分别代表以 DoDV 和 weak signal 为基准时, 横表头所示的各个集合对应识别方法的召回率。

从准确率来看, 无论是以 DoDV 还是 weak signal 为基准, ELO 方法在准确率上均高于其他方法; 从召回率来看, 无论以哪种弱信号识别方法为基准, YoY 和 YoY_weak 相较于其他集合都有较好的表现。这表明 ELO 方法不仅生成的弱信号结果数量较少, 而且更为准确, 在实际应用中, 能够减少不确定性决策带来的尝试风险。此外, 与 MoM/YoY 相比, 经过阈值过滤的低 ELO 值关键词集合 (MoM_weak/YoY_weak) 普遍表现出更高的准确率, 表明本文对 MoM/YoY 中弱信号关键词的判别方法是有效的。

在表 1 仅考虑多个单词关键词的比较中, 发现与含单个单词的关键词结果相比, weak signal 为基准对应的准确率和召回率都有所下降, 而 DoDV 为基准对应的准确率有所上升。结合关键词集合数量

表1 各个关键词集合的关键词数量和重叠比例

类型	集合	MoM(%)	YoY(%)	MoM_weak(%)	YoY_weak(%)	DoDV(%)	weak signal(%)	关键词数量
含单个单词的关键词	MoM	100.00	53.85	79.49	35.04	20.51	26.50	117
	YoY	14.45	100.00	9.40	93.12	27.98	19.50	436
	MoM_weak	100.00	44.09	100.00	44.09	25.81	27.96	93
	YoY_weak	10.10	100.00	10.10	100.00	30.05	18.72	406
	DoDV	4.00	20.33	4.00	20.33	100.00	14.33	600
	weak signal	4.05	11.10	3.39	9.92	11.23	100.00	766
仅含多个单词的关键词	MoM	100.00	51.22	95.12	46.34	31.71	12.20	41
	YoY	11.60	100.00	10.50	98.34	37.02	8.84	181
	MoM_weak	100.00	48.72	100.00	48.72	33.33	10.26	39
	YoY_weak	10.67	100.00	10.67	100.00	37.64	7.87	178
	DoDV	3.96	20.43	3.96	20.43	100.00	8.23	328
	weak signal	2.94	9.41	2.35	8.24	15.88	100.00	170

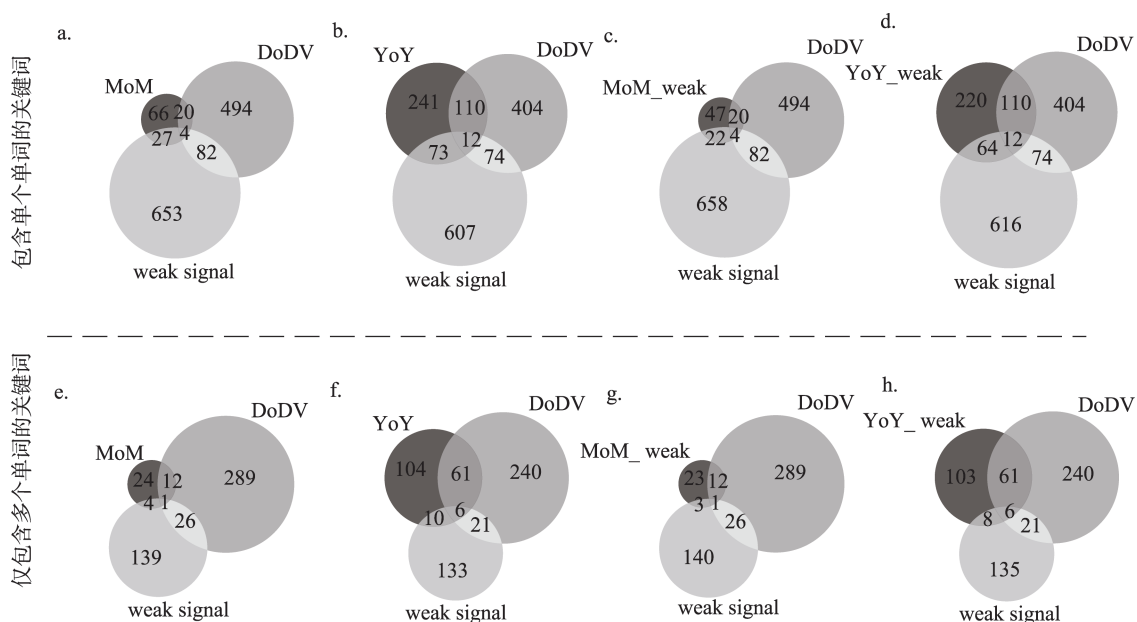


图3 基于ELO模型、信号强度过滤法(weak signal)与DoD/DoV组合图法的弱信号关键词对比分析

的变化可以看出, DoD/DoV组合图法能更多地识别出包含多个单词的关键词, 而弱信号强度过滤法倾向于识别单个单词的关键词。ELO方法则兼顾了这两类关键词的识别, 更符合实际需求。

从三种弱信号方法的弱信号关键词的维恩图(图3)来看, 以DoD/DoV组合图法的结果为基准, 可以发现YoY/YoY_weak与DoDV的关键词重合数(122)大于weak signal与DoDV的重合数(86), 而且这三个集合的重合关键词较少(12), 这表明ELO方法能识别出信号强度过滤法未能覆盖的关键词。

2.3.2 案例分析

本节使用案例分析方法进一步证明ELO方法的

有效性。从MoM_weak中选择三个具有不同文献数量级(历年文献数量分别为5+、20+、50+)的关键词“fracture tough”“resin transfer”和“preload”进行案例分析。

为了验证弱信号方法的识别结果, 本文定义两个验证指标: 关注热度和引用强度。其中, 关注热度指标用于衡量某个关键词在该领域学术文献中的出现频次, 反映了其在研究领域的受关注程度; 引用强度则用于计算含有该关键词的文献在一定时间段内的被引量(不同于在特定时间段内发表的论文在未来的总被引量), 衡量该关键词的学术影响力。图4展示了三个弱信号关键词的关注热度和引

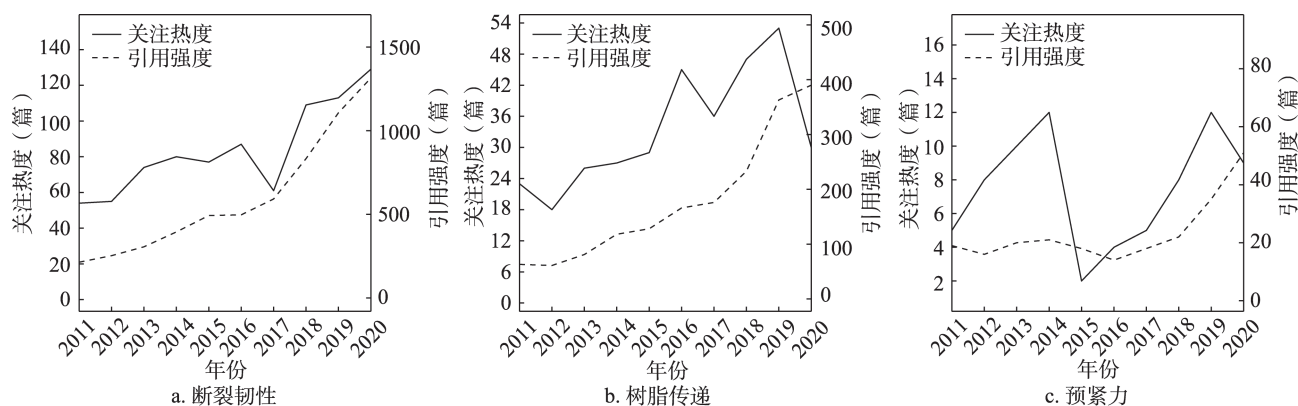


图4 代表性弱信号关键词的历年关注热度和引用强度演化图

用强度的演化趋势。可以看出,这三个关键词的关注热度虽然有波动,但均在2018年达到高峰,并显示出未来存在增长的趋势。从引用强度来看,这些关键词在2018年处于引用强度快速增长的早期阶段,显示出弱信号识别的前瞻性,符合预期要求。

(1) 断裂韧性(fracture tough)。通过文献检索与研读可知,碳纤维领域与断裂韧性相关的文献通常以碳纤维增强复合材料(carbon fibre-reinforced polymer, CFRP)的机械性质为主题,包括各种方法(增材制造^[39]、改变纤维角度取向和堆叠顺序^[40]等)、材料(氧化石墨烯、碳纳米管等^[41])对碳纤维增强复合材料的断裂韧性的影响。断裂是材料的损伤机理之一,断裂韧性能够衡量结构的能量吸收和损伤容忍能力,是断裂模式研究中的重要参数^[42]。

碳纤维增强复合材料作为有各向异性的复合材料,断裂模式相对复杂,早在2008年就有学者综合几次国际测试的结果制定了断裂韧性的测试标准,但其仍处于发展初步阶段,相关研究数量较少。由图4a可知,自2018年以来,断裂韧性在碳纤维领域的研究迅速发展至今,本文认为主要有两个方面的新发现推动了2018年以来断裂韧性相关的研究。其一,2016—2017年,碳纳米管优异的机械、电气和热性能的发现^[43-44],许多学者将碳纳米管引入复合材料层合板中^[45-46],有关碳纳米管与CFRP断裂韧性的研究不断展开^[47],包括控制碳纳米管在层压板中的分散^[48]、取向^[49]的技术,碳纳米管含量^[50]、长度^[51]等对断裂韧性的影响;其二,Papon等^[39]和Akasheh等^[52]于2018年首次提到增材制造(3D打印)碳纤维增强复合材料的断裂韧性,自此,许多学者开始探索3D打印碳纤维材料的断裂韧性。在

碳纳米管和增材制造两个方面研究的共同推动下,“断裂韧性”这一主题的关注热度与引用强度自2018年起开始快速提高。

(2) 树脂传递(resin transfer)。通过文献检索与研读可知,“树脂传递”一词通常出现在以“树脂传递模塑”为主题的文献中,树脂传递模塑(resin transfer moulding, RTM)方法起源于20世纪50年代的冷铸造工艺,RTM代表一种闭腔成型技术,可为制造的组件提供良好的尺寸精度和表面质量,其易加工性推动了RTM在航空航天工程中的广泛应用。

由图4b可知,树脂传递在碳纤维领域的研究在2016年出现过短暂峰值,但主要兴起在2018年以后。在2008年之前,相关文献对使用树脂传递模塑处理的碳纤维复合材料的研究较少。自2008年起,有学者开始研究树脂传递模塑处理或制备的碳纤维复合材料的力学特性^[53-54],并已有的一些在生产上的应用研究^[55]。2011—2017年,真空辅助树脂传递模塑制备的碳纤维复合材料及其性能被广泛研究^[56-57],但对其在工程上应用的研究较少;其中,2015—2016年,多壁纳米管在树脂传递模塑玻璃纤维混合复合材料中的应用被广泛研究^[58],相关研究数量在2015年达到高峰。2018年以来,一方面,真空辅助RTM性质的研究逐渐成熟,出现较多RTM制备的碳纤维复合材料在工程上的应用研究^[59];另一方面,热塑性树脂模塑传递工艺逐渐兴起^[59-61],各种新方法、新工艺改进RTM制备碳纤维复合材料的研究不断出现^[62-63]。随着这两类研究方向成为主流,这一关键词的关注热度与引用强度在2018年以来显著提高。

(3) 预紧力(preload)。预紧力是指在组件或结构中施加的初始力,在机械工程中,预紧力常见于螺栓连接中拧紧方法的优化上^[64],其主要目的是

确保螺栓在工作条件下不会松动,从而保证连接的可靠性和强度。

在机械工程领域,早在20世纪这一概念就被提出并用于研究螺栓拧紧方法的优化策略^[65],主要研究方向是如何找到不同状况下螺栓的最优预紧力。2009年,Heimbs等^[66]首次系统总结预紧力对CFRP材料的作用,碳纤维领域也开始这一方向的研究。直至近年,尤其在2018年增材制造在碳纤维领域兴起以后,随着碳纤维材料在工程上应用的推广,许多学者开始研究预紧力对螺栓、梁结构连接碳纤维复合材料时的增强作用^[67-68]或预紧CFRP对其他材料性能的增强作用^[69],预紧力的关注热度与引用强度不断提高。本文经过进一步文献调研发现,2022年以来,随着计算机科学的发展,机器学习、神经网络等计算机算法被用于计算最优预紧力^[70-71],该关键词的关注热度和引用强度进一步提升。

3 总结与展望

本文提出一种基于ELO模型的技术主题热度评估方法,该方法主要引入期望信号强度的概念,能够更深入地刻画技术主题热度。ELO方法能够综合研究主题的近期历史热度数据和同期所有研究主题的热度数据得到一个综合的期望热度值。不同于考虑时间因素的其他模型,ELO方法对于时间因素是自适应的,并且能够跨领域进行比较。相较于传统的热点主题识别,ELO方法不仅考虑给定研究主题本身的信号强度,而且考虑了与其他研究主题的相对信号强弱。

热度评估中的热点主题识别主要服务于情报决策者。情报工作者在使用传统方法时,需要先自定义筛选阈值再对结果进行解读,因此,必须考虑以下两个方面的问题。一方面,情报工作者需要向情报决策者解释阈值设定的合理性,无论是使用自定义的阈值还是使用统一的阈值(不同领域间指标的阈值是否能够统一还值得商榷)都需如此;另一方面,情报工作者需要说明为什么过去热度指标值与当前热度指标值发生或不发生变化所解释的结果是正确的,难以证实热度指标值的变化不是随机波动的结果。本文提出的ELO方法对上一个时间段的预期热度信号强度与实际热度信号强度进行比较,实时更新出每个主题最新的预期热度信号强度。该方法通过预期热度信号强度所代表的零模型生成期望热度信号强度,能够有效避免结果的随机性。

尽管ELO方法在动态性与可比性上显著优于传

统方法,其实践中仍需与语义挖掘等模块协同,以形成完整的情报分析闭环。针对实际应用需求,框架可通过三个方面扩展优化。其一,扩展主题内涵,利用语义模型(如BERT)合并同义/近义关键词,构建复合主题表征,提高热点主题识别的准确性。其二,丰富指标内涵,引入文献影响力(被引量、期刊影响因子)与颠覆性指标,设计多维度加权评估体系,从而更好地刻画主题的实际热度。其三,增强信号强度,通过引入跨领域信号增强弱信号识别。当主题在目标领域(如材料科学)初期文献稀疏时,可动态引入关联领域(如计算机领域)的高热度权重(如人工智能技术在其他学科的ELO值),加快本地ELO值的收敛速度,使其快速达到监测阈值,避免因文献积累滞后而错过早期识别窗口期^[72]。

不过,ELO方法也存在若干局限性。其一,虽然规避了时间阈值设定难题,却引入了权重的设定需求。权重的合理取值对主题ELO值能否及时反映热度至关重要,其确定过程仍具挑战性。其二,应用中,若不进行主题热度排名的恰当转换,ELO值极易受极端异常值干扰。本文采用的S形曲线转换法仅为应对之策的一种,在实际应用中,需依据具体场景灵活抉择适宜的转换方式。其三,在数据稀疏或新主题数量骤增的情形下,ELO值更新会陷入停滞或突变,严重削弱其对决策的有效支撑。其四,ELO值作为期望值,其功能主要局限于反映预期状况,在对主题热度进行深度解读与全面分析时存在固有缺陷,难以独自提供详尽且精准的决策信息。其五,基于文献频次的热度指标仍可能受出版滞后性的影响,例如,文献从投稿到公开发布的延迟可能导致早期弱信号的识别存在时间偏差。

综上所述,情报工作者向情报决策者提供建议时,既需要向其呈现技术主题过去与现在的热度信号强度等事实性情报,也需要引入自己的专业素养,对信号强度进行评估判断。报告内容除了描述事实性情报,还应该囊括信号强度的预期大小与过去相比,是下降还是上升。通过对比事实与预期的差异,情报决策者能从更多的信息中有效推测发展趋势,做好决策工作。下文列举三个方面问题以期引起未来更多理论和实证方面的探索。

(1) 如何运用微观、中观、宏观层面更多的科技情报和科技计量指标^[73],将ELO思想用于颠覆性技术早期识别、原创性评估等实际工作场景?这些场景或具回溯性,或具展望性,但都响应了科技情

报事业关切。

(2) 如何将ELO方法中的“期望”思维更好地融入领域知识类情报事实、情报机构事实和情报活动事实刻画, 以更好地为情报客户解决决策信息不完备的问题, 并提升科技情报工作品质^[74]?

(3) 如何进一步完善工程性问题, 构建相应的大屏展示和可视化平台, 并结合大模型技术, 以供情报工作人员更好地使用ELO方法辅助科技情报决策?

致谢 感谢北京大学王延飞教授在本文撰写过程中提出修改意见。

参 考 文 献

- [1] 熊坚, 林国栋, 蒋晶, 等. 红外技术在针灸领域应用现状和热点可视化分析[J]. 世界科学技术-中医药现代化, 2023, 25(7): 2479-2492.
- [2] 李薇, 张龙浩, 龚小清, 等. 基于文献计量的肺癌早期筛查技术前沿与热点分析[J]. 中国胸心血管外科临床杂志, 2022, 29(11): 1478-1484.
- [3] 徐晓日, 刘丹琳. 我国突发公共事件舆情治理研究的热点主题与演进趋势[J]. 行政与法, 2023(6): 15-27.
- [4] 孙艳姣. 智能制造专利热点技术竞争态势研究[J]. 制造技术与机床, 2023(10): 67-71.
- [5] 张新猛, 刘江鹏, 范亚茹, 等. 产业链视角下专利新兴技术主题识别[J]. 情报杂志, 2023, 42(8): 96-101, 55.
- [6] 刘盼盼, 王丽. 关系网络视角下新兴技术识别研究进展[J]. 图书情报工作, 2022, 66(11): 139-150.
- [7] 李昌, 杨中楷, 董坤. 基于多维属性动态变化特征的新兴技术识别研究[J]. 情报学报, 2022, 41(5): 463-474.
- [8] Jones B F. The burden of knowledge and the “death of the renaissance man”: is innovation getting harder?[J]. The Review of Economic Studies, 2009, 76(1): 283-317.
- [9] Tseng Y H, Lin Y I, Lee Y Y, et al. A comparison of methods for detecting hot topics[J]. Scientometrics, 2009, 81(1): 73-90.
- [10] 郝雯柯, 杨建林. 基于语义表示和动态主题模型的社科领域新兴主题预测研究[J]. 情报理论与实践, 2023, 46(2): 184-193.
- [11] 胡泽文, 李甜甜. 基于LDA主题模型和扎根理论的我国金融科技领域热点主题识别与进展分析[J]. 情报科学, 2023, 41(10): 99-111.
- [12] 李新根, 魏淑艳, 刘冬梅. 国内数字化治理研究的热点主题与演进趋势——基于CiteSpace的知识图谱分析[J]. 东南学术, 2022(2): 61-71, 247.
- [13] 丁敬达, 钟建兰. 新兴主题属性量化研究综述[J]. 图书情报工作, 2023, 67(9): 12-22.
- [14] 李红宇, 徐亚男, 张思洁. 我国职业教育研究热点识别与评析[J]. 河北师范大学学报(教育科学版), 2023, 25(6): 29-39.
- [15] 牛君, 程智超. 我国新能源产业发展研究现状与展望——基于文献计量学的可视化分析[J]. 生态经济, 2024, 40(1): 211-220.
- [16] 贾晓峰, 高芳, 耿喆. 基于文献计量的国际人体微生态研究主题分析[J]. 中国微生态学杂志, 2018, 30(12): 1402-1407.
- [17] 刘玉屏, 祝鹏, 胡雪珺. 国际中文教师研究热点主题、演进特征与未来展望[J]. 民族教育研究, 2024, 35(1): 157-166.
- [18] 刘成山, 杜怡然, 汪圳. 基于细粒度知识图谱的科技文献主题发现与热点分析[J]. 情报理论与实践, 2024, 47(5): 131-138.
- [19] 胡泽文, 韩雅蓉, 王梦雅. 基于LDA-Word2Vec的图书情报领域机器学习研究主题演化与热点主题识别[J]. 现代情报, 2024, 44(4): 154-167.
- [20] 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.
- [21] Abramo G, D' Angelo C A, Reale E. Peer review vs bibliometrics: which method better predicts the scholarly impact of publications[J]. Scientometrics, 2019, 121(1): 537-554.
- [22] 韩盟, 陈悦, 王玉奇, 等. 弱信号识别研究综述: 寻找微弱的未来信号[J]. 情报学报, 2023, 42(8): 996-1008.
- [23] Elo A E. The rating of chessplayers, past and present[M]. New York: Arco Publishing, 1978.
- [24] 聂秀萍. 基于时间序列的学科热点主题预测[D]. 北京: 中国农业科学院, 2019.
- [25] 邱均平, 温芳芳. 近五年来图书情报学研究热点与前沿的可视化分析——基于13种高影响力外文源刊的计量研究[J]. 中国图书馆学报, 2011, 37(2): 51-60.
- [26] 王山, 谭宗颖. 技术生命周期判断方法研究综述[J]. 现代情报, 2020, 40(11): 144-153.
- [27] 刘俊婉, 庞博, 徐硕. 基于弱信号的颠覆性技术早期识别研究[J]. 情报学报, 2023, 42(12): 1395-1411.
- [28] 裘惠麟, 邵波. 多源数据环境下科研热点识别方法研究[J]. 图书情报工作, 2020, 64(5): 78-88.
- [29] 荣国阳, 李长玲, 范晴晴, 等. 主题热度加速度指数——学科研究热点识别新方法[J]. 图书情报工作, 2021, 65(20): 59-67.
- [30] 方志超. 基于多元数据对象的科学计量学热点主题识别研究[D]. 大连: 大连理工大学, 2017.
- [31] 郑德俊, 程为. 基于三维主题特征测度的新兴主题识别研究[J]. 情报学报, 2024, 43(2): 167-180.
- [32] National Research Council. Intelligence analysis for tomorrow: advances from the behavioral and social sciences[M]. Washington, D.C.: The National Academies Press, 2011.
- [33] Priem J, Piwowar H, Orr R. OpenAlex: a fully-open index of scholarly works, authors, venues, institutions, and concepts[C]// Proceedings of the 26th International Conference on Science and Technology Indicators. Piscataway: IEEE, 2022: 1-5.
- [34] Grootendorst M. BERTopic: neural topic modeling with a class-based TF-IDF procedure[OL]. (2022-03-11). <https://arxiv.org/pdf/2203.05794>.

- [35] El Akrouchi M, Benbrahim H, Kassou I. End-to-end LDA-based automatic weak signal detection in web news[J]. *Knowledge-Based Systems*, 2021, 212: 106650.
- [36] 刘俊婉, 庞博, 徐硕. 基于弱信号的颠覆性技术早期识别研究[J]. *情报学报*, 2023, 42(12): 1395-1411.
- [37] Yoon J. Detecting weak signals for long-term business opportunities using text mining of Web news[J]. *Expert Systems with Applications*, 2012, 39(16): 12543-12550.
- [38] 韩盟, 陈悦, 王玉奇, 等. 基于异类数据和语义建构的新兴技术弱信号识别研究[J]. *情报学报*, 2024, 43(3): 302-312.
- [39] Papon E A, Haque A. Fracture toughness of additively manufactured carbon fiber reinforced composites[J]. *Additive Manufacturing*, 2019, 26: 41-52.
- [40] Naghipour P, Bartsch M, Chernova L, et al. Effect of fiber angle orientation and stacking sequence on mixed mode fracture toughness of carbon fiber reinforced plastics: numerical and experimental investigations[J]. *Materials Science and Engineering: A*, 2010, 527(3): 509-517.
- [41] Mishra K, Bastola K P, Singh R P, et al. Effect of graphene oxide on the interlaminar fracture toughness of carbon fiber/epoxy composites[J]. *Polymer Engineering & Science*, 2019, 59(6): 1199-1208.
- [42] Shrivastava R, Singh K K. Interlaminar fracture toughness characterization of laminated composites: a review[J]. *Polymer Reviews*, 2020, 60(3): 542-593.
- [43] Li N, Huang Y, Du F, et al. Electromagnetic interference (EMI) shielding of single-walled carbon nanotube epoxy composites[J]. *Nano Letters*, 2006, 6(6): 1141-1145.
- [44] Nguyen Dinh D, Nguyen P. The dynamic response and vibration of functionally graded carbon nanotube-reinforced composite (FG-CNTRC) truncated conical shells resting on elastic foundations[J]. *Materials*, 2017, 10(10): 1194.
- [45] Ou Y F, González C, Vilatela J J. Understanding interlaminar toughening of unidirectional CFRP laminates with carbon nanotube veils[J]. *Composites Part B: Engineering*, 2020, 201: 108372.
- [46] Su Y N, Zhou F, Wei X H, et al. Enhanced mechanical and electrical properties of carbon fiber/poly (ether ether ketone) laminates via inserting carbon nanotubes interleaves[J]. *Journal of Applied Polymer Science*, 2020, 137(19): 48658.
- [47] Zheng N, Huang Y D, Liu H Y, et al. Improvement of interlaminar fracture toughness in carbon fiber/epoxy composites with carbon nanotubes/polysulfone interleaves[J]. *Composites Science and Technology*, 2017, 140: 8-15.
- [48] Reia da Costa E F, Skordos A A, Partridge I K, et al. RTM processing and electrical performance of carbon nanotube modified epoxy/fibre composites[J]. *Composites Part A: Applied Science and Manufacturing*, 2012, 43(4): 593-602.
- [49] Li Q X, Church J S, Naebe M, et al. A systematic investigation in to a novel method for preparing carbon fibre-carbon nanotube hybrid structures[J]. *Composites Part A: Applied Science and Manufacturing*, 2016, 90: 174-185.
- [50] Zainol Abidin M S, Herceg T, Greenhalgh E S, et al. Enhanced fracture toughness of hierarchical carbon nanotube reinforced carbon fibre epoxy composites with engineered matrix microstructure[J]. *Composites Science and Technology*, 2019, 170: 85-92.
- [51] Du X S, Liu H Y, Xu F, et al. Flame synthesis of carbon nanotubes onto carbon fiber woven fabric and improvement of interlaminar toughness of composite laminates[J]. *Composites Science and Technology*, 2014, 101: 159-166.
- [52] Akasheh F, Aglan H. Fracture toughness enhancement of carbon fiber-reinforced polymer composites utilizing additive manufacturing fabrication[J]. *Journal of Elastomers & Plastics*, 2019, 51(7/8): 698-711.
- [53] Hsiao K T, Gangireddy S. Investigation on the spring-in phenomenon of carbon nanofiber-glass fiber/polyester composites manufactured with vacuum assisted resin transfer molding[J]. *Composites Part A: Applied Science and Manufacturing*, 2008, 39(5): 834-842.
- [54] Uddin N, Shohel M, Vaidya U K, et al. Bond strength of carbon fiber sheet on concrete substrate processed by vacuum assisted resin transfer molding[J]. *Advanced Composite Materials*, 2008, 17(3): 277-299.
- [55] Thouin M, Ghiasi H, Lessard L. Design of a carbon fiber bicycle stem using a novel internal bladder resin transfer molding technique[J]. *Advanced Composites Letters*, 2010, 19(1): 51-60.
- [56] Aurrekoetxea J, Agirregomezhorta A, Aretxaga G, et al. Impact behavior of carbon fiber/epoxy composite manufactured by vacuum-assisted compression resin transfer molding[J]. *Journal of Composite Materials*, 2012, 46(1): 43-49.
- [57] Lopez-Arraiza A, Amenabar I, Agirregomezhorta A, et al. Experimental analysis of drilling damage in carbon-fiber reinforced thermoplastic laminates manufactured by resin transfer molding[J]. *Journal of Composite Materials*, 2012, 46(6): 717-725.
- [58] Zeiler R, Kuttner C, Khalid U, et al. The role of multi-walled carbon nanotubes in epoxy nanocomposites and resin transfer molded glass fiber hybrid composites: dispersion, local distribution, thermal, and fracture/mechanical properties[J]. *Polymer Composites*, 2017, 38(9): 1849-1863.
- [59] Ball C A, Greydanus S, Swentek L, et al. Development of an epoxy carbon fiber reinforced roof frame using the high pressure resin transfer molding (HP-RTM) process[C]// *Proceedings of the WCX SAE World Congress Experience*. Warrendale: SAE International, 2020: 6415-6422.
- [60] Choi C W, Jin J W, Lee H, et al. Optimal polymerization conditions in thermoplastic-resin transfer molding process for mechanical properties of carbon fiber-reinforced PA6 composites using

- the response surface method[J]. *Fibers and Polymers*, 2019, 20(5): 1021-1028.
- [61] Lee J, Lim J W, Kim M. Effect of thermoplastic resin transfer molding process and flame surface treatment on mechanical properties of carbon fiber reinforced polyamide 6 composite[J]. *Polymer Composites*, 2020, 41(4): 1190-1202.
- [62] Lee J H, Son S M, Yoo J J, et al. Thermoplastic resin transfer molding of carbon fiber reinforced polyamide 6 composite with the improved processability using zeolite particle[J]. *Korea-Australia Rheology Journal*, 2023, 35(1): 39-45.
- [63] Xu X X, Wei K, Mei M, et al. An ultrasound-assisted resin transfer molding to improve the impregnation and dual-scale flow for carbon fiber reinforced resin composites[J]. *Composites Science and Technology*, 2024, 255: 110710.
- [64] Crococolo D, De Agostinis M, Fini S, et al. Optimization of bolted joints: a literature review[J]. *Metals*, 2023, 13(10): 1708.
- [65] Bickford J H, Nassar S. *Handbook of bolts and bolted joints*[M]. New York: Marcel Dekker, 1998.
- [66] Heimbs S, Heller S, Middendorf P, et al. Low velocity impact on CFRP plates with compressive preload: test and modelling[J]. *International Journal of Impact Engineering*, 2009, 36(10/11): 1182-1193.
- [67] Tian Z G, Zhi Q, Feng X Y, et al. Effect of preload on the weld quality of ultrasonic welded carbon-fiber-reinforced nylon 6 composite[J]. *Polymers*, 2022, 14(13): 2650.
- [68] Tong X, Meng L, Wan Y, et al. Preload relaxation behavior and its influence on the mechanical performance of bolted carbon fiber-reinforced thermoplastic sheet molding compound joints[J]. *Polymer Composites*, 2024, 45(10): 8810-8824.
- [69] Helal K, Yehia S, Hawileh R, et al. Performance of preloaded CFRP-strengthened fiber reinforced concrete beams[J]. *Composite Structures*, 2020, 244: 112262.
- [70] Zhang H B, Wang M W, Deng W, et al. Semi-physical simulation optimization method for bolt tightening process based on reinforcement learning[J]. *Machines*, 2022, 10(8): 637.
- [71] Foissac C, Daidié A, Segonds S, et al. Application of neural networks for smart tightening of aeronautical bolted assemblies[C]// *Proceedings of the International Joint Conference on Mechanics, Design Engineering & Advanced Manufacturing*. Cham: Springer, 2023: 421-429.
- [72] Trotta A, Ziosi M, Lomonaco V. The future of ethics in AI: challenges and opportunities[J]. *AI & Society*, 2023, 38(2): 439-441.
- [73] 步一, 许家伟, 黄文彬. 基于引文的科学文献定量评价: 引文影响力指标评述[J]. *图书情报知识*, 2021, 38(6): 47-59, 46.
- [74] 杜元清, 王延飞. 论情报事实[J]. *情报理论与实践*, 2024, 47(10): 1-9.

(责任编辑 魏瑞斌)