第 62 卷　第 5 期
2026 年 9 月

南京大学学报(自然科学)
JOURNAL OF NANJING UNIVERSITY
(NATURAL SCIENCE)

Vol. 62, No. 5
Sept., 2026

DOI:10.13232/j.cnki.jnju.2026.05.009

# 基于数据增强与对比学习的事件因果关系识别方法

李佳恩[1],李　旸[2*],李德玉[1,3],王素格[1,3]

(1.山西大学计算机与信息技术学院,太原,030006;2.山西财经大学金融学院,太原,030006;
3.计算智能与中文信息处理教育部重点实验室,山西大学,太原,030006)

**摘　要:**事件因果关系识别(Event Causality Identification,ECI)作为自然语言理解的基础任务,旨在从文本中判定事件对之间的因果关系,对问答系统、叙事生成等应用具有重要作用.现有方法存在明显局限:一是针对标注数据稀缺与分布不均衡的问题,若引入外部知识库,容易受其知识覆盖度低与噪声干扰;二是针对因果与非因果样本在语义空间中边界模糊问题,传统方法难以学习判别性表征.针对上述问题,提出一种基于反事实数据增强与监督式对比学习的ECI方法.构建了"知识库筛选-因果结构约束-LLM (Large Language Model)生成与校验"三重数据质量控制机制,即从结构化知识库中筛选语义相关的中间概念,注入多种基础因果图结构以约束生成逻辑,并利用大语言模型进行生成与多维度校验,从而批量构建高质量、多样化的反事实样本.进一步,引入监督式对比学习,显式优化语义表示空间,增强模型对因果边界的判别能力.实验结果表明,提出的方法在EventStoryLine和Causal-TimeBank基准数据集上的$F1$分数分别达到了74.2%和80.4%,显著优于现有主流模型.

**关键词:**事件因果关系识别,反事实数据增强,监督式对比学习,因果结构约束,语义表征优化

**中图分类号:**TP391　　**文献标志码:**A

# Event causality identification based on data augmentation and contrastive learning

Li Jiaen[1], Li Yang[2*], Li Deyu[1,3], Wang Suge[1,3]

(1. School of Computer and Information Technology, Shanxi University, Taiyuan, 030006, China;
2. School of Finance, Shanxi University of Finance and Economics, Taiyuan, 030006, China;
3. Key Laboratory of Computational Intelligence and Chinese Information Processing, Ministry of Education, Shanxi University, Taiyuan, 030006, China)

**Abstract:** Event Causality Identification (ECI), a fundamental task in natural language understanding, aims to determine causal relationships between event pairs from text and plays a crucial role in applications such as question-answering systems and narrative generation. Existing methods exhibit notable limitations: first, regarding the scarcity and imbalance of annotated data, incorporating external knowledge bases is prone to issues of limited coverage and noise interference; second, with respect to the blurred semantic boundaries between causal and non-causal samples, traditional approaches struggle to learn discriminative representations. To address these challenges, this paper proposes an ECI method based on counterfactual data augmentation and supervised contrastive learning. We design a three-stage data quality control mechanism "knowledge base filtering, causal structure constraint, and LLM (Large Language Model) generation and verification". Specifically, we first extract semantically relevant intermediate concepts from structured knowledge bases. Then, we inject various foundational causal graph structures to constrain the logical coherence of generated samples. Finally, we leverage LLMs to generate and

基金项目:国家自然科学基金(62376143,62473241),山西省基础研究计划(202503021211239)

收稿日期:2026－06－16

*通信联系人,E-mail:liyang@sxufe.edu.cn

multi-dimensionally verify these samples, enabling the batch construction of high-quality, diverse counterfactual examples. Furthermore, we integrate supervised contrastive learning to explicitly optimize the semantic representation space, thereby enhancing the model's capacity to discriminate causal boundaries. Experimental results demonstrate that our method achieves $F1$ scores of 74.2% and 80.4% on the EventStoryLine and Causal-TimeBank benchmark datasets, respectively, significantly outperforming current state-of-the-art models.

**Keywords:** event causality identification, counterfactual data augmentation, supervised contrastive learning, causal structure constraint, semantic representation optimization

ECI（Event Causality Identification）任务的核心目标是从自然语言文本中自动判断两个事件之间是否存在因果关联. 如图1所示，事件“crossfire”（交火）与“killed”（死亡）以及“injured”（受伤）之间分别存在明确的因果关系，即：（1）crossfire $\xrightarrow{\text{cause}}$ killed；（2）crossfire $\xrightarrow{\text{cause}}$ injured. 而句中的“investigation”（调查）与“trial”（审判）等其他事件则与“crossfire”（交火）无直接因果联系，在图中用虚线标示.

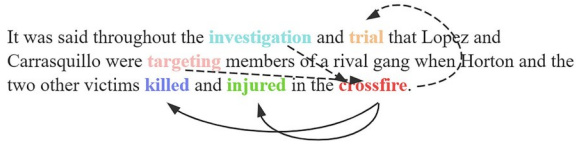



**图1　对ECI任务的介绍**

**Fig. 1　The introduction to the ECI task**

这项任务对于许多自然语言处理应用具有重要意义，例如，ECI已被应用于问答系统[1−3]、叙事生成[4]、文本总结[5]、机器学习理解[6]及事例图谱构建[7−8]等多个重要的研究方向. 此外，现实场景中也有许多实际应用，如事件预测[9−10]、策略优化[11].

现有的方法主要采用图结构来建模事件间的关系[12−14]，以发现事件及其之间的潜在关系. 具体地，使用文档级编码器初始化事件节点的嵌入，并采用图推理机制更新其嵌入向量，还结合了因果标签信息和事件对交互信息，以增强图中事件节点的表征学习. 同时，提示学习范式[15]通过充分挖掘预训练语言模型潜力为ECI任务提供了新思路. DPJL模型[16]设计了一个主提示任务，还设计了两个衍生提示任务. 尽管DPJL实现了较高的性能，但其设计涉及多提示的精细设计，并依赖主任务与衍生任务之间的正相关性. 另外，研究者利用外部知识提升ECI模型的性能[17−18]，但这些方法可能引入偏差与噪声.

为了克服上述限制，本文探索一种基于数据增强与对比学习的事件因果关系识别方法（DA-CL）. 其核心思想是构建“知识库筛选+因果结构约束+LLM（Large Language Model）生成与校验”三重机制，生成高质量反事实数据，达到数据增强的目的，并引入对比学习，采用对比损失与交叉熵损失结合的训练方式，进一步提升模型区分正负样例的能力.

## 1　相关工作

ECI任务中，早期的研究方法主要依赖明确的因果模式[19−20]、词汇和句法特征[20−21]以及因果指示或信号[22−23].

基于图的方法通常将ECI任务建模为节点分类问题，利用图神经网络，基于上下文语义学习事件节点的表示[24]. 例如，Fan et al[25]建立了事件、事件提及和上下文之间的明确联系，构建了事件节点表示和因果关系识别的共现图. 除了节点分类问题，一些研究还将ECI任务视为基于图的边预测问题. 例如，Zhao et al[12]使用文档级编码器初始化事件节点的嵌入向量，并利用图推理机制更新并预测因果边缘.

随着预训练语言模型（Pre-Trained Language Model，PLM）[26−28]的广泛应用，提示学习已成为各种NLP（Natural Language Processing）任务的新范式. 该方法将下游任务重新形式化为与预训练阶段相似任务的结构，有效缩小了PLM与下游任务之间的差距，促进知识从预训练模型向特定任务的迁移. 例如，Man et al[29]引入依赖路径生成方法，明确增强了因果推理过程. Hu et al[30]利用事件中心结构、事件相关结构及抽象意义表示

(Abstract Meaning Representation, AMR)解析器构建语义表示,进一步强化事件关联捕捉. 还设计了合适的提示模板,将ECI任务重新定义为填空任务. 例如,Liu et al[31]通过掩蔽语言模型来整合外部知识,并预测掩码位置的答案词. Liang et al[32]的ICCL模型融合了上下文学习概念,将任务中的特定实例加入提示模板,增强了因果推理.

知识增强的方法也是解决ECI任务的重要手段,其核心是整合外部知识库中丰富的事件信息,相关研究多采用ConceptNet[33],WordNet[34]等知识库[35−37]. 例如,Liu et al[38]和Liu et al[39]通过整合ConceptNet中的常识知识,提升了模型对因果关系的识别能力. Cao et al[40]将外部知识与模型表征融合,丰富从PLMs中所导出的特征表示. Zuo et al[41]通过生成额外训练数据的方式有效缓解了数据稀缺问题,为模型训练提供了充足的支撑,并为基于PLM的方法提供了重要的知识补充.

## 2 DACL方法

**2.1 模型整体架构** ECI任务的输入是一个三元组$(S, \langle e_1\rangle, \langle e_2\rangle)$,其中,$S$表示源文本,$\langle e_1\rangle$和$\langle e_2\rangle$是文本中标注的事件提及. 输出是一个二值分类标签$y\in\{0,1\}$,其中,$y=1$表示存在因果关系,$y=0$表示不存在. 图2展示了本文提出的DACL方法的模型整体架构. 为了缓解标注数据稀缺与分布不均衡问题,该方法构建了四个模块:概念引导的反事实数据增强模块(Concept-Guided Counterfactual Data Augmentation, CDA),生成高质量反事实样本;基于LLM引导的质量验证模块(LLM-Guided Quality Verification, LQV),对生成样本进行多维度质量评估与严格筛选,确保仅保留高质量的增强数据;监督对比因果学习模块(Supervised Contrastive Causal Learning, SCCL),显式拉近因果事件对、推远非因果事件对的语义表征距离,增强模型对因果边界的判别能力;因果预测模块(Causality Prediction Module, CPM)是模型的主干分类器,负责对事件对的因果关系进行最终判别,该模块与SCCL共享同一PLM编码器. 最后,模型采用对比损失与交叉熵损失联合优化的训练策略,实现对事件因果关系的有效识别.

**2.2 CDA模块** 该模块旨在解决ECI任务中的标注数据稀缺与分布不均衡问题. 与现有的数据增强方法不同,CDA不直接使用外部知识中的内容,而是基于原始标注样本生成标签翻转的反事实样本,减少远程监督带来的标签噪声问题. 同时,针对现有生成式数据增强方法缺乏因果逻辑约束的缺陷,CDA构建了"知识库筛选-因果结构约束-LLM生成与校验"三阶段数据生成机制,确保生成样本既具有语义相关,还要具有因果逻辑严谨、模式多样的特性.

(1)基于ConceptNet的知识库筛选阶段. 针对原始数据集中标注的因果事件对$(e_1, e_2)$,依托ConceptNet知识库,采用多跳路径检索算法探索事件对之间的潜在语义关联. 依据因果关联强度通常随路径长度增加而衰减的特性,设定最大路径长度阈值为$k$,当事件对在ConceptNet中的最短路径长度$L\leqslant k$时,保留该路径并剔除首尾事件节点. 为了筛选与事件对语义最相关的中间概念节点,通过计算事件对与路径节点的语义相似度筛选最优中间节点$c_m$:

$$c_m=\underset{c\in C}{\operatorname{argmax}}\left[\lambda\cdot\operatorname{sim}(e_1,c)+(1-\lambda)\cdot\operatorname{sim}(e_2,c)\right] \tag{1}$$

其中,$C$为路径中间节点集合,$\operatorname{sim}(\cdot)$为基于SBERT(Sentence-BERT)[42]的语义相似度计算函数,$\lambda$为平衡超参数. 当最短路径长度$L>k$时,激活大语言模型的推理能力,生成符合常识逻辑的中间节点.

(2)多因果结构约束阶段. 现有ECI研究多将因果结构用于模型推理阶段的关系建模,尚未将其作为生成约束引入数据增强过程. 为了解决这一问题,该阶段系统性地引入链式、对撞与叉状三种因果图结构,将其作为生成反事实样本的逻辑框架,确保生成样本的合理性与多样性.

表1详细阐述了这三种因果结构的图示、因果特征并提供了具体示例. 具体地,链式结构引入中间事件节点来建模多跳间接因果关系,旨在增强模型对长距离因果链的推理能力;对撞结构引入共同结果节点,训练模型识别具有共同后果的事件对,提升其对因果独立性的识别能力;叉状结构侧重于共同原因场景,使模型能够有效区分

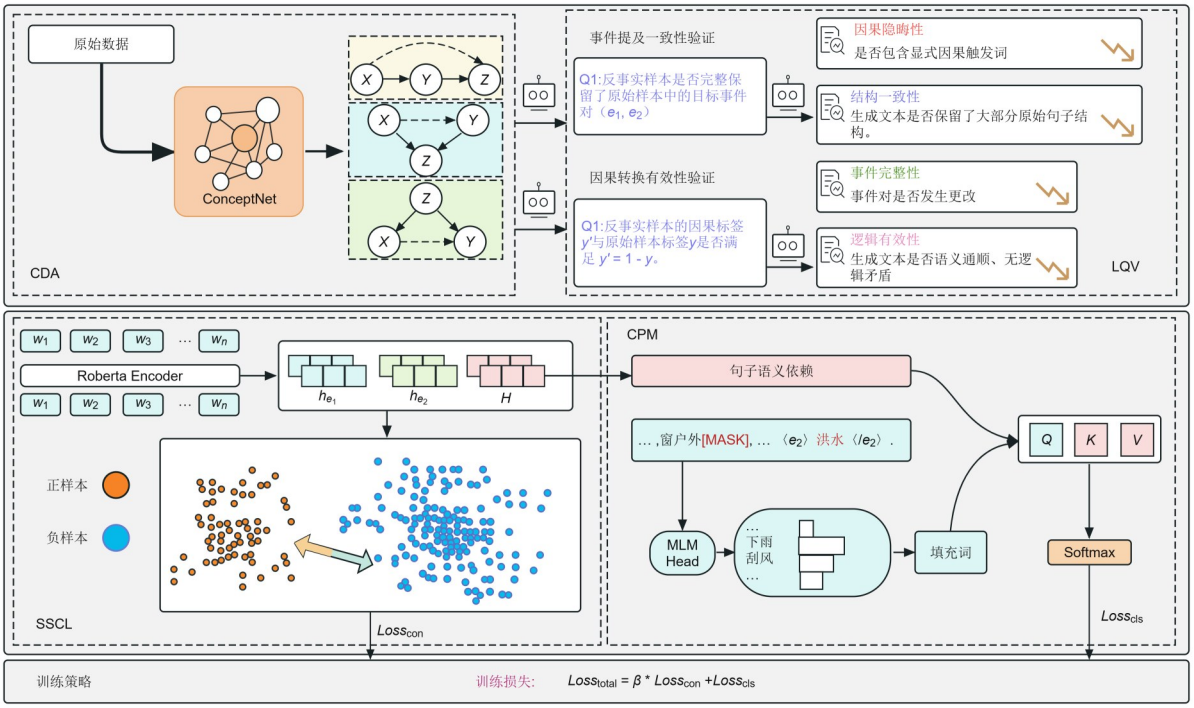



**图2　模型的整体框架**

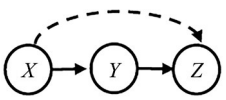
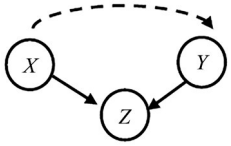
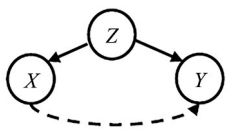
**Fig. 2　The overall framework of the model**

真实因果关系与单纯关联. 和完全依赖LLM的自由生成相比,结构化的指导能为数据生成过程提供明确的因果逻辑框架,不仅确保了生成样本在语义上的合理性,更显著提升了其因果逻辑的可靠性与生成过程的可控性,有效减少了LLM可能产生的语义偏离或逻辑错误,为后续的模型训练奠定了高质量、多样化的数据基础.

(3)基于反事实推理的样本生成阶段. 现有生成式数据增强方法大多是仅仅通过表层改写或自由续写构造增强样本,容易产生语义偏离、因果逻辑不成立的弱样本. 因此,此阶段在知识库筛选和因果结构约束的基础上,进一步构建基于反事实推理的样本生成方法. 具体地,给定一个原始样本$(S, e_1, e_2, y)$, 其中,$y \in \{0, 1\}$为因果标签,

**表1　常用因果结构的介绍**

**Table 1　Introduction to common causal structures**

| 因果结构 | 因果图 | 因果特征 | 举例 |
|---|---|---|---|
| 链状结构 |  | 在因果链中,中介事件$Y$连接事件$X$和事件$Z$,其中$X$导致$Y$,而$Y$随后导致$Z$. | 持续强降雨[$X$]导致城市排水系统超负荷运行[$Y$],进而使得城市主干道出现积水内涝[$Z$]. |
| 对撞结构 |  | 存在一个与事件$X$和$Y$都有联系的事件$Z$,其中$X$和$Y$均可导致$Z$. | 长期熬夜[$X$]会打乱生物钟,饮食不规律[$Y$]会影响营养吸收,二者均会导致内分泌失调[$Z$]. |
| 叉状结构 |  | 存在一个与事件$X$和$Y$都有联系的事件$Z$,其中$Z$可以导致$X$和$Y$. | 夏季突发强降雨[$Z$],市区多条低洼路段出现明显积水[$X$],同时路上的市民几乎都随身携带了雨伞或雨衣[$Y$]. |

该阶段在保持事件对$(e_1, e_2)$不变的条件下,通过添加中间节点$c_m$和因果结构,构造反事实样本$(S', e_1, e_2, 1-y)$.其中,$S'$是LLM在因果结构约束下生成的新句子,既保留了原始句子的核心语义,又通过因果结构和中间节点的引入实现了标签翻转.在生成过程中设计了严格的约束条件,要求LLM避免使用"导致""引发""造成""使得""由于"等显性因果指示词,使模型生成依赖上下文语义和逻辑推理的隐式因果表达.这一设计显著提升了生成样本的难度,避免模型过度依赖显式因果线索,促使模型学习更深层的因果逻辑关系.

通过上述三个阶段,实现了从语义相关性筛选、因果逻辑建模到反事实文本构造的逐层约束,不仅有效提升增强样本的因果合理性,同时,反事实样本与原始样本形成的"镜像对"还为模型提供了更具挑战性的边界样本,有助于增强模型对因果与非因果事件对的判别能力.

**2.3 LQV模块** 为了确保生成样本的质量与可靠性,设计LQV模块.该模块构建两阶段递进式质量验证与筛选机制,对LLM生成的反事实样本进行多维度验证,有效地过滤了低质量噪声样本,为后续的模型训练提供了高质量数据支撑.

在第一阶段验证中,为了规避事件提及被篡改、删减或替换所引发的样本分布偏移问题,同时确保因果标签严格反转,以提供有效的差异化监督信号,样本需通过两项核心校验.

(1)事件提及一致性验证.要求反事实样本必须完整保留原始样本的目标事件对$(e_1, e_2)$,确保事件表述的完整性与语义无偏性,不得添加或删减核心术语.

(2)因果转换有效性验证.要求反事实样本的因果标签与原始样本严格对立,即当原始标签为$y$时,反事实样本标签为$1-y$,保障了反事实转换的逻辑正确性.

第一阶段校验合格的样本仍可能存在因果表达简单、语义逻辑矛盾等问题,为此,在第二阶段构建细粒度质量评估体系,从因果隐晦性、事件完整性、结构一致性与逻辑有效性四个层面全面衡量样本质量.每个维度均由LLM进行自动化评分.

(1)因果隐晦性.评估生成文本是否避免使用"导致""因为"等显性因果连接词,确保因果关系需依赖深层语义推理获取,贴合多数真实文本的因果表达方式.

(2)事件完整性.进一步验证事件提及$e_1$和$e_2$的表述是否被完整保留,没有发生核心术语的添加或删除.

(3)结构一致性.评估生成文本是否至少保留了原始句子75%的句法结构.

(4)逻辑有效性.校验生成文本是否满足语义通顺、上下文连贯、因果场景合理的基本要求,无逻辑矛盾.

针对上述四个维度,分别建立评分为$score_i$ $(i=1,2,3,4)$,通过加权求和计算样本总分:

$$Score_{\text{total}} = \sum_{i=1}^{4} w_i \cdot score_i \tag{2}$$

其中,$w_i$为各维度的权重系数,满足$\sum w_i = 1$.仅当样本总分超过预设阈值$\theta$时,才被纳入最终的增强数据集.这种严格的质量控制机制有效避免了低质量样本对模型训练的负面影响,确保了增强数据的可靠性与有效性,为后续的监督对比学习提供了高质量的训练基础.

**2.4 SCCL模块** 为了增强模型对因果关系的判别能力,引入监督对比学习机制.该模块显式优化事件对在表征空间中的分布特性,拉近同类样本间的距离,同时推远异类样本,提升模型对因果边界的识别精度.给定输入句子$S=[w_1, w_2, \cdots, w_n]$及其中的两个事件提及$e_1$和$e_2$,利用预训练语言模型获取整个句子的上下文表征如下:

$$H = PLM(S) \tag{3}$$

其中,$H \in R^{n \times d}$,是句子中所有词的隐藏状态,$d$为隐藏层维度.

为了显式标记目标事件对在原始输入中的位置区间,保障语义表征的精准提取,采用$\langle e_1 \rangle$, $\langle /e_1 \rangle$, $\langle e_2 \rangle$, $\langle /e_2 \rangle$四个特殊标记,标注$e_1$和$e_2$在句子中的起止边界.对标记范围内词元的隐藏状态进行池化操作获得$h_{e_1}$和$h_{e_2}$:

$$h_{e_1} = \text{Pooling}\left(H\left[p_{e_1}:p_{e_1'}\right]\right) \tag{4}$$

$$h_{e_2} = \text{Pooling}\left(H\left[p_{e_2}:p_{e_2'}\right]\right) \tag{5}$$

其中，$[p_{e_1}:p_{e_{1'}}]$和$[p_{e_2}:p_{e_{2'}}]$分别表示事件对在句子中的位置范围，Pooling(·)为平均池化操作，用于将事件提及对应的多个词表征聚合为单一事件表征.

针对现有对比学习方法对比特征设计简单、无法全面捕捉事件对语义关系的缺陷，此模块构建了包含拼接特征、差分向量与逐元素交互特征的多维度对比输入.

$$z_{\text{raw}}=\left[h_{e_1},h_{e_2},h_{e_1}-h_{e_2},h_{e_1}\odot h_{e_2}\right] \tag{6}$$

其中，$\odot$表示逐元素乘操作. 拼接特征保留了每个事件的独立语义特征，差分向量捕捉了事件对之间的语义差异，细粒度交互特征表达了事件间的细粒度语义交互. 为了进一步优化特征空间分布，通过线性投影层将融合后的高维特征映射到适宜对比学习的低维空间：

$$z_{\text{cont}}=W_1 z_{\text{raw}}+b_1 \tag{7}$$

其中，$W_1\in R^{d'\times 4d}$为权重矩阵，$b_1\in R^{d'}$为偏置项，$d'$为投影后的特征维度.

在正负样本构建方面，设当前批次的样本集合为$X=\left\{\left(x_i,y_i\right)\right\}_{i=1}^{N}$，其中，$x_i$为查询样本，$y_i\in\{0,1\}$为因果标签. 基于监督信息，定义正样本集$P_i$和负样本集$N_i$如下：

$$P_i=\left\{x_j\in X\,\middle|\,y_j=y_i,j\neq i\right\} \tag{8}$$

$$N_i=\left\{x_j\in X\,\middle|\,y_j\neq y_i\right\} \tag{9}$$

为了进一步扩大样本多样性，此模块维护一个特征记忆库$M=\left\{\left(z_k,y_k\right)\right\}_{k=1}^{K}$，存储历史批次的投影特征及其标签，解决了小批次训练下对比样本不足的问题，能更有效地优化语义表征空间.

采用监督对比损失进行优化. 对于锚点样本$x_i$，其对比损失定义为：

$$L_{\text{con}}^{i}=-\frac{1}{|P_i|}\sum_{p\in P_i}\lg\frac{\exp\left(\text{sim}\left(z_i,z_p\right)/\tau\right)}{\sum_{a\in A_i}\exp\left(\text{sim}\left(z_i,z_a\right)/\tau\right)} \tag{10}$$

其中，sim(·)为余弦相似度，$A_i=P_i\cup N_i$，为锚点样本$x_i$的所有对比样本集合，$\tau$为温度超参数，$|P_i|$为正样本数量.

通过最小化$L_{\text{con}}$，显式地引导模型学习一个结构化的语义空间. 在此空间中，具有相同因果标签的事件对表征彼此靠近，而标签不同的事件对表征则相互远离，使模型在训练过程中逐渐学会生成判别性更强的事件关系表示.

**2.5　CPM模块**　该模块负责对事件对的因果关系进行最终判别. 需要强调的是，此模块与SCCL模块共享同一个预训练语言模型编码器.

给定句子$S$及其事件对$(e_1,e_2)$，随机选取其中一个事件用[MASK]标记进行掩码处理，得到掩码后的句子$\tilde{S}$，并使用预训练语言模型预测掩码位置的填充词符$v_{\text{pred}}$：

$$v_{\text{pred}}=\text{PLM}\left(p_{\text{mask}},\tilde{S}\right) \tag{11}$$

其中，$v_{\text{pred}}\in R^{1\times d}$为预测填充词符的嵌入，$p_{\text{mask}}$为掩码位置.

使用多头注意力机制捕捉该填充词与整个句子上下文间的深层语义依赖$z_{\text{cls}}\in R^{1\times d}$：

$$z_{\text{cls}}=\text{Attn}_{\text{multi-head}}\left(v_{\text{pred}},H,H\right) \tag{12}$$

进一步，使用前馈神经网络分类：

$$y_z=\text{ReLU}\left(z_{\text{cls}}W_1+b_1\right)W_2+b_2 \tag{13}$$

其中，$\{W_1,b_1,W_2,b_2\}$为可学习参数. 模型通过最小化标准交叉熵损失来优化分类任务：

$$L_{\text{cls}}=-\sum_{(e_1,e_2)\in S}y_{(e_1,e_2)}\lg\left(\text{softmax}\left(y_z W_y+b_y\right)\right) \tag{14}$$

其中，$S$表示数据集中所有句子的集合，$(e_1,e_2)$表示事件对，$y_{(e_1,e_2)}$表示事件对真实的因果标签.

**2.6　联合训练策略**　构建联合训练策略，使SCCL模块与CPM模块形成协同增效. 因此，将DACL模型的损失函数优化为：

$$L_{\text{total}}=L_{\text{cls}}+\beta\cdot L_{\text{con}} \tag{15}$$

其中，$\beta$为权重系数，用以平衡对比损失与分类损失的重要性.

$L_{\text{con}}$通过优化编码器，为模型引入一项额外的归纳偏置，使模型在训练过程中持续关注事件对之间的因果语义差别，为CPM提供判别性更强的特征，同时优化$L_{\text{cls}}$使模型进行更精确的因果分类. 这种机制使模型专注于学习因果关系的本质语义特征，提升模型的泛化能力和对边界的判别精度.

## 3　实验

**3.1　数据集介绍**　为了全面评估提出方法的有效性，在ECI领域两个广泛使用的基准数据集上

进行了实验,即 EventStoryLine v0.9 (ESC)[43]和 Causal-TimeBank (CTB)[44].

ESC 数据集包含了 22 个主题的 258 个文档,共标注 5334 个事件提及,从中提取了 7805 个句子内事件对,其中,1770 个被标注为存在因果关系,占总样本的 22.67%.

CTB 数据集包含 183 个新闻文档,共标注 6811 个事件提及和 9721 个句子内事件对,其中,仅有 298 个因果对,共占数据集的 3.06%,存在显著的类别不平衡问题.

本文采用和以往研究相同的方案,即在 ESC 数据集上采用 5 折交叉验证,在 CTB 数据集上采用 10 折交叉验证,并将精确率(Precision,$P$)、召回率(Recall,$R$)和 $F1$ 分数($F1$)作为主要评价指标,最终结果采用对所有折的结果取平均值.

**3.2 实验参数设置** 为了确保实验结果的可靠性和可复现性,系统阐述模型的核心实现细节与超参数配置. CDA 及 LQV 中均采用 Llama 3.1-8B[45]模型,用于生成高可信度反事实样本及多维度质量评估. 选用 RoBERTa-large 作为基础预训练语言模型,该模型包含 24 层 Transformer 编码器,隐藏层维度设置为 1024,注意力头数为 16. 注意力机制采用标准的多头自注意力架构,设置注意力头数为 16,每个头的维度为 64. 为了增强模型泛化能力,在注意力层和前馈网络层后均引入 dropout 正则化,dropout 率设置为 0.5. 采用 AdamW[46]优化器进行参数更新,初始学习率设置为 1e−5,训练批次大小设置为 32,测试批次大小为 20. 对比学习模块温度系数 $\tau$ 设置为 1.0,对比损失权重 $\beta$ 设置为 0.5.

**3.3 基线模型** 为了评估提出方法的有效性,将本文模型与现有主流方法进行比较. 对于 ESC 和 CTB 数据集,基线模型按照方法分为以下类别.

知识增强类方法:KnowDis[17],LearnDA[18],LKCER[47].

图神经网络类方法:ECLEP[14],LSIN[40],SemSIn[30].

提示学习类方法:DPJL[16],KEPT[38],DICP[48].

对比学习类方法:ICCL[32],GCKAN[36],DECLV[49].

**3.4 实验结果** 实验结果表明,DACL 模型在两个数据集上均取得突出的性能. 在 ESC 和 CTB 数据集上,DACL 的 $F1$ 分数分别达到 74.2%,80.4%,均优于当前主流模型. 尤其在 CTB 数据集上,DACL 展现出更卓越且均衡的识别能力,其 $F1$ 分数(80.4%)相较于次优模型 DECLV(72.4%)有显著提升,具体结果如表 2 所示,表中黑体字表示性能最优,下划线表示性能次优.

**表 2 在 ESC 和 CTB 数据集上的实验结果**

**Table 2 Experimental results on the ESC and CTB datasets**

| 技术范式 | 模型 | EventStoryLine | | | Causal-TimeBank | | |
|---|---|---|---|---|---|---|---|
| | | $P$ | $R$ | $F1$ | $P$ | $R$ | $F1$ |
| 知识增强类方法 | KnowDis (Zuo et al,2020) | 39.7% | 66.5% | 49.7% | 42.3% | 60.5% | 49.8% |
| | LearnDA (Zuo et al,2021b) | 42.2% | 69.8% | 52.6% | 41.9% | 68.0% | 51.9% |
| | LKCER (Su et al,2025) | 67.3% | 72.7% | 69.8% | 61.0% | <u>76.3%</u> | 67.3% |
| 图神经网络类方法 | ECLEP (Pu et al,2023) | 49.3% | 68.1% | 57.1% | 50.6% | 63.4% | 56.3% |
| | LSIN (Cao et al,2021) | 49.7% | 58.1% | 52.5% | 51.5% | 56.2% | 52.9% |
| | SemSIn (Hu et al,2023) | 64.2% | 65.7% | 64.9% | 52.3% | 65.8% | 58.3% |
| 提示学习类方法 | DPJL (Shen et al,2022) | 65.3% | 70.8% | 67.9% | 63.6% | 66.7% | 64.6% |
| | KEPT (Liu et al,2023a) | 50.0% | 68.8% | 57.9% | 48.2% | 60.0% | 53.5% |
| | DICP (Mu et al,2025) | <u>73.5%</u> | 69.4% | 71.3% | 62.4% | 73.5% | 67.3% |
| 对比学习类方法 | ICCL (Liang et al,2024) | 67.5% | 73.3% | 70.4% | 63.7% | 68.8% | 65.4% |
| | GCKAN (Ding et al,2024) | 50.9% | 60.6% | 55.3% | 52.2% | 60.7% | 56.1% |
| | DECLV (Su et al,2025) | 71.8% | **74.3%** | <u>73.8%</u> | <u>67.8%</u> | 71.3% | <u>72.4%</u> |
| 本文方法 | DACL | **74.8%** | <u>73.6%</u> | **74.2%** | **84.0%** | **77.6%** | **80.4%** |

总体上，提出的DACL模型系统性地融合了数据增强与对比学习机制．相较于仅依赖外部知识或提示学习范式，DACL构建“知识库筛选-因果结构约束-LLM生成与校验”三重数据生成机制，生成了高质量、多样化的反事实样本，有效缓解了数据稀缺问题．同时，SCCL模块显式优化了语义表示空间，增强了模型对因果边界的判别能力．这种协同优化机制使DACL能够更精准地捕捉复杂的因果关系，从而在具有挑战性的基准数据集上实现更优的性能．

**3.5　消融实验**　为了系统验证DACL方法中各核心模块的有效性，在CTB数据集上设计消融实验．通过移除关键组件来定量分析各模块对模型性能的贡献，实验结果如表3所示，共设计以下消融模型．

（1）－CDA表示移除概念引导的反事实数据增强模块，仅使用原始训练数据．移除后模型性能显著下降，充分证明反事实数据增强有效扩充了训练样本，对缓解数据稀缺问题至关重要．

（2）－SCCL表示移除监督对比学习模块，仅保留分类损失，即$L_{\text{total}}=L_{\text{cls}}$．结果表明，对比学习通过拉大因果/非因果样本的语义距离，能够帮助模型对样本进行判别性建模，提升分类边界清晰度，提高模型分类性能．移除该模块后，模型生成的事件语义表征判别力下降，影响CPM分类器的决策边界清晰度．

（3）－LQV表示移除LLM引导的质量验证模块．结果表明，若缺乏严格的质量控制，样本中会引入大量弱样本和语义矛盾样本，对模型造成干扰与误导，显著损害模型性能，直接验证了LQV对“弱样本”的过滤作用．

（4）－ConceptNet表示在CDA模块中，不从ConceptNet知识库中筛选中间节点，中间节点完全由LLM自行根据上下文语义生成．模型性能下降表明ConceptNet知识库中的多跳路径节点蕴含事件间的潜在因果关联，其筛选机制能为LLM生成提供更可靠的中间节点指导，减少因自由生成可能引入的语义偏差．

（5）－Causal structure表示从CDA模块中移除链式、对撞、叉状三种因果结构的约束，仅依赖LLM的无约束生成，结果模型的性能下降．表明显式注入因果结构能够确保生成数据的因果逻辑多样性和合理性，能够更好地模拟真实文本中因果关系的复杂表现形式．

**表3　CTB数据集上的消融结果**

**Table 3　Ablation results on the CTB dataset**

| Method | Causal-TimeBank | | | |
|---|---|---|---|---|
| | $P$ | $R$ | $F1$ | Δ |
| DACL | 84.0 | 77.6 | 80.4 | — |
| －CDA | 71.3 | 73.4 | 71.5 | －8.9 |
| －SCCL | 80.0 | 71.5 | 75.1 | －5.3 |
| －LQV | 81.2 | 72.3 | 76.4 | －4.0 |
| －ConceptNet | 83.1 | 75.6 | 78.6 | －1.8 |
| －Causal structure | 83.3 | 76.1 | 79.2 | －1.2 |

注：Δ表示消融后$F1$分数的下降幅度

**3.6　敏感性分析**　为了进一步分析LQV中各评价维度权重设置对模型性能的影响，在CTB数据集上对LQV各权重组合进行敏感性分析，实验结果如表4所示，表中黑体字表示性能最优，下划线表示性能次优．为了确定LQV第二阶段各评价维度的权重，根据不同质量维度在反事实样本筛选中的作用设置重要性系数，并通过归一化得到最终权重．具体地，设第$i$个维度的重要性系数为$\alpha_i$，则对应权重为：

$$w_i=\frac{\alpha_i}{\sum_{j=1}^{4}\alpha_j} \tag{16}$$

结果表明，反事实样本的逻辑连贯性和因果场景合理性对生成数据质量具有更重要的影响，若生成样本虽然语法通顺但因果逻辑不成立，容易形成弱样本并误导模型训练．综合优化策略取得了最优$F1$分数，表明LQV不宜过度依赖单一

**表4　CTB数据集上LQV权重敏感性的分析结果**

**Table 4　Analysis results of LQV weight sensitivity on the CTB dataset**

| $w_i$权重组合 | Causal-TimeBank | | | | $F1$ |
|---|---|---|---|---|---|
| | 因果隐晦性 | 事件完整性 | 结构一致性 | 逻辑有效性 | |
| 因果隐晦性主导 | 0.4 | 0.2 | 0.2 | 0.2 | 79.7 |
| 事件完整性主导 | 0.2 | 0.4 | 0.2 | 0.2 | 79.3 |
| 结构一致性主导 | 0.2 | 0.2 | 0.4 | 0.2 | 79.8 |
| 逻辑有效性主导 | 0.2 | 0.2 | 0.2 | 0.4 | <u>80.1</u> |
| 综合优化策略 | 0.25 | 0.15 | 0.3 | 0.3 | **80.4** |

质量维度,而应在保持事件完整性和因果隐晦性的基础上,重点关注结构一致性与逻辑有效性.综合来看,合理的多维度权重配置能够有效提升生成样本的筛选质量,进一步增强模型对事件因果关系的识别能力.

此外,为了探究对比损失与分类损失的最优权重配比,在CTB数据集上开展了超参数$\beta$的敏感性分析.实验结果如图3所示,$F1$分数呈先升后降的单峰变化趋势.当$\beta$=0.5时,取得80.4%的最优$F1$分数.

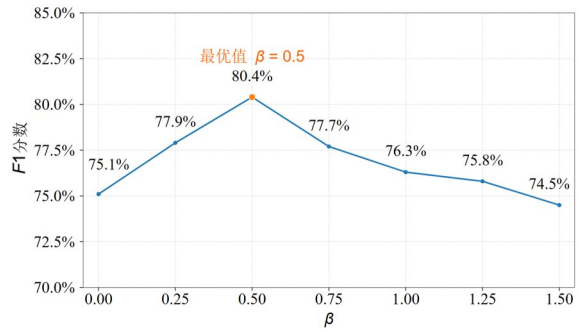



**图3 CTB数据集上DACL模型在$\beta$不同时的性能对比**

**Fig. 3 Performance of the DACL model with varying $\beta$ values on the CTB dataset**

这一结果说明,适当引入对比损失能够使模型更加有效地理解因果关系,提升模型对因果边界的判别能力.但过高的对比损失权重会使模型过度聚焦于事件对之间的语义差异,忽略上下文语境对因果关系的支撑作用,最终可能导致模型性能下降.

**3.7 案例分析** 通过具体案例,全面分析本文反事实数据增强方法的有效性与质量控制机制的过滤能力.

图4a通过查询ConceptNet知识库,获取连接"arrest"与"shot"的潜在路径,从中选取关键中间节点"resist",结合显式注入的链式因果结构,生成的样本逻辑严密,因果关系明确,可作为高质量的正例用于模型训练.不借助外部知识库筛选中间节点也不施加任何因果结构约束,由模型自由生成,生成的因果关系生硬、牵强,近乎噪声.

图4b展示了一个典型的"弱样本"被LQV过滤的过程,该样本的生成完全遵循本文的标准流程,句子整体语义通顺,无明显语法错误,但核心因果逻辑不成立,属于"人为构造的弱样本".若这类样本进入训练集会导致模型过度依赖时序线索判断因果关系,忽略上下文的整体语义逻辑,最终造成模型泛化能力下降.这一案例充分证明LQV对"弱样本"的过滤能力,它能识别出隐蔽性缺陷,是保障生成数据质量的关键防线.

## 4 结论

本文提出一种融合大语言模型反事实数据增强与监督式对比学习的ECI研究方法.该方法构建了"知识库筛选-因果结构约束-LLM生成与校验"三重质量控制流程,成功实现训练数据在数量

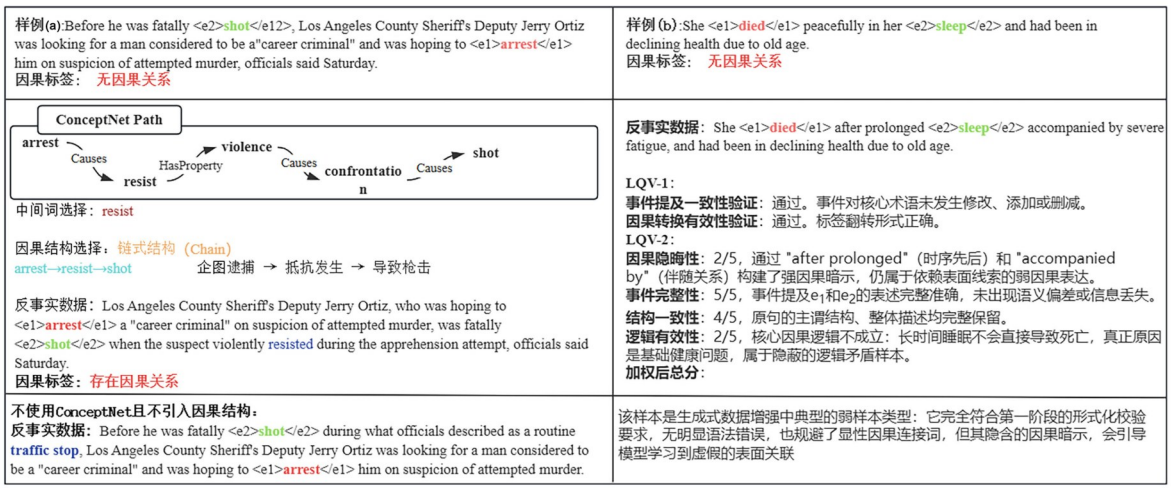



**图4 比较不同方式生成数据的质量**

**Fig. 4 The quality of data generated by different methods**

和质量上的协同提升.此外,本文引入监督对比因果学习模块,通过显式优化语义表征空间,有效提升模型对因果边界的判别能力.在ESC和CTB基准数据集上的实验结果表明,DACL方法超越了现有的主流模型,尤其在数据分布高度不平衡的CTB数据集上展现了卓越的适应性.

## 参考文献


[1] Breja M, Jain S K. Causality for question answering// Proceedings of the 28th International Conference on Computational Linguistics. Barcelona, Spain: International Committee on Computational Linguistics, 2020: 884－893.

[2] Oh J H, Torisawa K, Hashimoto C, et al. A semi-supervised learning approach to why question answering// Proceedings of the 30th AAAI Conference on Artificial Intelligence. Menlo Park, CA, USA: AAAI Press, 2016: 3022－3029.

[3] Zang C Q, Wang H Q, Pei M T, et al. Discovering the real association: multimodal causal reasoning in video question answering// 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE, 2023: 19027－19036.

[4] Ammanabrolu P, Cheung W, Broniec W, et al. Automated storytelling via causal, commonsense plot ordering// The 35th AAAI Conference on Artificial Intelligence. Online: AAAI Press, 2021: 5859－5867, DOI: 10.1609/aaai.v35i7.16733.

[5] Huang J H, Yang C H H, Chen P Y, et al. Causalainer: Causal explainer for automatic video summarization// 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Vancouver, Canada: IEEE, 2023: 2630－2636.

[6] Berant J, Srikumar V, Chen P C, et al. Modeling biological processes for reading comprehension// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PE, USA: Association for Computational Linguistics, 2014: 1499－1510.

[7] Chen Z Y, Chang C H, Chen Y P, et al. UHop: An unrestricted-hop relation extraction framework for knowledge-based question answering// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Volume 1. Long and Short Papers. Stroudsburg, PE, USA: Association for Computational Linguistics, 2019: 345－356.

[8] Al-Khatib K, Hou Y F, Wachsmuth H, et al. End-to-end argumentation knowledge graph construction// Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park, CA, USA: AAAI Press, 2020: 7367－7374.

[9] Radinsky K, Davidovich S, Markovitch S. Learning causality for news events prediction// Proceedings of the 21st International Conference on World Wide Web. New York, NY, USA: Association for Computing Machinery, 2012: 909－918.

[10] Preethi P G, Uma V, kumar A. Temporal sentiment analysis and causal rules extraction from tweets for event prediction. Procedia Computer Science, 2015, 48: 84－89.

[11] Balgi S, Pena J M, Daoud A. Personalized public policy analysis in social sciences using causal-graphical normalizing flows// Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park, CA, USA: AAAI Press, 2022: 11810－11818.

[12] Zhao K, Ji D H, He F Z, et al. Document-level event causality identification via graph inference mechanism. Information Sciences, 2021, 561: 115－129.

[13] Phu M T, Nguyen T H. Graph convolutional networks for event causality identification with rich document-level structures// Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PE, USA: Association for Computational Linguistics, 2021: 3480－3490.

[14] Pu R L, Li Y, Wang S G, et al. Enhancing event causality identification with event causal label and event pair interaction graph// Proceedings of the ACL 2023 Findings. Stroudsburg, PE, USA: Association for Computational Linguistics, 2023: 10314－10322.

[15] Liu P F, Yuan W Z, Fu J L, et al. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Computing Surveys, 2023, 55(9): 1－35.

[16] Shen S R, Zhou H, Wu T T, et al. Event causality identification via derivative prompt joint learning// Proceedings of the 29th International Conference on Computational Linguistics. Gyeongju, Korea (South): International Committee on Computational Linguistics, 2022: 2288－2299.

[17] Zuo X Y, Chen Y B, Liu K, et al. KnowDis: Knowledge enhanced data augmentation for event causality detection via distant supervision// Proceedings of the 28th International Conference on Computational Linguistics. Barcelona, Spain: International Committee on Computational Linguistics, 2020: 1544－1550.

[18] Zuo X Y, Cao P F, Chen Y B, et al. LearnDA: Learnable knowledge-guided data augmentation for event causality identification// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics, the 11th International Joint Conference on Natural Language Processing. Volume 1. Long Papers. Stroudsburg, PE, USA: Association for Computational Linguistics, 2021: 3558－3571.

[19] Hashimoto C, Torisawa K, Kloetzer J, et al. Toward future scenario generation: Extracting event causality exploiting semantic relation, context, and association features// Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistic. Stroudsburg, PE, USA: Association for Computational Linguistics, 2014: 987－997.

[20] Riaz M, Girju R. Recognizing causality in verb-noun pairs via noun and verb semantics// Proceedings of the EACL 2014 Workshop on Computational Approaches to Causality in Language. Stroudsburg, PE, USA: Association for Computational Linguistics, 2014: 48－57.

[21] Riaz M, Girju R. Toward a better understanding of causality between verbal events: Extraction and analysis of the causal power of verb-verb associations// Proceedings of the SIGDIAL 2013 Conference. Stroudsburg, PE, USA: Association for Computational Linguistics, 2013: 21－30.

[22] Do Q X, Chan Y S, Roth D. Minimally supervised event causality identification// Proceedings of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PE, USA: Association for Computational Linguistics, 2011: 294－303.

[23] Hidey C, McKeown K. Identifying causal relations using parallel Wikipedia articles// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Volume 1. Long Papers. Stroudsburg, PE, USA: Association for Computational Linguistics, 2016: 1424－1433.

[24] Beamer B, Girju R. Using a bigram event model to predict causal potential// Proceedings of the 10th International Conference on Computational Linguistics and Intelligent Text Processing. Berlin, Germany: Springer, 2009: 430－441.

[25] Fan C, Liu D X, Qin L B, et al. Towards event-level causal relation identification// Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, NY, USA: Association for Computing Machinery, 2022: 1828－1833.

[26] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Volume 1. Long and Short Papers. Stroudsburg, PE, USA: Association for Computational Linguistics, 2019: 4171－4186.

[27] Liu Z, Lin W, Shi Y, et al. A robustly optimized BERT pre-training approach with post-training// Proceedings of the 20th Chinese National Conference on Computational Linguistics. Huhhot, China: Chinese Information Processing Society of China, 2021: 1218－1227.

[28] Goswami A, Bhat A, Ohana H, et al. Unsupervised relation extraction from language models using constrained cloze completion// Proceedings of the Findings of the Association for Computational Linguistics. Stroudsburg, PE, USA: Association for Computational Linguistics, 2020: 1263－1276.

[29] Man H, Nguyen M, Nguyen T. Event causality identification via generation of important context words// Proceedings of the 11th Joint Conference on Lexical and Computational Semantics. Stroudsburg, PE, USA: Association for Computational Linguistics, 2022: 323－330.

[30] Hu Z L, Li Z X, Jin X L, et al. Semantic structure enhanced event causality identification//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Volume 1. Long Papers. Stroudsburg, PE, USA: Association for Computational Linguistics, 2023: 10901－10913.

[31] Liu J, Chen Y B, Zhao J. Knowledge enhanced event causality identification with mention masking generalizations//Proceedings of the 29th International Joint Conference on Artificial Intelligence. Menlo Park, CA, USA: AAAI Press, 2021: 3608－3614.

[32] Liang C, Wei X, Bang W. In-context contrastive learning for event causality identification// Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PE, USA: Association for Computational Linguistics, 2024: 868－881.

[33] Speer R, Chin J, Havasi C. ConceptNet 5.5: An open multilingual graph of general knowledge// Proceedings of the 31st AAAI Conference on Artificial Intelligence. Menlo Park, CA, USA: AAAI Press, 2017: 4444－4451.

[34] Miller G A. WordNet: A lexical database for English. Communications of the ACM, 1995, 38(11): 39－41.

[35] Chen S Y, Mao K Z. Explicit and implicit knowledge-enhanced model for event causality identification. Expert Systems with Applications, 2024, 238 (Part C): 122039.

[36] Ding L, Chen J T, Du P, et al. Event causality identification via graph contrast-based knowledge augmented networks. Information Sciences, 2024, 656: 119905.

[37] Wu S F, Zhao R H, Zheng Y F, et al. Identify event causality with knowledge and analogy//Proceedings of the 37th AAAI Conference on Artificial Intelligence, the 35th Conference on Innovative Applications of Artificial Intelligence, the 13th Symposium on Educational Advances in Artificial Intelligence. Menlo Park, CA, USA: AAAI Press, 2023: 13745－13753.

[38] Liu J T, Zhang Z Q, Guo Z, et al. KEPT: Knowledge enhanced prompt tuning for event causality identification. Knowledge-Based Systems, 2023, 259: 110064.

[39] Liu Y, Li G B, Lin L. Cross-modal causal relational reasoning for event-level visual question answering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(10): 11624－11641.

[40] Cao P F, Zuo X Y, Chen Y B, et al. Knowledge-enriched event causality identification via latent structure induction networks//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics, the 11th International Joint Conference on Natural Language Processing. Volume 1. Long Papers. Stroudsburg, PE, USA: Association for Computational Linguistics, 2021: 4862－4872.

[41] Zuo X Y, Cao P F, Chen Y B, et al. Improving event causality identification via self-supervised representation learning on external causal statement// Findings of the Association for Computational Linguistics. Stroudsburg, PE, USA: Association for Computational Linguistics, 2021: 2162－2172.

[42] Reimers N, Gurevych I. Sentence-bert: Sentence embeddings using Siamese Bert-networks// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, the 9th International Joint Conference on Natural Language Processing. Stroudsburg, PE, USA: Association for Computational Linguistics, 2019: 3980－3990.

[43] Caselli T, Vossen P. The event storyline corpus: A new benchmark for causal and temporal relation extraction//Proceedings of the Events and Stories in the News Workshop. Stroudsburg, PE, USA: Association for Computational Linguistics, 2017: 77－86.

[44] Mirza P, Tonelli S. An analysis of causality between events and its relation to temporal information// Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. Stroudsburg, PE, USA: Association for Computational Linguistics, 2014: 2097－2106.

[45] Grattafiori A, Dubey A, Jauhri A, et al. The llama 3 herd of models. https://arxiv. org/abs/2407. 21783, 2024－11－23.

[46] Loshchilov I, Hutter F. Decoupled weight decay regularization//International Conference on Learning Representations. New Orleans, LA, USA: ICLR, 2018: 1－8.

[47] Su Y, Zhang H, Zhang G J, et al. Enhancing event causality identification with LLM knowledge and concept-level event relations// Proceedings of the 31st International Conference on Computational Linguistics. Stroudsburg, PE, USA: Association for Computational Linguistics, 2025: 7403－7414.

[48] Mu L, Shen J, Ni L, et al. DICP: Deep in-context prompt for event causality identification// Proceedings of the Findings of the Association for Computational Linguistics. Stroudsburg, PE, USA: Association for Computational Linguistics, 2025: 2589－2599.

[49] Su Y, Zhang H, Fan Y, et al. Dynamic energy-based contrastive learning with multi-stage knowledge verification for event causality identification// Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PE, USA: Association for Computational Linguistics, 2025: 12249－12267.

(责任编辑 杨可盛)