



开放科学
(资源服务)
标识码
(OSID)

基于深度语义理解的“技术—知识”关联实体识别及演化分析研究

杨金庆^{1,2} 李嘉琦¹ 杨儒汉¹ 罗星雨¹ 程秀峰¹

- 华中师范大学信息管理学院 武汉 430079;
- 富媒体数字出版内容组织与知识服务重点实验室 北京 100038

摘要: [目的/意义] 从细粒度视角理解“技术—知识”关联实体, 构建专利文献中技术要素与知识要素识别的实现方案。[方法/过程] 选取传统机器学习模型 *HMM*、*CRF*, 深度学习模型 *BiLSTM-CRF*、*BERT-Softmax*、*BERT-CRF* 和 *BERT-BiLSTM-CRF* 进行任务训练学习, 以便确定性能最优的细粒度技术与知识实体识别模型。[结果/结论] 为了验证所构建技术与知识实体识别的理论框架, 以出版印刷领域文本作为实验验证场景, 从专利文本中随机抽取 7853 条有效语料句, 标注了 71626 个实体, 通过训练学习确定 *BERT-BiLSTM-CRF* 为性能较好的实体识别模型, 其对知识与技术实体识别综合性能 F1 值为 0.82。此外, 运用训练出的最优模型从 66665 篇专利文本的第一权利要求、权利要求、独立权利要求和技术功效中识别出 4769296 对知识与技术实体关联组合体, 并分析了技术演化路径和“技术—知识”关联网络结构的演化规律。

关键词: 科技情报; 学科技术关联; 实体识别; BERT

中图分类号: G35; TP391

Research on Related Entity Recognition and Evolution Analysis of “Technology-Knowledge” Based on Deep Semantic Understanding

YANG Jinqing^{1,2} LI Jiaqi¹ YANG Ruhan¹ LUO Xingyu¹ CHENG Xiufeng¹

- School of Information Management, Central China Normal University, Wuhan 430079, China;
- Key Laboratory of Rich-media Knowledge Organization and Service of Digital Publishing Content, Beijing 100038, China

基金项目 富媒体数字出版内容组织与知识服务重点实验室开放基金项目“基于深度语义理解的技术谱系自动构建研究”(ZD2023/11-03); 中央高校基本业务费项目“基于技术要素深度语义理解的创新路径识别及引导策略研究”(CCNU24ZZ140); 中央高校基本业务项目“基于知识网络科学知识角色转变研究”(CCNU23XJ013)。

作者简介 杨金庆(1991-), 博士, 副教授, 主要研究方向为科技情报、学科知识演化, E-mail: jinq_yang@163.com; 李嘉琦(2001-), 硕士研究生, 主要研究方向为信息检索、用户行为; 杨儒汉(2000-), 硕士研究生, 主要研究方向为政务智能问答; 罗星雨(2002-), 硕士研究生, 主要研究方向为科技情报、学科知识演化; 程秀峰(1981-), 博士, 教授, 主要研究方向为信息组织与检索、大数据分析及应用。

引用格式 杨金庆, 李嘉琦, 杨儒汉, 等. 基于深度语义理解的“技术—知识”关联实体识别及演化分析研究[J]. 情报工程, 2024, 10(5): 73-84.

Abstract: [Objective/Significance] Understand the “technology-knowledge” related entities from a fine-grained perspective, and construct a implementation scheme for identifying technology elements and knowledge elements in patent documents. [Methods/Processes] The traditional machine learning models *HMM* and *CRF*, and the deep learning models *BiLSTM-CRF*, *BERT-Softmax*, *BERT-CRF*, and *BERT-BiLSTM-CRF* are selected for the task training and learning in order to identify the fine-grained technology with optimal performance and the knowledge entity recognition model. [Results/Conclusions] In order to validate the theoretical framework of technology and knowledge entity recognition, this paper takes the text in the field of publishing and printing as the experimental validation scenario, randomly selects 7853 valid corpus sentences from patent texts, and annotates 71626 entities, and determines that the *BERT-BiLSTM-CRF* is the entity recognition model with better performance through training, and the F1 value of its comprehensive performance for knowledge and technology entity recognition is 0.82. In addition, this paper applies the trained optimal model to identify 4769296 pairs of knowledge-technology entity association combinations from the first claim, claims, independent claims and technical effects of 66665 patents, and analyzes the technology evolution paths and the evolution pattern of the “technology-knowledge” association network structure. We also analyzed the technology evolution path and the evolution law of “technology-knowledge” association network structure.

Keywords: Scientific and Technological Information; Subject Technology Correlation; Entity Recognition; *BERT*

引言

新一轮科技革命和产业变革加速演进，科学与技术的交互日益紧密，共同驱动科技创新发展。尽管科学与技术拥有独立的知识结构和演化过程，但是在科技创新中有多层次交融。技术的本质是对科学发现的利用，新技术源自新的科学发现实践应用，或是现有技术的改进、调整和组合^[1]。同样，新科学的发现也是依赖于科学与技术之间的合作，例如青霉素、晶体管等产品的研发，表现出科学与技术之间不仅遵循自身的发展规律，而且还相互影响，既促进科学进步，又催生新技术^[2]。目前，科学与技术之间的关系一般包括技术与科学的依赖关系、结合关系、互动关系以及知识传递关系^[3]。对技术而言，科学为技术提供强有力的理论基础，是技术开发的基本依据。从构成要素来看，技术由经验型要素、知识型要素和实体应用型

要素构成^[4]，反映出新技术的形成离不开科学知识 with 现有技术的双重支撑。

目前，“技术—知识”关联主要由引用关系连接的期刊论文和专利文献表示^[5]。可是，技术本质上是人类为满足自身的需要，在实践活动中根据实践经验或科学原理所创造或发明的各种手段和方式方法的总和。技术本身由多种要素构成，可理解为主体知识、经验、技能与客体（工具、机器设备等）要素的统一^[6]。专利文献作为技术核心内容呈现的载体形式，不仅包含了表述工具、机器设备等的技术要素，还包含了表示主体知识及科学原理的知识要素。

“知识要素”是指专利文本中表述自然科学原理以及理论知识等的知识实体，“技术要素”是指表示技术实践中的方法、技巧、手段和工具等的技术实体^[7]。鉴于此，本文试图借助深度语义理解技术构建专利文献中技术与知识实体识别的实现方案，识别出专利文本中共联关

联的知识与技术实体对，并对细粒度“技术—知识”实体关联组合进行初步的演化分析。

1 相关研究

1.1 基于深度语义理解的科技文献实体识别研究

科技文献作为知识的重要载体，包括论文和专利两大类^[8]，是科学技术演化进程和发展水平的直观反映。如何高效地对科技文献内容中的知识与技术实体进行识别与抽取，是生物医学领域、社交媒体领域、图书情报等领域的重要研究主题^[9]。实体识别（*Named Entity Recognition, NER*）是自然语言处理（*NLP*）中的一个基本任务，目标是从文本中识别出具有特定意义的实体（如人名、地点、组织名），并将其分类为预定的类别^[10]。科技文献是蕴含大量知识的非结构化文本，其中涉及多种类型实体、长实体链，呈现出更新快、专业术语密集、语义识别困难、重叠关系复杂^[11]等特点，导致科技文献实体识别相对于通用领域实体识别难度大幅度提升。此外，不同领域科技文献在语言风格、结构和术语等方面的差异也要求开发适用于各自特点的实体识别方法和模型。

在早期科技文献实体识别研究中，基于统计的机器学习方法，如支持向量机（*SVM*）、隐马尔可夫模型（*HMM*）、条件随机场（*CRF*）等被广泛使用^[12]。随着科技文献术语的复杂化和结构的多样化发展，基于统计的机器学习方法在语义理解方面逐渐显得不足，而深度学习的发展为实体识别带来了突破性进展。深度学

习的方法在大规模数据集训练基础上学习丰富的语言规律和语义关系，能够捕捉和建模复杂的、非直观的语言现象和语义模式，进而提升实体识别性能。典型的深度学习方法包括递归神经网络（*RNN*）、卷积神经网络（*CNN*）、双向长短期记忆网络（*BiLSTM*）以及 *BERT* 和其他基于 *Transformer* 的模型^[13]。目前较为流行的实体识别方法主要为基于 *BiLSTM* 和 *CRF* 框架的变体或以 *BERT* 作为预训练语言模型进行优化的方法。如，Gao 等^[14] 构建了一种适用于稀少标记样本数据集的技术实体识别的对抗性多任务神经网络，利用 *BERT* 捕捉文本深层次语义特征，再通过引入对抗性训练机制增强模型对实体边界和关系的识别能力。Ma 等^[15] 提出了 *BiLSTM* + *CRF* 模型与基于特征的命名实体知识库融合的实体抽取方法，提取文献中时间、地点及技术实体，以分析最具潜力的中国脆弱生态修复技术。于润羽等^[16] 基于关键词—字符 *LSTM* 和注意力机制提出了一种能够考虑文本中的上下文信息以及全局信息的命名实体识别算法，应用于科技学术会议论文数据的识别，实验结果显示该方法优于对比算法。Chen 等^[17] 提出一种基于外部医学知识和自注意力机制的 *MFA-BERT-BiLSTM-CRF* 模型，能够有效地捕捉丰富的医学语义特征和全局上下文信息，优化了复杂医学文献命名实体识别的性能。王宇晖等^[18] 提出基于 *Transformer* 与技术词信息的知识产权实体识别方法，包含技术词信息融合层、知识产权语义抽取层和实体标签优化层三个层次，有效解决了知识产权实体的多义性问题，以及 *Transformer* 模型的相对位置感知弱等问题。Zhao 等^[19] 提出了针对中国专利文本

的联合实体和关系提取的 *BERT-BiLSTM-DGAT* 模型,能够准确捕获中国专利文本中的复杂长实体和重叠关系,提高了实体和关系提取的准确性。

随着技术的不断发展,针对科技文献实体识别任务的研究越来越多,识别效果也不断提升,但仍面临一些挑战,如多类型实体的识别、领域内实体识别精度的差异,以及如何处理专业术语的多义性问题等。为解决这些问题,未来科技文献实体识别研究应向深层次语义、跨领域适应性、自动学习和调整的实体识别方向发展。

1.2 “科学—技术”关联视角下科学演化规律研究

科学是由人类对认识客体的知识体系、产生知识的活动、科学方法等按一定层次、一定方式所构成的一种动态知识系统,它不仅仅是一套固定的知识,更是一种持续发展和进步的过程^[20-21]。技术既是一种知识体系,也是实物的集合,包括技能、方法、物理设备和软件等^[22]。科学与技术关联本质是科学知识体系和技术体系内部知识要素之间的复杂关系,表现为技术专利与科学论文通过相互引用、知识耦合、学者—发明人等途径产生的关联^[23]。在科学与技术的关联关系上,学术界持有不同见解,Bundy^[24]认为科学与技术虽有各自独特的知识积累结构,但知识流动的路径可以是双向的。Bush^[25]认为基础研究是技术进步的先驱,科学为基础研究,而技术则将科学原理应用于实践。Guan等^[26]提出了双螺旋理论,认为科学与技术的信息交互反馈中呈现螺旋上升的发展趋势。总体而言,科学与技术是独特而相互

作用的共同体,二者共同塑造了科技发展的脉络和走向。因此,测量科学与技术之间的互动转化,识别二者间潜在的推动或启示作用,是厘清科学演化规律和模式的关键,对预测科技趋势、指导科研政策和推动技术创新具有重要意义。

当前,科学技术关联研究中,学者们通常以文献和专利分别表征科学和技术,利用科学计量学工具和网络分析技术结合历史分析的方法进行探讨,主要包括专利的论文引文分析法,论文的专利引文分析法和专利发明人—论文作者关联关系分析法^[8]。如,Verbee等^[27]通过分析生物技术和信息技术领域论文与专利间的引用网络、合作模式和知识流动,探讨了领域科学知识的发展脉络以及科学技术之间的交互。韩芳^[28]基于不同技术领域专利与文献的引用关系,揭示了科学理论与技术成果之间的相互转化过程和规律。研究发现,科技专利的增长与科学文献的增长密切相关,反映出共演化特征。Ke^[29]针对生物技术和药物专利领域,对不同时间段、不同国家及机构类型的专利引用进行了详细探讨,揭示了科学与技术之间的关联和演变,为科技政策制定和国际合作提供了依据。刘雨梦等^[30]构建了基础科学(*Science*)、技术科学(*Engineering science*)及工程应用(*Engineering*)的SESE研究模型,应用在辽宁省重大装备协同创新中心和智能型新能源汽车协同创新中心中,以分析科学技术的互动行为和共演过程。吴宁等^[31]利用多层网络方法,基于专利和论文数据,构建了“科学—技术”知识流动微观网络分析框架,揭示了科学与技术之间的知识流动网络结构特征,为理解科技联系与

挖掘创新点提供了新视角。

随着科学技术关联研究的深入,少数学者开始从微观层面进行研究,但整体而言,现有科技关联视角下的科学演化规律和模式研究多以引文为纽带,通过专利与论文间的引证映射,揭示科学与技术的知识关联和演化规律。这类研究侧重于文献层次的引用关系和主体层次的相似关系的计量分析,未能充分从语义角度揭示科学技术内在关联,难以透视文献中知识单元间的直接联系,限制了对科学演化规律和模式的全面剖析与理解。未来的研究应更注重从深度语义角度出发,以文本内容为基础,知识与技术实体为核心,探索科学技术之间错综复

杂的联系,更有效地推动科技创新和应用。

2 数据来源及标注过程

为了更好地构建细粒度“技术—知识”关联实体识别模型,笔者选取了技术与知识交叉较大的出版印刷领域作为实验场景,并从商业专利库 *Patsnap* 中运用分类号搜索的方式获取 9880 份专利文本,随机抽取 7853 条有效语料句。本文采用了 *Label studio* 数据标注工具对有效语句标注技术实体(机器、工具等)和知识实体(理论、知识等),共标注了 71626 个实体,两种实体标注的示例如图 1 所示。



图 1 技术和知识实体在专利文本中的示例

3 实体识别结果分析

3.1 “技术—知识”关联实体识别模型构建

细粒度技术实体和知识实体的识别需要在对专利文本语义进行深刻理解的基础上,对输入模型的语句中的每个词组进行处理,预测两类实体标签。本文拟通过参照当下命名实体识别任务中的常见语义理解框架,构建“技术—知识”关联实体识别模型。目前,命名实体识别的深度语义理解框架前后经历两个阶段:一是以卷积神经网络(CNN)和循环神经网络(RNN)为文本特征编码器,学习单词序列的隐层表示,再借助CRF完成实体标签分类,此

阶段以 *LSTM-CRF* 实体识别模型为代表;二是借助以 *BERT* 预训练模型为底层的文本特征编码器,利用预训练模型强大的文本语义理解能力预测实体标签。

结合以往实体识别研究结论,本文以 *BERT* 为底层文本特征编码器,构建 *BERT-CRF* 或 *BERT-BiLSTM-CRF* “技术—知识”关联的两种类型实体的识别模型。模型训练数据构建环节,本文采用“*BIO*”数据标注模式,将技术实体(机器、工具等)和知识实体(理论、知识等)分别标注为 *B-tool/app\I-tool/app*, *B-knowledge\I-knowledge*, 适应以 *BERT* 为底层的技术要素深度学习识别架构,如图 2 所示。

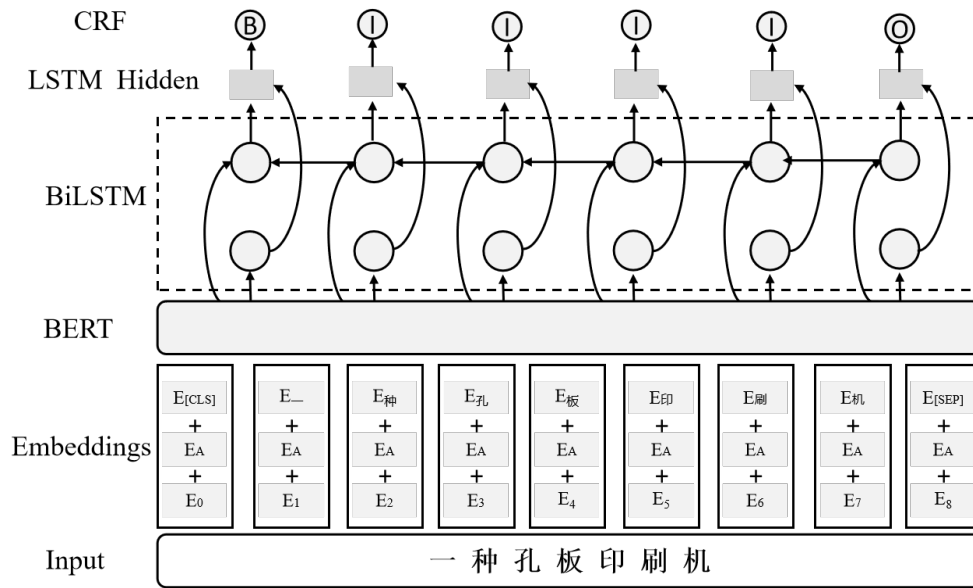


图2 “技术—知识”关联实体识别深度学习框架

由图2可知，输入有效的专利文本序列后由嵌入层 (*Embedding layer*) 将序列嵌入转化为向量表示，随后借助 *BERT* 或 *BERT-BiLSTM* 文本特征编码能力对输入的文本向量表示进行建模，以获取文本深层语义信息，最后通过 *CRF* 分类算法对技术与知识两类实体标签进行分类预测。

3.2 基于深度语义理解的模型训练与评价

3.2.1 实验设置及评价指标

本文标注的 71626 个实体中包含 55444 个技术实体和 16182 个知识实体，按 1:1:3 划分为测试集、验证集和训练集，两者在测试集、验证集和训练集中存在数据不均衡现象，如表 1 所示。

表1 两类实体在测试集、验证集和训练集中的分布情况

数据划分	技术实体	知识实体
训练集	25272	8799
验证集	15486	3507
测试集	14686	3876

在模型训练过程中，本文以出版印刷领域的数据集为基础，采用 *Adam* 优化器微调 *BERT* 预训练模型，对该领域专利文本中的语义进行深层次理解，并设定 *BERT* 和 *CRF* 的超参数，如表 2 所示。

表2 *BERT/CRF* 超参数设定结果

超参数	值
优化器	Adam
<i>BERT</i> 学习率	5e-5
<i>CRF</i> 学习率	5e-3
隐层维度	768
Epoch	10
Batch_size_train	32
Batch_size_eval	32
Max_len	150

为了对模型识别性能进行评价，本文采取实体级的评价方式，精确度 (*precision*) 是指正确识别的技术与知识实体数与模型识别出的技

术与知识实体总数的比例，召回率 (*recall*) 是指正确识别的技术与知识实体数与数据集中实际存在的技术与知识实体总数的比例，F1 值是精确度和召回率的调和平均值，可衡量模型的综合性能，计算公式如下：

$$precision = \frac{\text{正确识别的技术与知识实体数量}}{\text{模型识别的技术与知识实体数量}} \quad (1)$$

$$recall = \frac{\text{正确识别的技术与知识实体数量}}{\text{数据集中实际的技术与知识实体数量}} \quad (2)$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (3)$$

3.2.2 模型训练与对比评价

为了确定性能最好的“技术—知识”关联实体识别模型，本文先后对 *HMM*、*CRF*、*BiLSTM-CRF*、*BERT-Softmax*、*BERT-CRF* 和 *BERT-BiLSTM-CRF* 等深度学习框架进行训练和学习，并对模型识别性能进行评价，如表 3 所示。

表 3 技术与知识实体识别模型的性能对比评价

模型	实体类型	Precision	Recall	F1
HMM	不区分实体	0.66	0.58	0.62
	知识实体	0.48	0.54	0.51
	技术实体	0.70	0.59	0.64
CRF	不区分实体	0.68	0.85	0.76
	知识实体	0.42	0.80	0.56
	技术实体	0.74	0.86	0.79
BiLSTM-CRF	不区分实体	0.79	0.80	0.79
	知识实体	0.56	0.67	0.61
	技术实体	0.84	0.82	0.83
BERT-Softmax	不区分实体	0.84	0.76	0.80
	知识实体	0.81	0.48	0.60
	技术实体	0.85	0.84	0.84
BERT-CRF	不区分实体	0.84	0.80	0.82
	知识实体	0.78	0.57	0.66
	技术实体	0.85	0.87	0.86
BERT-BiLSTM-CRF	不区分实体	0.81	0.82	0.82
	知识实体	0.77	0.60	0.67
	技术实体	0.82	0.88	0.85

对比各模型的 F1 值，不难发现 *BERT-CRF* 和 *BERT-BiLSTM-CRF* 两者对知识与技术两类实体综合识别性能相当，F1 值均为 0.82。也可看出，数据不均衡对知识实体的识别性能产生一定的影响，为了使得两类实体的识别性能都较为可用，*BERT-BiLSTM-CRF* 模型可作为“技

术—知识”关联实体识别的实现路径。

4 “技术—知识”实体关联组合的演化规律分析

专利文本作为科技创新的重要载体，可视为由知识单元实体和技术单元实体按照一定的

逻辑和功能组合而成。知识与技术实体在一定程度上反映着“技术—知识”组合的演化规律。本文按照“链接—链条—网络”的分析思路先后对“技术—知识”组合、“技术—知识”演化路径和“技术—知识”关联网络结构演化进行规律揭示。具体而言,运用知识与技术实

体识别性能最优的 *BERT-BiLSTM-CRF* 模型从 1985—1998 年和 2010—2022 年跨度内 66665 篇专利文本的第一权利要求、权利要求、独立权利要求和技术功效中识别出知识与技术实体组合共 4769296 对,其各年份关联组合数量分布如图 3 所示。

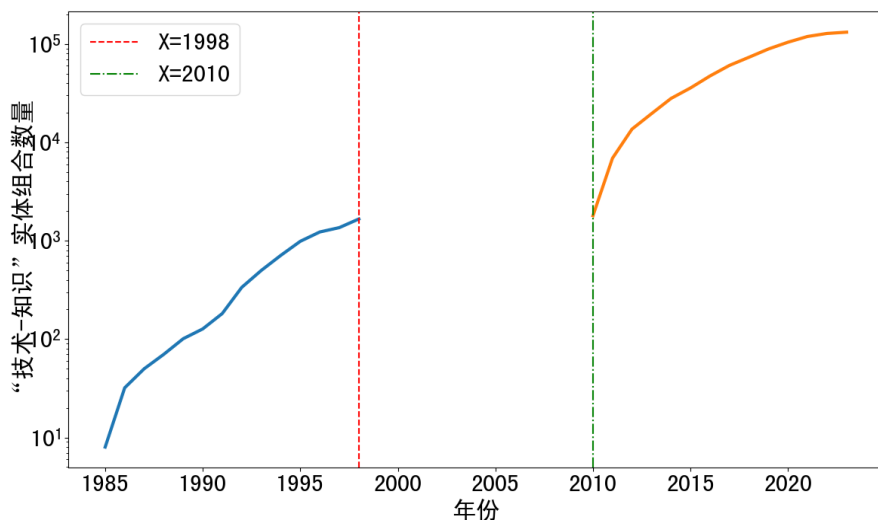


图3 “技术—知识”实体关联组合数量历时分布情况

观察图3中“技术—知识”实体关联组合对数在1985—1998年和2010—2022年阶段均呈现上升趋势,但2010—2022年呈现相对较快的增长率,表明新的“技术—知识”组合创新不断涌现,在出版印刷领域“技术—知识”关联趋势逐渐增强。本文选取1985—1998年和2010—2022年时间跨度内的“技术—知识”实体关联关系,以时间变量为导引表达不同实体之间的继承关系,进而可视化生成细粒度“技术—知识”实体之间的演化路径,如图4所示。

此外,“技术—知识”实体关联对链接形成“技术—知识”关联网络,关联网络的结构变化一定程度也反映了出版印刷领域的技术演化规律。本文为控制网络规模,选取了共现频

次大于2的知识与技术实体关联对,进而构建技术—知识关联网络,并计算了典型网络结构指标的历时变化情况,如图5所示。

观察图5可知,1985—1998年时间跨度内关联网络的平均中心度增长较为缓慢,而2010—2022年增长较为迅速。结合图3可知,尽管近十多年“技术—知识”关联对不断增加,但“技术”与“知识”节点间的相互关联趋势也在快速上升,表明出版印刷领域技术创新更倾向于“技术”与“知识”相互驱动的模式。中介中心度在初始成长阶段下降较为明显,反弹后仍呈现下降趋势,关联网络中节点平均影响力也在持续下降,表明整个关联网络的连通性变大,也侧面表明了“技术”与“知识”关联趋势增强。

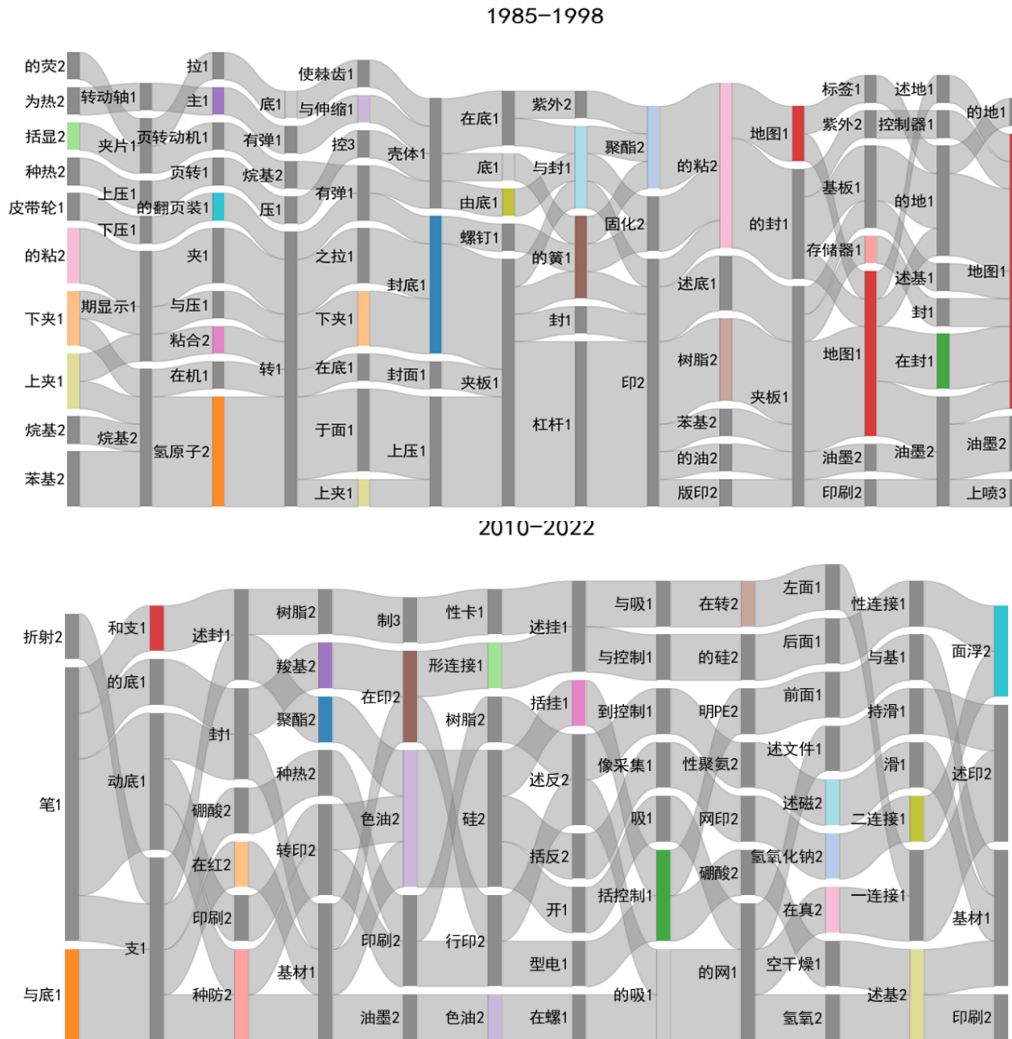


图 4 技术演化路径可视化示例图 (1985—1988 年, 2010—2022 年)

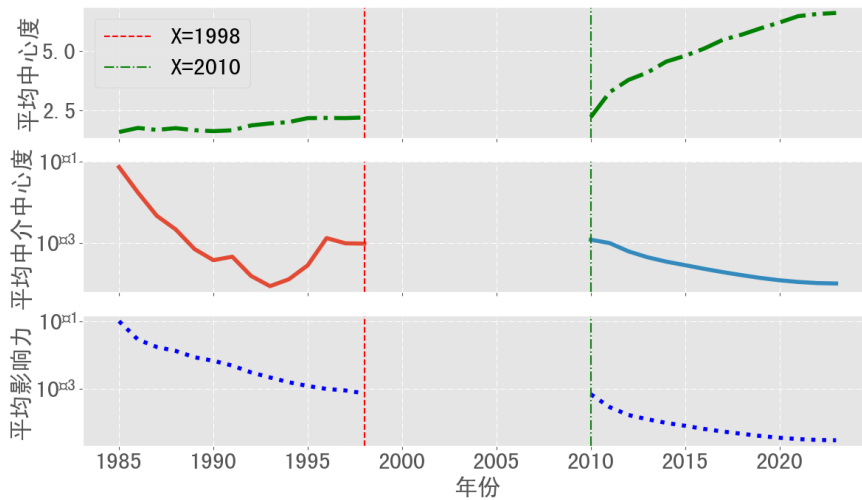


图 5 典型网络结构指标值历时可视化分析

5 结语

“技术—知识”关联实体识别对推动科技创新、制定科技政策,以及促进知识的转化具有极其重要的意义。以往科学技术关联研究主要以专利文献引用关系为连接,忽视了文本内容中所蕴含的深层语义信息。针对以上问题,本文从细粒度角度,识别专利文献中的技术与知识要素,通过技术与知识之间原有的内在逻辑关系,深入语义层面挖掘科技之间的关联与互动。为了更为准确识别出技术与知识实体,本文选取了传统机器学习模型 *HMM*、*CRF*,深度学习模型 *BiLSTM-CRF*、*BERT-Softmax*、*BERT-CRF* 和 *BERT-BiLSTM-CRF* 进行任务训练学习,获取性能最优的细粒度技术与知识实体模型框架。实验结果表明以 *BERT* 为底层文本特征编码器的预训练模型识别性能较好,尤其 *BERT-BiLSTM-CRF* 和 *BERT-CRF* 组合框架,对两类实体识别的综合性能 *F1* 值为 0.82。为进一步探析“科学—技术”关联视角下科学演化规律及模式,本文运用训练出的最优模型 *BERT-BiLSTM-CRF*,从 1985—1998 年和 2010—2022 年两个时间跨度内的专利文本中识别“技术—知识”实体关联组合体,并将各年份关联组合数量分布可视化,分析可知两阶段内实体关联组合体数量呈现上升趋势,2010—2022 年增长率相对较快。在此基础上,本文以时间变量为导引表达不同实体之间的继承关系,构建了细粒度“技术—知识”关联网络结构,计算并可视化显示典型网络结构指标的历时变化情况,结果表明 1985—1998 年期间关联网络

的平均中心度增长变化率低于 2010—2022 年期间,中介中心度在初始成长阶段下降较为明显,反弹后仍呈现下降趋势。关联网络中节点平均影响力也在持续下降。实验分析发现尽管近十多年“技术—知识”关联对不断增加,但“技术”与“知识”节点间的相互关联趋势也在快速上升,表明出版印刷领域技术创新更倾向于“技术”与“知识”相互驱动的模式。本文试图从专利本身发掘“技术—知识”实体间的关联,初步分析了“技术—知识”实体关联组合的演化规律,仍存在不足之处,主要包括实验训练数据集存在客观的数据不均衡问题以及技术与知识实体标注过程中需要较强的专业背景知识,未来在继续提升细粒度“技术”与“知识”实体识别准确率的基础上,深入分析技术和知识关联背后的机理。

参考文献

- [1] 林苞,雷家骕.基于科学的创新与基于技术的创新——兼论科学—技术关系的“部门”模式[J].科学学研究,2014,32(9):1289-1296.
- [2] 林苞,雷家骕.基于科学的创新模式与动态——对青霉素和晶体管案例的重新分析[J].科学学研究,2013,31(10):1459-1464.
- [3] 刘小玲,谭宗颖,张超星.国内外“科学—技术关系”研究方法述评——聚焦文献计量方法[J].图书情报工作,2015,59(13):142-148.
- [4] 陈其荣.自然辩证法导论[M].上海:复旦大学出版社,1995:188-208.
- [5] LI X, WANG T, PANG Y, et al. Review of Research on Named Entity Recognition[C]// SUN X, ZHANG X, XIA Z, et al. Advances in Artificial Intelligence and Security. Cham: Springer International

- Publishing, 2022: 256-267.
- [6] 马昱堃. 作为有机体的技术——评《技术的本质：技术是什么，它是如何进化的》[J]. 智能社会研究, 2023, 2(2): 171-182.
- [7] 邬金鸣, 胡智杰, 姚茹, 等. 面向技术创新过程表征和描述的技术要素关联概念模型[J]. 图书情报工作, 2024, 68(10): 1-15.
- [8] PAN W, JIAN L, LIU T. Knowledge generation and diffusion in science & technology: an empirical study of SiC-MOSFET based on scientific papers and patents[J]. Technology Analysis & Strategic Management, 2024, 36(7): 1587-1603.
- [9] 何玉洁, 杜方, 史英杰, 等. 基于深度学习的命名实体识别研究综述[J]. 计算机工程与应用, 2021, 57(11): 21-36.
- [10] LEI J, TANG B, LU X, et al. A comprehensive study of named entity recognition in Chinese clinical text[J]. Journal of the American Medical Informatics Association, 2014, 21(5): 808-814.
- [11] ZHAO Y, LI K, WANG T, et al. Joint entity and relation extraction model based on directed-relation GAT oriented to Chinese patent texts[J]. Soft Computing, 2024(28): 7557-7574.
- [12] 刘浏, 王东波. 命名实体识别研究综述[J]. 情报学报, 2018, 37(3): 329-340.
- [13] ALI S, MASOOD K, RIAZ A, et al. Named entity recognition using deep learning: A review[C]//2022 International Conference on Business Analytics for Technology and Security (ICBATS). IEEE, 2022: 1-7.
- [14] GAO H, WANG T, LUO W, et al. Adversarial Multitask Learning for Technology Entity Recognition[C]//2018 IEEE International Conference on Progress in Informatics and Computing, 2018: 134-139.
- [15] MA J X, YUAN H. Bi-LSTM+CRF-based named entity recognition in scientific papers in the field of ecological restoration technology[J]. Proceedings of the Association for Information Science and Technology, 2019, 56(1): 186-195.
- [16] 于润羽, 杜军平, 薛哲, 等. 面向科技学术会议的命名实体识别研究[J]. 智能系统学报, 2022, 17(1): 50-58.
- [17] CHEN W Q, HAN P, ZHONG Y L, et al. Named Entity Recognition from[C]//Chinese Medical Literature Based on Deep Learning Method 2023 China Automation Congress (CAC), 2023: 9332-9337.
- [18] 王宇晖, 杜军平, 邵莹侠. 基于 Transformer 与技术词信息的知识产权实体识别方法[J]. 智能系统学报, 2023, 18(1): 186-193.
- [19] ZHAO Y, LI K C, WANG T, et al. Joint entity and relation extraction model based on directed-relation GAT oriented to Chinese patent texts[J]. Soft Computing, 2024: 1-18.
- [20] 钱时惕. 什么是科学——科学与人文漫话之一[J]. 物理通报, 2009(10): 56-58.
- [21] 库恩. 科学革命的结构[M]. 北京: 北京大学出版社, 2012: 58-72.
- [22] BASALLA G. The Evolution of Technology[M]. Cambridge: Cambridge University Press, 1989: 1-20.
- [23] 王颖洁, 张程烨, 白凤波, 等. 中文命名实体识别研究综述[J]. 计算机科学与探索, 2023, 17(2): 324-341.
- [24] BUNDY A. Australian and New Zealand information literacy framework[EB/OL]. (2004-01-01) [2024-04-23]. <http://www.anziil.org.index.htm>.
- [25] BUSH V. Science: The Endless Frontier[M]. Washington DC: National Science Foundation, 1960: 18-19.
- [26] GUAN J, HE Y. Patent-bibliometric analysis on the Chinese science—technology linkages[J]. Scientometrics, 2007, 72(3): 403-425.
- [27] VERBEEK A, DEBACKERE K, LUWEL M. Science

- cited in patents: A geographic “flow” analysis of bibliographic citation patterns in patents[J]. *Scientometrics*, 2003, 58(2): 241-263.
- [28] 韩芳. 基于专利引文的“科学—技术关系”及技术演化轨迹研究[D]. 北京: 北京邮电大学, 2017.
- [29] KE O. An analysis of the evolution of science-technology linkage in biomedicine[J]. *Journal of Informetrics*, 2020, 14(4): 101074.
- [30] 刘雨梦, 郑稣鹏. 基础科学、技术科学及工程应用的“链合”演化研究——国家级协同创新中心纵向案例[J]. *科学学研究*, 2022, 40(10): 1745-1755.
- [31] 吴宁, 杨艳萍. 基于多层网络的科学-技术微观知识流动分析框架研究[J]. *农业图书情报学报*, 2023, 35(11): 40-52.