arXiv:2303.00344v1 [cs.CL] 1 Mar 2023

利用外围上下文与时变增强进行文内引文分类

Priyanshi Gupta[1],Yash Kumar Atri[1],Apurva Nagvenkar[2], Sourish Dasgupta[2], Tanmoy Chakraborty[3]

[1]IIIT-Delhi，印度， [2]Crimson AI，印度， [3]IIT Delhi，印度{priyanshig,yashk}@iiitd.ac.in，{apurva.nagvenkar,sourish.dasgupta}@crimsoni.ai, tanchak@iitd.ac.in

摘要。 引文在确定研究论文之间的关联方面发挥着关键作用。它揭示了指示性、支持性或对比性研究中的重要信息。文内引文分类任务有助于推断这些关系；然而，现有研究仍不成熟，尚需进一步审视。目前用于文内引文分类的数据集和方法仅使用标记了引文的句子，导致模型忽略领域知识和邻近的上下文句子。本文提出一个名为 3Cext 的新数据集。除被引句子外，该数据集还利用邻近句子提供话语信息，以分析对比和蕴含关系，并提供领域信息。我们提出 PeriCite，这是一种基于 Transformer 的深度神经网络，可融合外围句子和领域知识。与最佳基线相比，我们的模型在 3Cext 数据集上取得了最先进水平， $+0.09$ F1。我们进行了广泛的消融实验，以分析所提出的数据集和模型融合方法的有效性。

关键词： 引文分类 · 文献计量学 · Transformer。

1 引言

过去几十年来，人们一直关注用于研究评估的引文分析。研究人员强调，需要采用新的方法，考量引证句的各种组成部分。一种广为人知的定性方法是分析研究论文中提及某项研究的句子，以确定引用背后的目的，从而评估其科学影响力。引文的上下文，即提及被引文献的文本，已被证明是引文意图的有效指标 [25]。衡量研究论文的科学影响力，需要从根本上理解引文意图。评估科学出版物重要性的一个好方法，是确定研究中为何引用相关文献，以及这些引文有多重要。

以往的引文语境分类方法采用了从低粒度到高粒度的各种标注技术。由于缺乏标准化的方法和标注方案，

极难对早期系统进行比较。3C共享任务 [12,13] 使用了学术引文类型（ACT）数据集中的一部分，通过查看引文句或包含引文的文本，将参考文献标记分类为“功能”或“目的” [19]。传统的引文分析仅依据引文数量，只考虑定量因素。引文语境分析在识别引文时面临的最大障碍之一，是缺乏多学科数据集，也没有能够充分表征其功能及影响的中等粒度到细粒度方案 [8]。为应对这一挑战，Kunnath 等人 [12]提出了一项名为 3C共享任务的统一任务，使用同一数据集比较多种引文分类方法，以弥补引文语境分类的不足。该任务第二轮迭代 [13] 的主要区别在于，子任务包含全文数据集。然而，即使有了全文，与引文句相关的元数据仍不足以理解引用的理由。

为缓解上述局限性，我们提出了一个名为 3Cext 的新数据集，以及一个名为 PeriCite 的新模型，该模型结合了增强和外围上下文的优势。实验表明，被引句子高度依赖外围上下文，通过对比或蕴含信息来强化论证。我们的主要贡献如下：

1. 我们扩展了 3C 数据集 [13] –3Cext。该数据集在提供引用句的同时，通过外围句子提供对比和蕴含信息，从而增加了更多话语信息。2. 我们提出了一种新颖的模型 PeriCite，该模型利用空间融合和跨文本注意力，关注外围句子的上下文信息，并采用时变增强来应对训练期间的类别不平衡。3. 我们还将所提出的模型与各种基线方法进行了比较，并通过消融实验和误差分析展示了该模块的有效性。

2 相关工作

引文对于说服至关重要，因为它们为作者提出的论断提供了依据。了解作者是否赞同或反对被引出版物中的论断非常重要，因为并非所有引文的用途都相同。为了根据语境对引文进行分类，此前已有大量研究考察了科学话语中使用的语言。研究者提出了多种框架来对引文意图进行分类 [16]。早期自动分类引文意图的工作采用了多种策略，包括基于规则的系统 [7,18], 、基于语言模式的机器学习技术[9], ，以及根据引文语境手工构建的特征。Teufel 等人[25]介绍了如何将学术文章中的引文标注为 12 个类别，并使用机器学习技术复现标注结果。这些类别

被划分为四个顶层组，分别是神经类别、明确指出缺陷的引文、进行对比或比较的引文，以及赞同、采用或与引用该文的作品相一致的引文。Abu-Jbara 等人 [1] 使用线性支持向量机，并利用词汇、结构和句法特征进行分类。此外，研究者还探讨了用于在所提及的出版物中定位引文片段的基于特征的技术 [5,4] 。使用基于卷积神经网络的模型[24]联合预测被引片段和引文功能，取得了更好的效果。

这些举措大多提供了过于细致的引文类别，其中一些在论文中很少使用。因此，它们对科学论文的自动化分析几乎没有帮助。Jurgens 等人 [10]开发了一种包含六个类别的方法，以纳入早期研究，并提出了一套更精确的分类方案，扩展了此前所有基于特征的引文意图分类工作。作者还新增了六个类别和 1,941 个来自计算语言学研究的样本，并采用了三个原有特征——基于模式、基于主题和基于原型论证。此外，他们还使用了结构拓扑、词汇语义、语法、字段位置和值，以及使用特征。

上述所有方法都使用人工设计的特征对数据进行分类。Cohan 等人 [3] 提出了一种神经多任务学习方法，使用非上下文化嵌入（GloVe [17]）和上下文化嵌入（ELMo [23], 双向长短期记忆网络和注意力方法）对引文意图进行分类。作者使用了两项辅助任务来支持主要分类任务，以实现多任务学习。他们近期的研究 [3] 仅纳入三个引文类别和来自计算机科学与医学领域的 11,020 个实例，构建了新的数据集（SciCite）。Beltagy 等人 [2] 发布了 SciBERT，这是一个基于 Semantic Scholar 中 1.14 百万篇论文进行预训练的模型。为支持该领域的研究，近期一项分析 [11] 评估了这一主题的 60 篇研究论文、研究人员在开展工作时遇到的困难，以及仍有待解决的知识空白。

3 方法论

本节将讨论我们提出的模型 PeriCite。该模型由两个堆叠的 Transformer 组成，每个 Transformer 包含四个模块。它利用跨文本注意力捕捉被引句子与外围句子之间的语篇关系。PeriCite 还包含时变增强，用于根据标签损失合成生成数据，并通过空间融合融合堆叠 Transformer 的最终表示。图 1 展示了 PeriCite 的示意图。

3.1 跨文本注意力

仅基于单个输入文本进行注意力计算可能无法向模型提供充分的信息。然而，当与外围上下文注意力融合时，模型能够学习与标签相关的重要片段。为融合外围

上下文注意力与主要输入文本，我们提出了跨文本注意力（CTA）。它计算主要文本与外围上下文之间的成对权重。

给定自注意力机制如下

$$Attn_{\text{self}}(\mathbf{x}) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V,$$

首先，针对主输入文本生成查询 $Q \in \mathbb{R}^{N\times d_k}$、键 $K \in \mathbb{R}^{N\times d_k}$和值 $V \in \mathbb{R}^{N\times d_v}$，其维度分别为 d_k 和 d_v。接下来，为计算跨文本注意力分数，使用以下方式计算主输入与外围上下文之间的成对权重：

$$Attn_{bidir}(x_s, x_t) = Attn_{cross}(x_t, x_s) + Attn_{cross}(x_s, x_t),$$
$$Attn_{cross}(x_t, x_s) = softmax\left(\frac{Q_t K_s^T}{\sqrt{d_k}}\right)V_s,$$
$$Attn_{cross}(x_s, x_t) = softmax\left(\frac{Q_s K_t^T}{\sqrt{d_k}}\right)V_t$$

其中，x_s 表示主输入文本的上下文表示，x_t 表示外围文本的上下文，Q、K 和 V 在 s 和 t 的条件下，分别表示基于主文本和外围文本的查询、键和值。随后，我们对注意力头进行线性投影，以捕获语言理解信息。最后，将计算得到的注意力权重传递至前馈层。

3.2 空间融合

由于融合模块结合了来自两个外围编码器的生成式注意力向量，因此确定哪个向量更为重要、它如何作用于主上下文以及各向量之间如何交互至关重要。基于这一认识，我们采用了基于空间融合（SF）的融合策略。空间融合最初由 Li 等人 [15] 提出，用于融合多幅图像以解析深层特征。在我们的设置中，我们将其扩展为融合文本，从而形成深层上下文特征。

按照 [27], 中的建议，在空间融合中，全局特征由自注意力层建模，该层由卷积层和降维层构成。这两者的结合能够捕捉特征集之间的局部关系和全局关系。此外，我们将获得的特征传入一系列卷积块，以在融合过程中增强上下文信息。最后，生成一张有序加权平均图，将特征与源文本合并。

多尺度特征矩阵记为 a，其中 a 是张量列表$a \in \{1, 2, \cdots, a\}$，$a = 4$。融合机制有两个输入：一个来自引用句，记为 P_1^a；另一个来自外围句子，记为 P_2^a。融合后生成空间注意力 [15] C_f^a，并将其输入最终解码层。

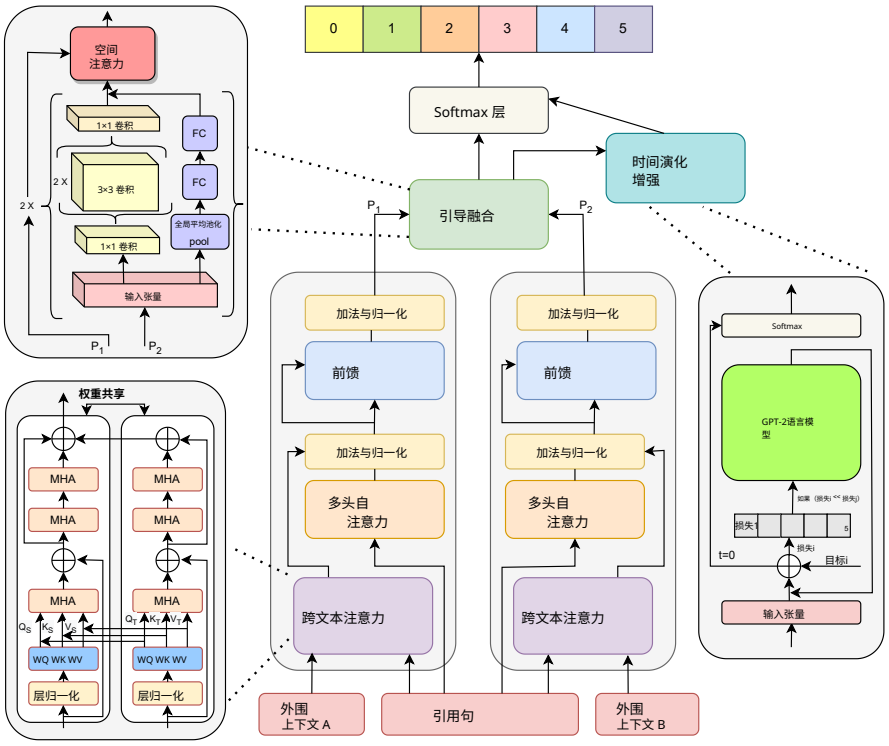

图1. 示意图 PeriCite：它由两个并行堆叠的 Transformer 块组成。跨文本注意力在主上下文与外围句子之间计算。Transformer 块的输出通过引导融合进行融合。基于小批次中标签表示的时变增强会生成合成训练样本。

有序加权图通过对 P_1^a和 P_2^a进行 L1 归一化和 softmax 计算得到，分别生成权重矩阵 W_1^a和 W_2^a。最终的有序加权图通过公式 1 计算如下：

$$S_j^a(m,n) = \frac{P_j^a(m,n)}{\sum_{i=1}^{j} ||P_i^a(m,n)} \tag{1}$$

此处，对 P_j^a和 P_i^a计算 L1 范数，其中 j的取值范围为 $[0,2]$集。特征集 P_j^a和 P_i^a中的位置由 $(m,\ n)$ 表示，固定向量大小为 V。 $P_j^a(m,\ n)$ 也输出一个大小为 V 的向量。

根据公式 2，利用 α_k^m对权重矩阵 W_1^a和 W_2^a上的特征向量 P_1^a和 P_2^a进行加权，从而进一步增强，

$$P_j^a(m,n) = \alpha_j^a(m,n) \times P_j^a(m,n) \tag{2}$$

最后，通过将其与增强后的特征集进行投影，使用式 3 计算融合向量 C_f^a

$$C_f^a(m,n) = \sum_{i=1}^{j} P_j^a(m,n) \tag{3}$$

3.3 时间演化增强

数据增强有助于生成合成数据并平衡数据集。然而，大多数数据增强技术都是基于对少数类进行随机变换来生成合成数据的。如表 2 所示，我们提出的数据集 3Cext 也存在严重的类别不平衡问题，其中多数类的样本数量是第二大类的三倍。处理类别不平衡最常见的两种方法，是让训练集中的所有类别都具有相同的代表性，或者增加少数类的样本，使其数量与多数类相当。然而，对于行内引文分类任务，这两种方法都会进一步降低模型在结构复杂性和信息传播方面的性能。为解决这些局限性，我们提出了时变增强（TEA）。

在每个时间步 t，TEA 都会计算每个小批次 m 中的标签表示，记为 $s_i = [l_1, l_2, l_3, l_4, l_5]$。对于在时间步 t计算的损失，模型会计算给定小批次 m 中每个标签对应的损失 l_i，并构建损失与标签之间的关系 $loss \rightarrow label$，记为 $losslabel_i$。给定分布 $losslabel_i$，TEA 会为给定小批次中损失最高的少数类别合成生成训练样本。数据样本通过 GPT2 语言模型 [20]进行增强。
$loss \rightarrow label$ 表示独立于各类别样本数量的全局表示，只考虑给定小批次的表示。该方法有助于模型在每一步控制每个标签的损失。每个标签的表示损失是引导 TEA 在每个时间步演化的因素，有助于模型在训练阶段学习到均衡的表示。

4 实验

4.1 数据集

本节将详细介绍我们提出的 3Cext 数据集。Kunnath 等人 [12] 介绍了 ACT 数据集，其中包含由 883 位作者标注的 11，233 条引文。被引标签被屏蔽，并以“#AUTHOR TAG”表示被引对象的位置。此外，3C 数据集包含全文以及表示粒子引文类别的标签（参见表 2）。

在我们的工作中，我们扩展了 3C 数据集，以收录更多话语信息，从而更好地解释句子中出现引文的原因。我们的直觉是，被引句子大多与相邻句子构成蕴含或对比关系。为了捕捉外围句子，我们提取了数据集中与 COREID（唯一论文编号）对应的全文文件，以继续探索这一发现。

第一句	引用句	第二句
[CITE] 描述一种混合推荐系统 推荐系统 该 利用本体来提高 用户画像构建 过程的准确性，从而提升 推荐结果的实用性。	#AUTHOR TAG 使用不同的 通过表示.....策略 将用户画像表示为词袋 and 根据每个词项进行加权， 依据用户兴趣 从领域本体中派生 ogy	Razmerita 等人 [CITE] de- 描述 OntobUM，这是一款基于本体的 推荐系统，整合了 三个本体：i） 用户本体， 它 构 的特征 用户及其关系， ii）领域本体，它 定义了该领域
基于内容的推荐 系统 [CITE] 依赖于预先- 现有领域知识来 推荐更相似的项目 用户似乎喜欢的那些 项目。它们通常会生成 描述用户的用户模型 兴趣的特征 [引用]。	此应用程序接口支持多种应 用，包括 智能图书 推荐系统， 智能主题 矿工 [引文]， 技术-主题框架 工作 #AUTHOR TAG，一个系 统，用于预测技术的传 播在各个领域中的情况 研究社区，以及 实用 本体 Evolu- 化框架 [CITE]	<EOF>

表 1. 3Cext 数据集实例。第一句表示前置句子，引用句表示主要的被引句子，第二句表示后置句子。如果被引句子是首句或末句，我们会将第一句或第二句标记为 EOF。

数据集	类别					
	背景	比较-对比	扩展	未来工作	动机	用途
3Cext	1318	380	294	221	137	50

表 2. 各类别的实例数量。类别包括背景、比较-对比（COM-CONT：比较与对比）、扩展（EXTENSN：扩展）、未来工作（FUTR：未来工作）、动机（MOTIVN：动机）和用途。

对于给定文档，我们会定位主要引用句的位置，并找出其前置句和后置句。我们采用正则表达式、莱文斯坦距离等启发式方法，以及句号识别和作者姓名识别等硬性规则，来标记这三个句子。在我们的数据集中，第一个句子表示引文前的前置句，第二个句子表示引文后的后置句。六个类别分别对应标签 0 至 5：0 - 背景，1 - 比较与对比，2 - 扩展，3 - 未来，4 - 动机，5 - 用途（如 [12] 中所示）。表 1 展示了 3Cext 中的一个样例，表 2 列出了各类别的实例数量。

4.2 实现细节

数据集设置： 在数据集部分，我们首先通过移除停用词、转换为小写并删除所有特殊字符来预处理数据。我们将训练和测试实例中的句子截断为 256 个词元长度。我们使用 2400 个实例作为训练集，600 个实例作为测试集。

基线系统	数据集					
	3C			3Cext		
	精确率	召回率	F1分数	精确率	召回率	F1分数
多层感知机	0.30	0.38	0.30	0.39	0.32	0.39
LSTM-Attention-Scaffold	0.27	0.48	0.46	0.30	0.55	0.39
Transformer	0.38	0.37	0.34	0.36	0.38	0.37
DistilBERT	0.43	0.48	0.37	0.46	0.55	0.40
BART	0.41	0.46	0.34	0.47	0.52	0.41
T5	0.41	0.43	0.35	0.43	0.48	0.38
SciBERT	0.45	0.51	0.46	0.52	0.53	0.51
PeriCite						
PeriCite（含 SF）	0.38	0.37	0.34	0.36	0.38	0.37
PeriCite 搭配 CTA	0.34	0.36	0.32	0.38	0.41	0.37
PeriCite 搭配 TEA	0.36	0.43	0.41	0.38	0.42	0.39
PeriCite 搭配 TEA、SA、CTA	0.46	0.44	0.42	0.60	0.63	0.60

表3. 在 3C 和 3Cext 数据集上的性能基准。我们提供了六种分类基线方法，以及 PeriCite 的不同消融版本。

5 基线方法

我们将详细讨论基线系统。对于基于语言模型（LM）的基线方法，我们使用 3C 和 3Cext 的训练样本对 LM 进行微调，直至收敛。

1. **多层感知机：** 我们使用三个堆叠的全连接层 [6]，并采用 softmax激活函数。我们使用 Glove嵌入作为输入表示，并以交叉熵作为损失。2. **带注意力和支架机制的双向长短期记忆网络 [3]：** 该基线使用带注意力机制的双向长短期记忆网络，以 Glove 作为输入嵌入，以 Elmo 作为上下文表示。我们使用一个包含 20 个节点的多层感知机来执行支架任务。对于该基线，我们保留所有原始超参数。3. **Transformer：** 我们使用原始 Transformer [26] 架构作为基线运行。编码器和解码器分别使用 4 个堆叠层，最大序列长度为 32，激活函数采用 softmax，损失采用交叉熵。4. **DistilBERT [22]：** 它遵循与 BERT 相同的架构，但在预训练过程中利用知识蒸馏减少了参数数量。我们使用 Hugging Face 移植的模型作为基线。5. **SciBERT：** SciBERT [2] 采用标准的 BERT 架构，并在科学文献上进行预训练。其超参数与 Transformer 基线相似。6. **BART：** 与 SciBERT 类似，BART [14] 使用双向编码器和自回归解码器。它在 Books 和维基百科数据上进行预训练。我们使用与 SciBERT 基线相似的超参数。

7. **T5：** 它是一种基于预训练 Transformer 的编码器-解码器语言模型[21]。它在各种有监督和无监督下游任务上作为文本到文本 Transformer 进行预训练。

6 分析

我们对提出的 PeriCite 模型进行消融实验，以展示各模块的有效性。我们证明，仅外围上下文就能通过提供上下文信息显著提升模型性能。加入 TEA 后，各类别的性能均得到提升，这表明受控合成生成训练数据能够改善系统的整体性能。表3显示，与最佳基线相比，结合 CTA、SA 和 TEA 后，我们的模型取得了 $+0.09$ 个 F1分数点的提升。每个模块都带来了性能改进。引入 CTA 后，与基础 Transformer 网络相比，我们的模型的召回率提高了 $+0.02$。加入 TEA 后，与 Transformer 相比，F1分数提高了 $+0.02$，召回率提高了 $+0.04$。

模型	类别	精确率	召回率	F1分数
PeriCite	背景	**0.67**	**0.83**	0.74
	比较对比	**0.49**	**0.28**	0.36
	扩展	0.30	0.09	0.14
	未来	0.33	0.17	0.22
	动机	0.55	0.31	0.40
	USES	0.61	0.61	0.62

表4. 所提出模型的分类性能指标。我们报告每个类别的精确率、召回率和F1分数。

表6展示了 PeriCite 各类别的性能，表明我们的模型能够捕捉所有类别的上下文信息。与表6中表现最佳的基线（SciBERT）的混淆矩阵相比，我们发现该基线严重偏向多数类，几乎将所有其他类别都预测为 0。然而，我们的模型能够较为均衡地预测所有类别。我们还在表5中分析了模型的预测错误。对于第二行中的句子，模型可能会被短语“我们评估相似性”分散注意力，从而认为它属于“用途”类别。第三行也可能会受到短语“遵循反思过程”的干扰。这些类别的训练样本数量极少，可能是导致错误标注的原因。提供更多上下文信息和大量定性训练样本，有助于提升模型性能。

正文	真实	预测
所谓的情景预见（#AUTHOR TAG，2010），以及自传式记忆和心智理论，也构成了我们走神内容的大部分（Spreng 和 Grady，2009），因为我们会预想未来的某项活动，或思考可能的未来选项，以便选择合适的行动	1	2
我们使用余弦相似度来评估两个语义向量的相似性 #AUTHOR TAG，因为该度量依据向量的方向而非大小向量空间中的主题权重，从而能够比较章节数量不同的编辑产品相关联的编辑产品	0	5
遵循 #AUTHOR TAG (1996) 提出的反思过程，在新的版本中，将4个问题的首词添加为视觉提示	0	5
不过，一些较新的模型也纳入了领域专家，以便定义教育游戏的学习内容（#AUTHOR TAG 等人，2017）	4	1

表5. PeriCite 在 3Cext 数据集上的分类错误示例。

		预测					
		背景	**Com-Cast**	**扩展**	**未来**	**动机**	**用途**
Actual	**背景**	280/320	13/0	5/0	1/0	9/0	27/0
	Com-Cast	46/73	21/0	1/0	0/0	3/0	3/1
	扩展	25/34	0/0	3/0	1/0	2/0	3/0
	未来	9/12	1/0	0/0	2/0	0/0	0/0
	动机	28/55	3/0	1/0	1/0	17/0	5/0
	Uses	30/79	5/0	0/0	1/0	0/0	59/16

表6。 PeriCite / SciBERT 上的混淆矩阵（最佳基线）。这些类别分别为背景、Com-Cast：比较-对比、扩展、未来、动机和用途。

7 结论

本文提出了一个新的数据集 3Cext。我们在 3C 数据集的基础上引入外围上下文信息，以分析对比性和蕴含性信息。我们还提出了一种新颖的模型 PeriCite，它利用跨文本注意力关注引文输入和外围句子中的上下文信息。我们还引入了时变增强，在每个时间步为少数类生成合成数据，并采用空间融合来关注输入空间中的关键信息。与表现最佳的基线相比，我们提出的模型在 3Cext 上取得了新的最先进水平，F1分数提高了 $+0.09$。我们还进行了广泛的消融实验，以分析所提出的数据集和模型融合方法的有效性。未来工作中，一个值得探索的方向是利用各章节的话语信息，为文内引文提供更多上下文。相邻句子中的对比性或蕴含性信息对于理解引文意图背后的推理至关重要。基线推荐、科学论文推荐等其他任务，也能随着该任务性能的提升而获得显著改善。

8 致谢

本文所报告的研究由 Crimson AI Pvt. Ltd. 资助。

参考文献

1. Abu-Jbara, A., Ezra, J., Radev, D.：引文的目的与极性：迈向基于自然语言处理的文献计量学。载于：NAACL。第596–606页（2013） 2. Beltagy, I., Lo, K., Cohan, A.：SciBERT：用于科学文本的预训练语言模型。arXiv 预印本 arXiv:1903.10676（2019） 3. Cohan, A., Ammar, W., van Zuylen, M., Cady, F.：科学出版物中引文意图分类的结构性支架。载于：NAACL。第3586–3596页。ACL，明尼阿波利斯，明尼苏达州（2019年6月） 4. Cohan, A., Goharian, N.：利用词嵌入和领域知识对引文进行语境化，以实现科学摘要。载于：ACM SIGIR。第1133–1136页（2017） 5. Cohan, A., Soldaini, L., Goharian, N.：生物医学文献中引文文本与被引文本片段的匹配：一种面向检索的方法。载于：NAACL。第1042–1048页（2015） 6. Gardner, M.W., Dorling, S.：人工神经网络（多层感知器）——大气科学中的应用综述。大气环境 **32**(14-15)，2627–2636（1998） 7. Garzone, M., Mercer, R.E.：迈向自动引文分类器。载于：加拿大计算智能研究学会会议。第337–346页。施普林格（2000） 8. Hern´andez-Alvarez, M., G´omez, J.M.：引文影响分类：面向科学文献。载于：2015 IEEE ICCSE第18届计算科学与工程国际会议。第307–313页。IEEE（2015） 9. Ikram, M.T., Afzal, M.T.：利用语言模式进行基于方面的引文情感分析，以更好地理解科学知识。科学计量学 **119**(1)，73–95（2019年4月）。https://doi.org/10.1007/s11192-019-03028- 10. Jurgens, D., Kumar, S., Hoover, R., McFarland, D., Jurafsky, D.：通过引文框架衡量科学领域的演变。计算语言学协会汇刊 **6**，391–406（2018） 11. Kunnath, S.N., Herrmannova, D., Pride, D., Knoth, P.：引文语义分类的元分析。定量科学研究 **2**(4)，1170–1215（2022） 12. Kunnath, S.N., Pride, D., Gyawali, B., Knoth, P.：2020年 WOSP 3C 引文上下文分类任务概览。载于：科学出版物挖掘第8届国际研讨会论文集。第75–83页。计算语言学协会（2020） 13. Kunnath, S.N., Pride, D., Herrmannova, D., Knoth, P.：2021年 SDP 3C 引文上下文分类共享任务概览。计算语言学协会（2021） 14. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L.：BART：用于自然语言生成、机器翻译和理解的去噪序列到序列预训练。arXiv 预印本 arXiv:1910.13461（2019） 15. Li, H., Wu, X.J., Durrani, T.：NestFuse：一种基于嵌套连接和空间/通道注意力模型的红外与可见光图像融合架构。IEEE 仪器与测量汇刊 **69**(12)，9645–9656（2020） 16. Moravcsik, M.J., Murugesan, P.：关于引文功能与质量的一些结果。科学社会研究 **5**(1)，86–92（1975） 17. Pennington, J., Socher, R., Manning, C.：GloVe：用于词表示的全局词向量。载于：EMNLP。第1532–1543页。ACL，多哈，卡塔尔（2014年10月）。https://doi.org/10.3115/v1/D14-1162，https://aclanthology.org/D14-1162

18. Pham, S.B., Hoffmann, A.：一种使用提示短语进行科学引文分类的新方法。载于：澳大利亚联合人工智能会议。第759–771页。施普林格（2003）19. Pride, D., Knoth, P., Harag, J.：Act：一个用于大规模引文类型标注的标注平台。载于：ACM/IEEE 数字图书馆联合会议。第329–330页。IEEE（2019）20. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.：语言模型是无监督的多任务学习器（2019）21. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J., 等：探索统一文本到文本转换器的迁移学习极限。《机器学习研究杂志》**21**(140)，1–67（2020）22. Sanh, V., Debut, L., Chaumond, J., Wolf, T.：DistilBERT，BERT 的蒸馏版本：更小、更快、更便宜、更轻量。CoRR **abs/1910.01108**（2019）23. Sarzynska-Wawer, J., Wawer, A., Pawlak, A., Szymanowska, J., Stefaniak, I., Jarkiewicz, M., Okruszek, L.：利用深度上下文化词表示检测形式思维障碍。《精神病学研究》**304**，114135（2021）24. Su, X., Prasad, A., Kan, M.Y., Sugiyama, K.：用于引文功能和来源追溯的神经多任务学习。载于：2019年 ACM/IEEE 数字图书馆联合会议（JCDL）。第394–395页。IEEE（2019）25. Teufel, S., Siddharthan, A., Tidhar, D.：引文功能的自动分类。载于：EMNLP 2006年自然语言处理实证方法会议论文集。第103–110页（2006）26. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.：注意力就是你所需要的一切。NeurIPS 神经信息处理系统进展**30**（2017）27. Vs, V., Valanarasu, J.M.J., Oza, P., Patel, V.M.：图像融合转换器。载于：2022年 IEEE 国际图像处理会议（ICIP）。第3566–3570页。IEEE（2022）