计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
63(10): 2697−2714, 2026

基于多粒度语境感知和迭代图优化的文档级事件因果关系识别

曾泽凡[1,2] 司悦航[1,2] 成 清[1,2,3] 刘 忠[1,2,3]
[1](国防科技大学系统工程学院 长沙 410073)
[2](大数据与决策重点实验室(国防科技大学) 长沙 410073)
[3](湖南先进技术研究院 长沙 410073)
(zzfzad2021@163.com)

Multi-Granularity Context-Aware and Iterative Graph Optimization for Document-Level Event Causality Identification

Zeng Zefan[1,2], Si Yuehang[1,2], Cheng Qing[1,2,3], and Liu Zhong[1,2,3]
[1](*College of Systems Engineering*, *National University of Defense Technology*, *Changsha* 410073)
[2](*Laboratory for Big Data and Decision* (*National University of Defense Technology*), *Changsha* 410073)
[3](*Hunan Advanced Technology Institute*, *Changsha* 410073)

Abstract Event causality identification (ECI) aims to detect causality between events in textual contexts. Compared with sentence-level ECI which focuses on single sentences, document-level ECI (DECI) needs to process more complex contexts such as long texts and cross-sentence events. Current DECI approaches face two key challenges: learning high-quality context-aware event representation to capture event-specific causal semantic and effectively integrating structural and semantic information for causal reasoning. To address these challenges, we propose MAIGO (multi-granularity context-aware and iterative graph optimization), a novel DECI framework that achieves context-aware event representation through multi-granularity attention mechanisms and dynamically optimizes the event causality graph structure via iterative embedding refinement. MAIGO enhances event embeddings through sentence-level, document-level, and graph-level attention, capturing event-event, event-sentence, and sentence-sentence interactions to produce better representation for events. The framework further constructs an event causality graph (ECG) and jointly optimizes both the ECG structure and event embeddings via masked attention in an iterative optimization process. Extensive experiments and analysis on two benchmark datasets and a multilingual dataset demonstrate MAIGO's effectiveness, achieving performance gains of 4.1% (causality existence identification) and 7.3% (causality direction identification) in inter-sentence ECI over state-of-the-art models. Further analysis confirms the advantages of the multi-granularity context-aware enhanced representation strategy and iterative graph optimization mechanism.

Key words event causality; document-level; context-aware; attention mechanism; event causality graph; iterative optimization

摘 要 事件因果关系识别旨在检测文本中事件间的因果关系。与面向单个句子内事件的句子级事件因果关系识别相比,文档级事件因果关系识别需要处理长文本和跨句事件等更加复杂的情形。当前,文档级事件因果关系识别方法面临2个关键挑战:1)学习高质量的上下文感知事件表征以捕捉事件特定的

收稿日期:2025-11-12;**修回日期**:2026-04-29
基金项目:湖南省研究生科研创新项目(CX20250115)
This work was supported by the Hunan Provincial Graduate Research and Innovation Foundation (CX20250115).
通信作者:成清(chengqing@nudt.edu.cn)

因果语义；2）有效整合结构与语义信息进行因果推理。针对这2个局限性，提出基于多粒度语境感知与迭代图优化（multi-granularity context-aware and iterative graph optimization, MAIGO）的文档级事件因果关系识别模型。该模型通过多粒度注意力机制实现语境感知事件表征，并借助拓扑-嵌入的相互促进来动态迭代优化事件因果图结构。利用句子级、文档级和图级注意力增强事件嵌入，捕获事件-事件、事件-句子及句子-句子间的交互作用，从而生成更优质的事件表征。进一步地，构建事件因果图，并通过掩码注意力在迭代优化过程中联合优化事件因果图结构与事件嵌入。在2个基准数据集和1个多语言数据集上的实验分析证明了MAIGO的有效性，在跨句因果关系识别任务中相较最先进模型分别实现了4.1%（因果关系存在识别）和7.3%（因果关系方向识别）的性能提升。深入分析证实了多粒度语境感知增强表示策略与迭代图优化机制的优势。

关键词 事件因果关系；文档级；语境感知；注意力机制；事件因果图；迭代优化

中图法分类号 TP391

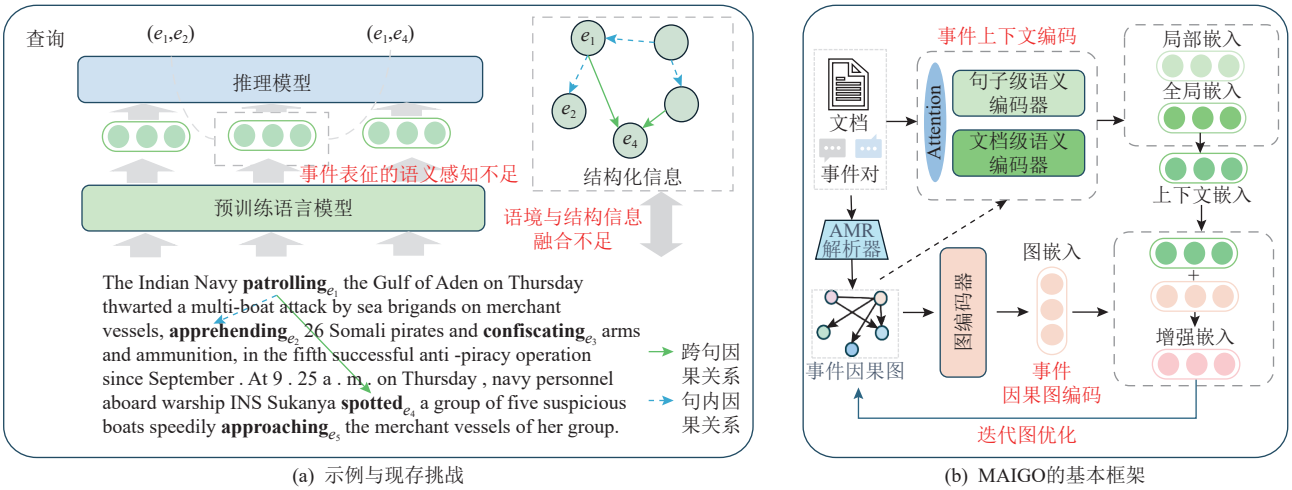
DOI: 10.7544/issn1000-1239.202550803 **CSTR:** 32373.14.issn1000-1239.202550803

事件因果关系识别（event causality identification, ECI）旨在从文本中自动识别事件间的因果关系，从而揭示这些事件背后隐含的逻辑联系。理解事件因果关系对于历史事件解读、未来情景预测以及流程分析支撑有着关键作用[1]。ECI在机器阅读理解[2]、知识图谱构建[3]和智能问答系统[4]等领域具有广泛应用。

在ECI任务中，事件通过触发词（即“事件提及”）进行表征。根据查询事件是否出现在同一句子中，ECI任务通常划分为句子级事件因果关系识别（sentence-level event causality identification, SECI）和文档级事件因果关系识别（document-level event causality identification, DECI）两类，前者主要针对句内因果关系进行分析，而后者则主要面向跨句因果关系。ECI任务的典型案例如图1(a)所示，虚线箭头表示句内因果关系，实线箭头表示跨句因果关系。从方法论的特性来看，SECI主要依赖于短文本范围内的显式因果提示与句法结构特征；而DECI则面临更为复杂的挑战，需在噪声干扰环境下实现上下文信息整合、深层语义理解和长距离依赖中的隐式推理。

早期的因果关系识别研究主要集中于句子级，普遍采用基于特征模式的识别方法。这类方法主要依赖词汇信号[5-6]、时序特征[7-8]及共现关系[9-10]等线索进行因果推断。随着深度学习技术的兴起，更先进的方法应运而生。这些方法通过深度编码文本语义，能够更好地捕捉上下文信息[11]。基于Transformer的预训练语言模型（pre-trained language model, PLM）[12]为事件因果识别方法的发展带来了革命性突破。这些模型经过大规模语料训练，不仅增强了对语义的理解能力，还能生成高质量的事件与上下文表征，从而显著提升了因果关系识别的性能[13-14]。

SECI在2020年之前就已得到广泛研究，而DECI



(a) 示例与现存挑战 (b) MAIGO的基本框架

Fig. 1 Example of the DECI task, existing challenges, and the basic framework of MAIGO

图1 DECI任务的示例、现存挑战与MAIGO的基本框架

直至2022年后才引起学界广泛关注。现有文档级事件因果关系识别方法主要可分为2类：1)基于提示微调的方法[15-17]将ECI构建为问答任务，通过精心设计的提示模板，对预训练语言模型进行微调或优化；2)事件图推理技术[18-21]通过构建以事件为中心的图结构来整合语义与结构信息，并借助图嵌入模型进行因果推理。随着大语言模型(large language model, LLM)的发展，出现了少样本(few-shot)甚至零样本(zero-shot)的ECI研究方法[22-23]。然而，当前基于LLM的ECI研究主要集中于SECI任务，在DECI方面的应用探索相对有限。此外，这些方法仍受因果幻觉(causal hallucination)问题的制约，会生成大量的虚假因果关系，导致其识别的准确率与有监督模型相比存在显著差距。因此，此类方法尚未得以广泛应用。

总而言之，尽管性能正在不断提升，现有的DECI方法仍面临2个关键局限性(参见图1(a))。

1)事件特定上下文捕捉能力不足。现有方法在事件表征学习过程中，通常对文档进行统一编码，忽视了不同事件所依赖的独特上下文环境。这使得模型难以捕捉关键的因果语义与上下文，从而引入了噪声与歧义，最终影响了事件嵌入质量与因果关系表示的有效性。

2)上下文-结构信息融合机制薄弱。现有推理方法未能充分整合上下文语义与图结构关系信息。这种不足不仅阻碍了对因果交互的有效建模，使细微的因果线索难以被察觉，也削弱了模型学习稳健因果语义的能力，从而导致识别准确率下降，并增加了将相关性误判为因果关系的风险。

为应对上述挑战，本文提出一种基于多粒度语境感知与迭代图优化(multi-granularity context-aware and iterative graph optimization, MAIGO)的文档级事件因果关系识别模型。MAIGO通过多粒度注意力机制感知上下文，捕获事件-事件、事件-句子及句子-句子间的语义交互，在句子与文档2个层级生成事件嵌入，从而提升事件表征质量。进一步地，基于抽象语义表示构建事件因果图，采用图注意力网络(graph attention network, GAT)进行结构嵌入，并设计了一种迭代优化算法以联合优化图拓扑结构与事件表征。通过融合上下文与结构信息，上述策略增强了模型对文档级事件因果关系的表征和推理能力。本文的主要贡献有3个方面：

1)提出一种基于注意力机制的多粒度语境感知事件增强编码策略。该方法能够捕捉事件、句子及文档之间的交互作用，生成特定于不同上下文的事件表征，有效减少嵌入表示中的噪声。

2)设计一种迭代图优化方法，通过事件嵌入进行图结构推理，以学习事件因果图(event causality graph, ECG)拓扑，同时利用掩码注意力机制通过ECG反向优化事件嵌入更新。这种双向迭代机制增强了模型对因果语义结构的学习能力。

3)在3个数据集上的广泛实验验证表明，MAIGO在多个文档级ECI评估指标上均优于当前先进基线模型。消融研究与深入分析进一步证实了模型框架设计的合理性。

图1(b)展示了MAIGO的基本框架。MAIGO模型的源代码已在Gitee仓库中公开，开源地址为https://gitee.com/KGAWS/maigo-master。

1 相关工作

本节系统梳理了文本中SECI、DECI及文本事件表征学习3个方向的研究进展。

1.1 SECI

SECI是ECI研究的早期重点，聚焦同一句子内的事件因果关系。现有SECI方法主要可分为4类：1)基于特征模式匹配的方法[6,24]通过词汇与句法模式检测因果关系，虽具有可解释性强、计算效率高的优点，但难以捕捉深层语义信息；2)深度语义编码方法[13,25-26]采用神经网络模型构建蕴含上下文语义的事件嵌入表示，显著提升了因果关系识别的准确性，但如何学习高质量的事件表征是一项关键挑战；3)基于提示微调的方法[27-28]借助预训练语言模型将因果识别任务转化为掩码预测或分类问题，在泛化能力方面表现突出，但在处理复杂长文档时仍面临挑战；4)外部知识增强方法[29-31]通过引入知识图谱等外部资源来加强跨领域因果推理能力，但其性能受限于外部知识与文本内容之间的语义冲突。

在实际应用场景中，事件因果关系往往跨越多个句子甚至整个段落。因此，ECI研究重心正逐渐从SECI向DECI转移，以更好地应对现实场景中文档和篇章级别因果关系识别的迫切需求。

1.2 DECI

DECI的研究最早可追溯至2017年，但其真正取得显著进展则是在2022年之后。当前DECI研究主要围绕基于事件图推理的模型与基于提示的微调方法两大方向展开。基于事件图推理的模型[15-17]通过构建结构化图来表征事件与因果关系：以事件为节点、因果关系为边，借助图嵌入技术实现文档级因果

识别。这类方法能有效捕捉全局上下文信息,并降低跨句噪声干扰。然而,随着文档复杂度的提升,此类模型面临图结构规模膨胀及图噪声累积等挑战。基于提示微调的方法[18-21]沿用了 SECI 任务中类似的技术路线。但在处理跨句因果关系时,由于事件对之间存在大量无关词汇噪声,该方法难以有效捕捉关键语义信息并排除干扰,从而影响对因果语义的准确理解。

总体而言,基于事件图推理的方法能够更有效地学习文档中的因果结构,建模长文本中远距离事件间的因果关系,并有效过滤无关事件带来的噪声干扰,但其在深层语义信息捕捉方面存在不足。基于提示微调的方法虽擅长建模语义关联,却易受长文本跨度与无关事件的影响。本文提出的迭代图优化方法,旨在通过语境感知编码与图结构编码的交互优化实现优势互补,进而达成语义信息与因果结构的深度融合。

1.3 文本事件表征学习

事件表征学习采用机器学习或深度神经网络模型将文本中的事件编码为嵌入向量。主要包括文本嵌入方法和图嵌入方法。文本嵌入方法仅根据事件提及的上下文生成数值化表示,例如 Word2Vec[32]等静态模型生成固定表征;双向 LSTM(Bi-directional LSTM, BiLSTM)[33]等序列模型侧重局部语境但难以捕捉长距离依赖;BERT[12], RoBERTa[34]等预训练语言模型则通过注意力机制捕获全序列依赖关系。图嵌入方法采用图卷积网络、图注意力网络等图编码器,通过聚合邻域节点信息以增强事件表征。

此外,部分研究通过引入额外信息提升事件表征学习质量:语义结构增强编码方法[13,35]利用依存关系、句法路径、语义图等结构化语义数据来丰富事件表征;事件信息增强方法[36-37]则通过捕捉语义、位置、时间及参与者等多维要素,为事件提供更细粒度的上下文描述;标签信息增强方法[26,38]借助因果标签优化事件对嵌入,采用对比学习、图交互、反事实推理等方法提升区分度。

然而,现有增强编码方法大多采用单轮增强范式,即通过一次性注入额外特征完成嵌入优化。若增强的辅助信息质量不可靠,反而可能导致表征质量下降。此外,现有方法未能充分考虑事件在句子层级与文档层级的表征差异,忽略了事件所处上下文的特异性,这些局限都可能在嵌入向量中引入噪声。本文提出的迭代优化机制同步提升外部增强信息(事件因果图)与事件嵌入的质量,并通过多粒度语境感知机制,在表示学习过程中精细考量不同句子对事件嵌入的差异化贡献,从而有效抑制嵌入噪声。

2 MAIGO 模型

本节全面阐述 MAIGO 框架的内容,包括其架构设计、技术细节以及推理流程。首先介绍多粒度语境感知的事件增强编码方法,然后说明事件因果图的初始化方法和图编码策略,最后详细讲解用于联合优化事件嵌入和事件因果图的迭代图优化算法。

2.1 MAIGO 框架概述

MAIGO 框架由 3 个核心模块组成。1)多粒度语境感知编码模块利用预训练语言模型处理事件及其所在句子,通过自注意力机制获取事件的句子级嵌入,并结合带掩码注意力的 BiLSTM 生成事件的文档级嵌入。2)事件因果图编码模块通过抽象语义表示(abstract meaning representation, AMR)解析构建语义图,将其与事件对齐生成 ECG,并利用图注意力网络(GAT)生成图嵌入。这些图嵌入与查询事件对的句子级和文档级嵌入结合后输入因果推理器,用于预测事件因果关系。3)迭代图优化模块以迭代方式优化 ECG 和事件嵌入。它根据事件嵌入和因果推理器的预测结果更新 ECG,然后基于更新后的 ECG 重新生成图嵌入与上下文嵌入,通过掩码注意力和信息聚合实现优化。

图 2 给出了 MAIGO 框架各个模块之间的关系,其中带有小火焰图标的组件表示它们在训练过程中进行了优化,而没有图标的组件则在整个训练过程中保持不变(未进行参数更新)。三角形标记了在迭代优化过程中被更新的嵌入向量。

2.2 多粒度语境感知的事件增强编码

多粒度语境感知编码利用基于注意力机制的文本编码器和序列编码器,同时捕捉局部(句子级)和全局(文档级)不同粒度下的事件语境,以实现事件表征的增强和去噪。

2.2.1 句子编码

首先对输入文本的句子进行编码,记为 $\mathcal{D}=\{X_1,X_2,\cdots,X_{n_s}\}$,其中 n_s 表示句子的数量。每个句子 X_i 由一系列词元(token)构成,其中 $x_{k,i}$ 表示句子 X_i 中的第 k 个词元,且 $x_{0,i}$ 是该句子的[CLS]词元。n_i 表示 X_i 的长度。为了捕捉事件在句子中的位置信息,在每个事件提及的 2 侧插入特殊标记 [E] 和 [/E]。随后,句子通过预训练语言模型进行编码:

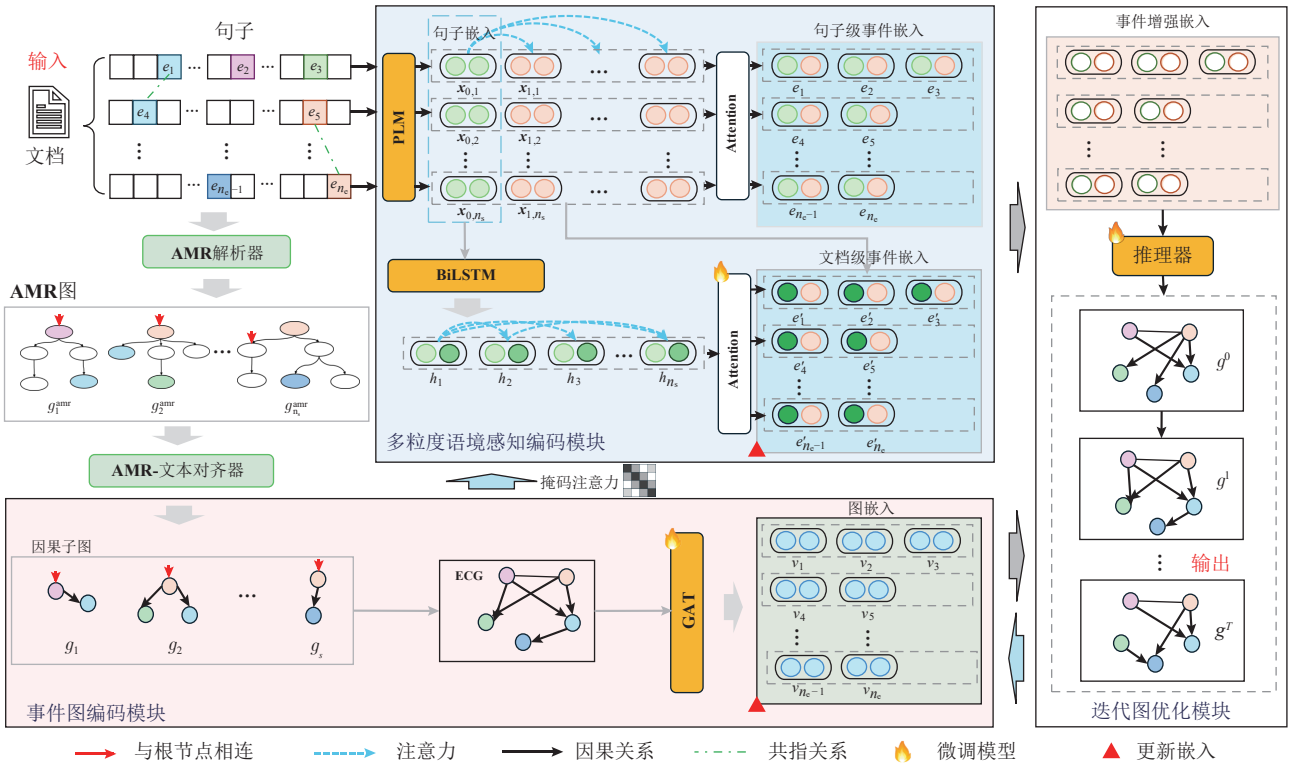


Fig. 2 Framework of MAIGO model

图 2 MAIGO 模型框架

$$\boldsymbol{X}_i = PLM(X_i) = (\boldsymbol{x}_{0,i}, \boldsymbol{x}_{1,i}, \cdots, \boldsymbol{x}_{n_i,i}), \tag{1}$$

其中$\boldsymbol{x}_{0,i} \in \mathbb{R}^d$表示句子的[CLS]词元的嵌入，称为局部句子嵌入，$\boldsymbol{x}_{k,i}(k \in \{1,2,\cdots,n_s\})$表示第$k$个词元的嵌入。$d$表示预训练语言模型的嵌入维度。

2.2.2 句子级事件编码

对于某个事件提及$e_p \in \{e_1, e_2, \cdots, e_{n_e}\}$（$n_e$是句子$X_i$中的事件数，此处省略了上标$i$），其在$X_i$中对应的词元为$[x_{m,i}, \cdots, x_{m+l_p,i}]$（$l_p$为事件$e_p$的跨度）。$e_p$被编码为对应的嵌入：

$$\boldsymbol{e}_p = \frac{1}{l_p+1}\sum_{k=0}^{l_p} \boldsymbol{x}_{m+k,i} \in \mathbb{R}^d \text{。} \tag{2}$$

在句子X_i中，事件e_p的事件特定句子嵌入计算为

$$\bar{\boldsymbol{x}}_{0,i}^{p} = softmax\left(\frac{\boldsymbol{e}_p \boldsymbol{x}_{0,i}^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{x}_{0,i}, \tag{3}$$

其中$\bar{\boldsymbol{x}}_{0,i}^{p}$表示注意力机制的输出，事件嵌入$\boldsymbol{e}_p$被视为查询（query），而句子嵌入$\boldsymbol{x}_{0,i}$同时作为键（key）和值（value）。该编码方式确保即使在同一句子中，每个事件也拥有其独特的局部句子嵌入，从而能够更精确地捕捉与每个事件相关的特定句子级语义。

然后，事件e_p的句子级嵌入最终被定义为

$$\boldsymbol{e}_p^{\text{sent}} = \left[\boldsymbol{e}_p; \bar{\boldsymbol{x}}_{0,i}^{p}\right], \tag{4}$$

其中“;”表示拼接操作，$\boldsymbol{e}_p^{\text{sent}} \in \mathbb{R}^{2d}$。

2.2.3 文档级事件编码

为了将每个句子的局部句子嵌入$\boldsymbol{x}_{0,1}, \boldsymbol{x}_{0,2}, \cdots, \boldsymbol{x}_{0,n_s}$编码为文档级嵌入，采用结合注意力机制的 BiLSTM：

$$\boldsymbol{H} = (\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_{n_s}) = BiLSTM(\boldsymbol{x}_{0,1}, \boldsymbol{x}_{0,2}, \cdots, \boldsymbol{x}_{0,n_s}), \tag{5}$$

其中$\boldsymbol{H} \in \mathbb{R}^{n_s \times d}, \boldsymbol{h}_i \in \mathbb{R}^d$表示句子$\boldsymbol{X}_i$的隐状态（即全局句子嵌入）。对于位于不同句子$\boldsymbol{X}_i$和$\boldsymbol{X}_j$中的事件$e_p$和$e_q$，跨句子的注意力分数$s_{ij}$计算为

$$s_{ij} = \frac{(\boldsymbol{h}_i \boldsymbol{W}_{\mathrm{Q}})(\boldsymbol{h}_j \boldsymbol{W}_{\mathrm{K}})^{\mathrm{T}}}{\sqrt{d}}, \tag{6}$$

其中$\boldsymbol{W}_{\mathrm{Q}}$和$\boldsymbol{W}_{\mathrm{K}} \in \mathbb{R}^{d \times d}$是可学习的参数矩阵，为简洁起见，省略了所有偏置项。这里的注意力机制可以有效地过滤文档中无关内容（其他句子）的噪声，提升句子嵌入的质量。

为了约束句子之间的选择性相互关注，引入跨句子的注意力掩码矩阵$\boldsymbol{M}^p = (M_{ij})^p_{n_s \times n_s}$：

$$M_{ij}^{p} = \begin{cases} 1, & \text{if } |j-i| = 1 \text{ 或 } i = 1 \text{ 或 } G_{pq} = 1, \\ 0, & \text{其他}, \end{cases} \tag{7}$$

其中$\boldsymbol{G} = (G_{pq})_{n_e \times n_e}$表示事件因果图$\mathcal{G}$的邻接矩阵，$G_{pq}$表示矩阵的元素。该矩阵通过 AMR 初始化。由于文档中的首句话通常起到引入或总结的作用，它更可能与所有其他句子相连接。此外，相邻的句子被允许相互关注。同时，句子也可以关注那些包含通

过 ECG 相连的事件所在句子。掩码后的注意力分数计算为

$$s'_{ij}=\begin{cases}s_{ij}, & M^p_{ij}=1,\\ -\infty, & M^p_{ij}=0。\end{cases} \tag{8}$$

在应用掩码后，经过 softmax 运算得到归一化的注意力权重：

$$\alpha^p_{ij}=\frac{\exp(s'_{ij})}{\sum_{k=1}^{n_s}\exp(s'_{ik})}, \tag{9}$$

其中α^p_{ij}表示句子X_i关于句子X_j的注意力权重。以事件e_p为中心的句子X_i的上下文嵌入$\boldsymbol{g}^p_i$计算为

$$\boldsymbol{g}^p_i=\sum_{j=1}^{n_s}\alpha^p_{ij}\boldsymbol{h}_j, \tag{10}$$

$\boldsymbol{g}^p_i\in\mathbb{R}^d$是所有句子嵌入的加权和，但主要由掩码矩阵允许关注的那些句子共同决定。最后，将原始的全局句子嵌入$\boldsymbol{h}_i$与上下文嵌入$\boldsymbol{g}^p_i$拼接：

$$\boldsymbol{h}^p_i=\left[\boldsymbol{h}_i;\boldsymbol{g}^p_i\right], \tag{11}$$

其中$\boldsymbol{h}^p_i\in\mathbb{R}^{2d}$表示文档级的句子嵌入。事件$e_p$的文档级嵌入最终定义为

$$\boldsymbol{e}^{\mathrm{doc}}_p=\left[\boldsymbol{e}_p;\boldsymbol{W}_i\boldsymbol{h}^p_i\right], \tag{12}$$

其中$\boldsymbol{W}_i\in\mathbb{R}^{d\times 2d}$为可学习的映射矩阵，$\boldsymbol{e}^{\mathrm{doc}}_p\in\mathbb{R}^{2d}$。

2.3 ECG 编码

ECG 是有向图，记作$\mathcal{G}=\{\mathcal{V},\mathcal{R}\}$，其中节点集合$\mathcal{V}=\{v_1,v_2,\cdots,v_{n_e}\}$对应于事件集合$\mathcal{E}=\{e_1,e_2,\cdots,e_{n_e}\}$，而边集合$\mathcal{R}=\{r_1,r_2,\cdots\}$则表示节点(事件)之间的因果关系。这些因果关系被编码在邻接矩阵$\boldsymbol{G}$中。ECG 对于识别文档级因果关系至关重要，不仅有利于跨句的事件关系建模，其结构还能帮助发现额外的潜在因果联系。

2.3.1 ECG 初始化

首先构建初始 ECG。尽管采用全连接图、基于初始事件嵌入的初始化或随机初始化等方法在技术上是可行的，但这些方法往往会引入大量噪声、导致高昂的计算成本，或者无法有效捕捉文档的语义结构。为了更有效地利用结构化的语义信息，将 AMR 解析技术与经验规则相结合，构建初始 ECG($\mathcal{G}^0$)。

对于每个 AMR 图$\mathcal{G}^{\mathrm{amr}}_i$，使用 Python 的networkx工具包搜索事件节点之间的连接路径。如果存在路径并满足以下 2 个条件，则在节点之间添加 1 条边(即潜在的因果关系)：

1)2 个节点之间路径上的所有边均属于表 1 所列的关系类型。

Table 1 Potential Edge Types in AMR Graphs for Constructing Event Causality Graphs

表 1 用于构建事件因果图的 AMR 图中潜在边类型

类别	AMR 边类型
手段(Means)	instrument，manner，purpose，…
修饰(Modifiers)	mod，poss
运算符(Operators)	op-X(X为数字)
介词(Prepositions)	prep-X(X为数字)
核心角色(Core Roles)	arg0，arg1，arg2，arg3，arg4
情境角色(Contextual Roles)	condition，cause，comparison

2)2 个节点之间的路径满足以下条件之一。①所有路径均在K_1跳以内；②所有路径均在K_2跳以内，且路径中包含显式的因果线索词(如 cause, result in, lead to 等)或与因果相关的边类型(在本方法中，为了防止假阳性率过高，仅选择 cause, causation, condition, purpose, result, effect 这 6 种)。其中K_1和K_2($K_1\leqslant K_2$)作为超参数使用。K_1保证了长句子或复杂语义结构中的因果关系能够被捕获，而K_2则侧重于识别显式的因果联系。如果 2 个事件节点之间路径的所有边方向一致，则构建 1 条相同方向的边；否则，分别构建$e_i\to e_j$和$e_j\to e_i$两条边。

在识别出所有潜在的事件因果关系后，保留每个 AMR 图的根节点，去除非事件节点，并忽略边的类型，将$\mathcal{G}^{\mathrm{amr}}_i$简化为有向图$\mathcal{G}_i$。由此得到一系列$\mathcal{G}_1,\mathcal{G}_2,\cdots,\mathcal{G}_{n_s}$。然后，以根节点为枢纽将这些子图合并为统一的$\mathcal{G}^0$，在合并后去除根节点，但保留与之相连的事件。尽管边的类型被省略，但在该过程中仍会记录该因果关系是基于条件 1 还是条件 2 建立的。此外，对$\mathcal{G}^0$应用基于字符串匹配和嵌入相似度的指代消解，以识别相同的事件。图 3 展示了基于 AMR 解析器和图合并构建初始 ECG 的示例。

2.3.2 基于 ECG 的事件编码

为了捕捉 ECG 中的事件及其因果关系信息，使用有向图注意力网络对图进行编码，并为事件节点生成其图级别嵌入。这些嵌入与句子级嵌入和文档级嵌入相结合，形成增强的嵌入。ECG 根据增强的嵌入进行迭代更新。

在第t次迭代中，图$\mathcal{G}^t$中节点e_i的嵌入表示为$\boldsymbol{v}_i(t)$，其邻接节点表示为$\mathcal{N}_i(t)=\mathcal{N}^{\mathrm{in}}_i(t)\cup\mathcal{N}^{\mathrm{out}}_i(t)$，其中$\mathcal{N}^{\mathrm{in}}_i(t)$和$\mathcal{N}^{\mathrm{out}}_i(t)$分别表示$e_i$的入邻居和出邻居集合。需要注意的是，$e_i$既属于$\mathcal{N}^{\mathrm{in}}_i(t)$也属于$\mathcal{N}^{\mathrm{out}}_i(t)$。对于任意邻居$e_j\in\mathcal{N}_i(t)$，$e_j$和$e_i$之间的注意力系数由式(13)计算。

$$a_{ij}(t)=LeakyReLU(\boldsymbol{W}_{\mathrm{f}}[\boldsymbol{W}_{\mathrm{a}}\boldsymbol{v}_i(t);\boldsymbol{W}_{\mathrm{a}}\boldsymbol{v}_j(t);c_{ij}(t)]), \tag{13}$$

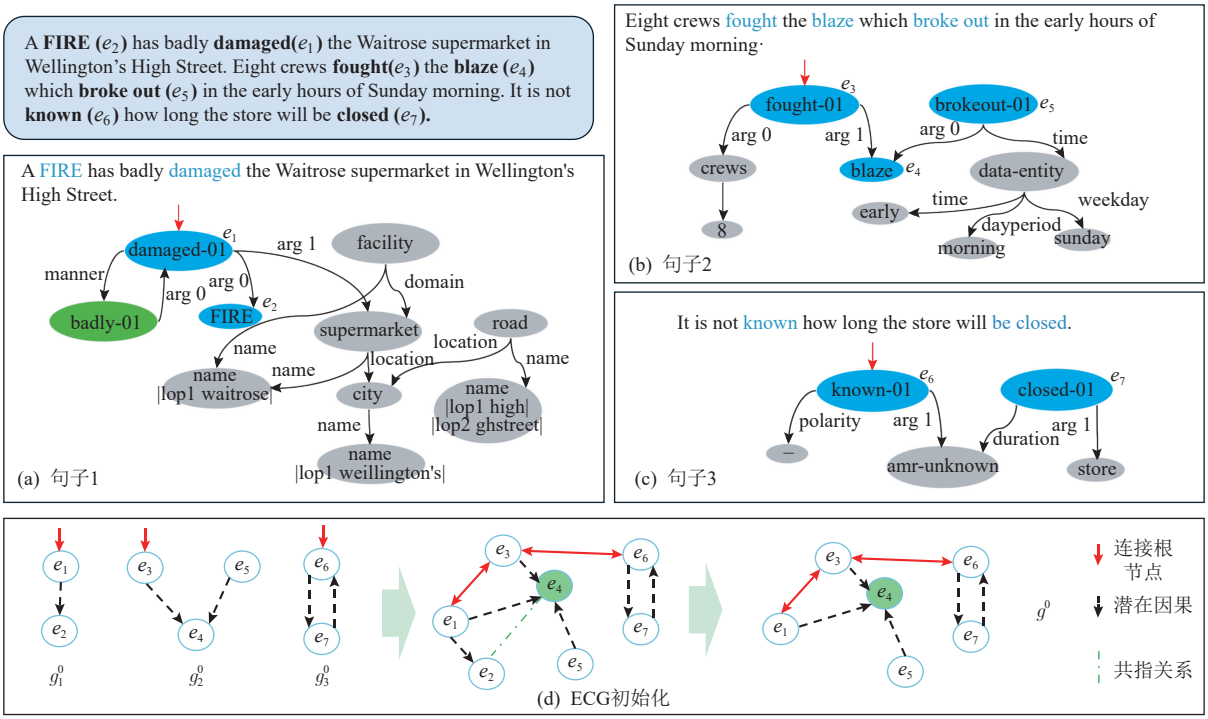


Fig. 3 Example of ECG initialization using AMR parser and graph merging

图 3 使用 AMR 解析器和图合并进行 ECG 初始化的示例

其中$\boldsymbol{W}_{\mathrm{a}} \in \mathbb{R}^{d\times 2d}$，$\boldsymbol{W}_{\mathrm{r}} \in \mathbb{R}^{1\times 3d}$是可学习的参数矩阵，$c_{ij}(t)$表示事件$e_i$和$e_j$间的边（关系）类型，$c_{ij}(t) \in \{\text{intra_in}, \text{intra_out}, \text{inter_in}, \text{inter_out}\}$。在编程实现中，通过设置不同的$c_{ij}$来对图注意力网络进行参数化。

然后，所有注意力系数通过式（14）进行归一化。

$$w_{ij}(t) = softmax(a_{ij}(t)) = \frac{\exp(a_{ij}(t))}{\sum_{k\in\mathcal{N}_i}\exp(a_{ik}(t))}。\tag{14}$$

采用L个注意力头从邻接节点聚合信息。第t次迭代中，节点e_i的输出表示更新为

$$\tilde{\boldsymbol{v}}_i(t) = \sigma\left(\frac{1}{L}\sum_{l=1}^{L}\sum_{j\in\mathcal{N}_i} w_{ij}^{l}(t)\boldsymbol{W}_i^{l}\boldsymbol{v}_j(t)\right),\tag{15}$$

其中$\boldsymbol{W}_i^{l} \in \mathbb{R}^{2d/L\times 2d}$是第$l$个注意力头的参数矩阵。

2.4 事件增强表征

在第t次迭代中，通过将事件的上下文嵌入$\boldsymbol{e}_i^{\bullet}(t)$与基于 ECG 的节点嵌入$\tilde{\boldsymbol{v}}_i(t)$进行融合，得到事件的增强表征$\boldsymbol{e}_i^{+}(t)$，融合时使用可学习的加权系数$\mu$：

$$\boldsymbol{e}_i^{+}(t) = \mu\boldsymbol{e}_i^{\bullet}(t) + (1-\mu)\boldsymbol{W}^{\mathrm{t}}\tilde{\boldsymbol{v}}_i(t),\tag{16}$$

这里，$\bullet$可以是“sent”和“doc”，前者用于句内因果关系，后者用于跨句因果关系。加权因子μ作为平衡因子，控制上下文嵌入与图嵌入在融合表示中的相对权重，在实验中作为超参数进行调整。$\boldsymbol{W}^{\mathrm{t}} \in \mathbb{R}^{2d\times 2d}$为可学习的投影矩阵，用于图嵌入-语境（上下文）嵌入的空间转换。

为了识别事件之间的因果关系，事件对的联合表示定义为

$$\boldsymbol{E}_{ij}(t) = [\boldsymbol{e}_i^{+}(t); \boldsymbol{e}_j^{+}(t)]。\tag{17}$$

2.5 迭代图优化

为了同时学习高质量的事件增强表示与 ECG 的结构，MAIGO 使用了一种迭代式图优化方法。具体而言，模型首先基于当前的 ECG 结构更新事件对的联合表示$\boldsymbol{E}_{ij}$，然后再利用更新后的表示进一步优化 ECG 结构，从而实现事件表征与图结构的相互促进。在此过程中，采用多层感知机（multi-layer perceptron, MLP）作为推理模块，对$\boldsymbol{E}_{ij}(t)$进行建模，用于预测事件对之间的因果关系概率，并据此更新 ECG。

$$\boldsymbol{p}_{ij}(t) = softmax(\boldsymbol{W}_2^{\mathrm{r}} ReLU[\boldsymbol{W}_1^{\mathrm{r}}\boldsymbol{E}_{ij}(t)]),\tag{18}$$

其中$\boldsymbol{W}_1^{\mathrm{r}}$与$\boldsymbol{W}_2^{\mathrm{r}}$为可学习参数矩阵。模型的输出$\boldsymbol{p}_{ij} = (p_{ij}^1, p_{ij}^2, p_{ij}^3)$对应 3 类关系：前向因果关系$e_i \to e_j$、反向因果关系$e_j \to e_i$以及无因果关系。随后，按照算法 1 中的步骤，迭代更新 ECG 结构与事件表征。

算法 1. MAIGO 框架中的迭代图优化。

输入：文本$\mathcal{D} = \{X_1, X_2, \cdots, X_{n_s}\}$，标注的事件提及$e_1, e_2, \cdots$，超参数$K_1, K_2$，阈值$\theta$；

输出：最终的事件因果图$\mathcal{G}$。

① 使用 AMR 对每个句子进行解析，生成语义图

$\mathcal{G}_0^{\text{amr}}, \mathcal{G}_1^{\text{amr}}, \cdots, \mathcal{G}_{n_s}^{\text{amr}}$;

② 将 AMR 图映射到文本词元, 识别事件节点及连接路径;

③ 对每个句子图$\mathcal{G}_i^{\text{amr}}$构建因果子图$\mathcal{G}_i$;

④ 将所有子图$\mathcal{G}_i$合并为初始 ECG($\mathcal{G}^0$)(即邻接矩阵$\boldsymbol{G}^0$);

⑤ 对句子进行编码, 获得局部句子表示$\boldsymbol{X}_i$(式(1))和全局句子表示$\boldsymbol{h}_i$(式(4));

⑥ 初始化迭代计数器: $t=0$, 设$\Phi(t)=\infty$;

⑦ while 未收敛 ($\Phi(t)\geqslant\delta$) do

⑧ for all 事件对(e_i, e_j)($i\neq j$)

⑨ 依据式(17)计算增强事件表征$\boldsymbol{E}_{ij}(t)$;

⑩ 依据式(18)计算因果关系概率$\boldsymbol{p}_{ij}(t)$;

⑪ 依据$\boldsymbol{p}_{ij}(t)$和阈值θ更新邻接矩阵$\boldsymbol{G}^0$;

⑫ end for

⑬ 更新 ECG: $\mathcal{G}^t \to \mathcal{G}^{t+1}, t \leftarrow t+1$;

⑭ end while

⑮ 输出最终的$\mathcal{G}=\mathcal{G}^t$。

本文采用一种自适应的迭代策略来确定终止条件: 对于训练集$\mathcal{D}_{\text{train}}$中的每个样本$\mathcal{D}_k$, 在第$t$次迭代时记录聚焦损失$\mathcal{L}(t)$, 其公式为

$$\mathcal{L}=-\sum_{j=1}^{N}\sum_{i=1}^{C}\beta_i^j(1-p_i^j)^{\gamma}y_i^j\lg(p_i^j), \tag{19}$$

其中C=3 为$\boldsymbol{p}$的输出维度, y_i^j为真实因果标签, p_i^j为预测概率, N为标注事件对数量。γ调节聚焦强度, $\beta_i^j\in[0,1]$为类别平衡因子。然后, 计算损失的变化率:

$$\Delta\mathcal{L}_k(t)=\left|\frac{\mathcal{L}_k(t)-\mathcal{L}_k(t-1)}{\mathcal{L}_k(t)}\right|。 \tag{20}$$

当$\Delta\mathcal{L}_k(t)$连续 3 次迭代低于阈值ϵ时, 迭代停止, 并记录对应的迭代次数T_k。接着, 通过式(21)计算平均图结构差异δ。

$$\delta=\frac{1}{|\mathcal{D}_{\text{train}}|}\sum_{\mathcal{D}_k\in\mathcal{D}_{\text{train}}}\sum_i\sum_j\left|G_{ij}^{T_k}-G_{ij}^{T_k-1}\right|, \tag{21}$$

其中$|\mathcal{D}_{\text{train}}|$表示训练集样本数, 参数$\delta$被用作图差异的阈值。对于测试集$\mathcal{D}_{\text{test}}$中的每个样本, 在每次迭代$t$后, 计算图结构差异$\Phi(t)$:

$$\Phi(t)=\sum_i\sum_j\left|G_{ij}^t-G_{ij}^{t-1}\right|。 \tag{22}$$

当$\Phi(t)\leqslant\delta$时, 迭代停止, 当前的 ECG($\mathcal{G}^t$)被输出作为最终结果。

2.6 计算复杂度分析

区分 3 个模块对 MAIGO 框架的计算复杂度进行分析。设文档包含n_s个句子, 平均句子长度为$\bar{n}_i$, 每句平均事件数为$\bar{n}_e$, 则事件总数满足$n_e\leqslant n_s\cdot\bar{n}_e$。嵌入维度为$d$, ECG 图中节点的平均度数为$\bar{d}$, 迭代次数为$T$。不同模块的计算开销可以归纳如下。

1)多语境感知编码。包括 2 个部分: ①预训练语言模型编码, 复杂度为$O(n_s\bar{n}_iLd)$; ②跨句子的掩码注意力, 复杂度为$O(n_s^3d)$。因此, 总复杂度为$O(n_s\bar{n}_iLd+n_s^3d)$。

2)ECG 编码。在因果图上通过 GAT 进行编码, 复杂度为$O(n_s\bar{n}_eL\bar{d}d)$。

3)迭代图学习。每次迭代包含因果推理和 ECG 更新, 复杂度为$O(n_s^2\bar{n}_e^2d^2+n_s^3d)$, 若迭代$T$次, 则总复杂度为$O(T(n_s^2\bar{n}_e^2d^2+n_s^3d))$。

综合上述各部分, 整体复杂度为$O(n_s^3d+Tn_s^2\bar{n}_e^2d^2)$。在稀疏图结构($\bar{d}\approx5$)、较少的迭代次数($T\in\{3,4,5\}$)以及典型参数范围($n_s\approx\bar{n}_i\leqslant50$, $\bar{n}_e\leqslant5$)下, 计算量是可接受的。整体计算瓶颈主要在迭代图学习部分。但由于 ECG 稀疏且迭代次数有限, 该框架在单块 24 GB 显存的 GPU 上即可高效完成训练。

3 实验与结果分析

为了评估 MAIGO 在 ECI 中的有效性, 并验证其框架设计的合理性, 选择了 3 个 DECI 基准数据集, 并与先进的基线方法进行了全面的比较实验。实验主要关注以下 5 个研究问题(research question, RQ):

问题 1. MAIGO 在 DECI 任务上是否具有性能优势;

问题 2. 多粒度语境感知是否增强了事件表征的质量;

问题 3. 通过 ECG 整合语义表示和结构表示对 ECI 性能的贡献如何;

问题 4. 迭代图优化机制是如何通过协同优化图结构与事件表征来提升模型性能的;

问题 5. MAIGO 的超参数敏感性如何。

3.1 实验设置

3.1.1 实验数据集

为进行全面实验, 选取了 3 个广泛使用的 ECI 基准数据集:

1)ESL(Event StoryLine v0.9)[9] 包含 258 篇新闻文档, 涵盖 22 个主题, 共标注 5 334 个事件提及, 其中包括 1 770 个句内因果事件对和 3 855 个跨句因果事件对。依据标准标注规范, 将标记为“PRECONDITION”的事件对归类为“原因”关系, 将标注为“FALLING-ACTION”的事件对归类为“结果”关系。采用标准数

据划分方案，将最后 2 个主题留作验证集，其余 20 个主题使用 5 折交叉验证进行评估。

2）MAVEN-ERE[39] 是基于多样化维基百科主题构建的大规模事件关系抽取数据集。该数据集采用增强的标注体系，涵盖事件共指、时序、因果及子事件等多种关系类型，共包含 4 480 篇文档、103 193 个事件及 57 992 个因果事件对。遵循标准标注惯例，将“CAUSE”与“PRECONDITION”标签均视为因果关系。沿用原有评估协议，将验证集作为测试集使用。

3）MECI[40] 是多语言 ECI 数据集，源自在 5 种语言（英语、丹麦语、西班牙语、土耳其语和乌尔都语）的维基百科文本。该数据集为多语言识别及跨语言迁移学习提供了宝贵资源，共包含 3 591 篇文档、46 000 个事件和 11 000 个因果事件对。由于目前缺乏适用于丹麦语、乌尔都语和土耳其语的 AMR 解析器，所以仅在英语和西班牙语数据集上进行实验。

需要说明的是，尽管 CTB（Causal-TimeBank）[41] 等基准数据集在 ECI 研究中被广泛采用，但其标注特征使其不适用于评估本文的文档级任务，因为其中仅包含 18 条标注的跨句因果关系，不足以支持对 DECI 性能的可靠评估。

3.1.2 基线模型

选取了 30 个模型作为基线方法，涵盖以下 5 类技术路线：

1）基于特征模式匹配的方法。OP[9]（2017），LIP/LR+[42]（2019）。

2）深度语义编码方法。BERT[12]（2019），Longformer[43]（2020），RoBERTa[34]（2021），SemSIn[13]（2023），ECLEP[26]（2023），CF-ECI[38]（2023），DiffusECI[44]（2024），KADE[45]（2023），SENDIR[46]（2023）。

3）基于 LLM 问答的方法。LLaMA-2-7B，GPT-3，GPT-3.5-turbo，GPT-4，GPT-4o-mini，DeepSeek，DeepSeek-R1，Dr. ECI[47]（2025）。

4）基于提示微调的方法。KEPT[27]（2023），DAPrompt[16]（2023），DPJL[28]（2022），GenECI[48]（2022），HOTECI[17]（2024），LCKER（2025）[49]。

5）事件图推理方法。RichGCN[19]（2021），ERGO[18]（2022），CHEER[50]（2023），PPAT[51]（2023），iLIF[20]（2024）。

6）多语言 ECI 专用方法。Meta-MK[52]（2024），GIMC[53]（2024）。

为确保评估的公平性，实验中的模型性能报告遵循以下原则：若基线模型的原始文献已报告其在对应数据集上的评估结果，则直接采用文献报道数值；若原始文献未报告该数据集结果但其源代码已公开，则通过复现实验获取结果；对于未开源代码或在特定数据集上未报告结果的模型，则不予纳入比较。此外，受调用成本限制，部分 LLM 只应用在特定数据集上。

3.1.3 评估方案

本研究采用精确率（precision，P）、召回率（recall，R）与 $F1$ 分数（$F1$-score）作为模型性能评估指标。从因果关系的“存在性”与“方向性”2 个维度对模型进行评估：在因果关系存在性识别（existence identification，EI）任务中，只要正确识别出存在因果关联的事件对即视为正确预测；而在因果关系方向性识别（direction identification，DI）任务中，需同时准确判断事件对的因果方向方可计为正确。需要说明的是，由于部分模型不具备方向识别能力，因此 2 个任务中所选取的基线模型并不完全相同。

3.1.4 实现细节

实验在 2 块 NVIDIA GeForce RTX 3090Ti 显卡上完成。所有配置均经过精心选择，以确保整体实验，尤其是预训练语言模型或 LLM 加载与推理过程中的最佳性能和兼容性。所有可训练参数均从正态分布中随机初始化。采用英文数据集，并采用 BERT-base-uncased 作为预训练语言模型（嵌入维度 $d=768$），其学习率设置为 2E−5，其他模块（GAT 层、线性投影层和 MLP 推理器）的学习率设置为 1E−4。对于西班牙语则使用多语言预训练模型 mBERT（multilingual-BERT）进行编码。使用带有 LeakyReLU 的双层 MLP 作为推理模块，dropout 设置为 0.4。实验中固定批处理大小（batch size）$\epsilon=0.01$。批处理大小设置为 1。除 2.5 节介绍的迭代终止策略外，还设置了最大迭代次数为 10。借鉴文献 [18, 20] 的方法，在使用 GAT 编码 ECG 时采用 4 个注意力头。对于聚焦损失函数，β 值设置为 0.25, 0.75, 0.75，分别对应无因果关系、正向因果关系和反向因果关系的权重参数，而聚焦参数 γ 设置为 2。对于英文数据集，采用 HanLP 作为 AMR 解析器；对于西班牙语数据集则使用 amr-eager-multilingual 作为 AMR 解析器。

所有可调超参数均根据开发集上的最优 $F1$ 分数进行选择。在超参数优化过程中，对 K_1 在 {3, 5, 7}，K_2 在 {4, 6, 8}、μ 在 {0.4, 0.5, 0.6, 0.7, 0.8}，θ 在 {0.6, 0.7, 0.8} 范围内进行搜索。经过开发集上的交叉验证，最终在 ESL 数据集中设置参数为 $K_1=5$，$K_2=6$，$\mu=0.7$，$\theta=0.7$；在 MAVEN-ERE 和 MECI 数据集中设置为

$K_1=5$, $K_2=6$, $\mu=0.8$, $\theta=0.6$。

对于基于 LLM 的方法，通过调用 API 进行问答实验。在存在性识别任务中，使用提示模板“Is there a causal relationship between e_i and e_j?”；在方向性识别任务中，使用提示模板“Can e_i cause e_j?”。此外，当使用 GPT-4o-mini 和 2 个版本的 DeepSeek 模型时，引入了方向一致性验证机制：仅当 LLM 在双向提示下的因果关系判断结果一致时，才认为因果关系存在。

3.2 整体性能（问题 1）

表 2~5 展示了在 ESL 和 MAVEN-ERE 数据集上、2 种评估模式下 MAIGO 与基线模型的性能对比。结果表明：在跨句因果识别任务中，MAIGO 在 2 个数据集的存在性识别和方向性识别性能上均优于所有基线模型，其 $F1$ 分数在 ESL 数据集较次优基线分别提升 2.3 个百分点和 7.6 个百分点，在 MAVEN-ERE 数据集提升 5.8 个百分点和 7.0 个百分点，这验证了 MAIGO 的优越性。此外，在整体性能方面，MAIGO 同样取得了最先进的结果，其存在性识别和方向性识别的 $F1$ 分数在 ESL 数据集较次优基线分别具有 0.1 个百分点和 7.3 个百分点、在 MAVEN-ERE 数据集提升 7.3 个百分点和 7.6 个百分点的优势。

MECI 数据集上的实验结果如表 6 所示。结果表明，在本研究所覆盖的英语与西班牙语这 2 种语言，以及使用的特定 AMR 解析工具条件下，MAIGO 都展现出明确的性能提升：在 2 种不同语言中均达到最先进水平（高出次优模型 0.1 个百分点和 5.6 个百分点）。这验证了本文提出的语义嵌入与图嵌入融合及其交互增强机制具备一定的跨语言泛化能力。

同时需要指出的是，模型的性能与事件因果图的初始化质量高度相关，而后者直接依赖于 AMR 解析器的性能。具体来说，在 MECI 数据集上，相比使用简单编码器（mBERT）的方法，MAIGO 在西班牙语数据集中的性能提升幅度（13.6 个百分点）显著低于英语（17.4 个百分点）。因此，对于缺乏高质量 AMR 解析器的语言，MAIGO 的效果可能会受到目标语言 AMR 解析工具可用性与质量的制约，存在不确定性。

3.3 消融研究与深入分析（问题 2~4）

在本节中，采用比较分析和消融研究，逐一检查 MAIGO 模型中各组成部分对模型性能的贡献。以下关键组成部分在消融研究中被排除：

1）使用注意力机制增强事件上下文表示，即多粒度语境感知（w/o context_att）；2）采用掩码注意力机制控制跨句子的文档级注意力（w/o mask_att）；3）通过 ECG 引入结构化事件关系信息（w/o ECG）；4）通过迭代优化事件嵌入和 ECG（w/o iteration）。需要说明的是。对于 3），使用事件的上下文表示作为推理模块的输入，用于关系预测，保留事件编码器和推理模

Table 2 Experimental Results of Causal Existence Identification on ESL Dataset

表 2 ESL 数据集上的因果存在性识别实验结果 %

类型	模型	句内因果			跨句因果			总体性能		
		P	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
特征模式匹配	OP	22.5	**98.6**	36.6	8.4	**99.5**	15.6	10.5	**99.2**	19.0
	LIP	38.8	52.4	44.6	35.1	48.2	40.6	36.2	49.5	41.9
	LR+	37.0	45.2	40.7	25.2	48.1	33.1	27.9	47.2	35.1
深度语义编码	BERT	60.4	45.7	52.0	30.6	39.1	34.3	37.2	41.2	39.1
	RoBERTa	62.7	45.4	52.7	32.7	38.3	35.3	39.7	40.6	40.1
	Longformer	47.7	69.3	56.5	26.1	55.6	35.5	31.4	60.0	41.2
	SemSIn	64.2	65.7	64.9	-	-	-	-	-	-
	ECLEP	49.3	68.1	57.1	-	-	-	-	-	-
	CF-ECI	47.1	66.4	55.1	-	-	-	-	-	-
	DiffusECI	65.3	78.3	**71.4**	61.9	59.9	60.9	63.0	64.1	63.5
	KADE	61.5	73.2	66.8	51.2	74.2	60.5	51.9	70.6	59.8
	SENDIR	65.8	66.7	66.2	33.0	90.0	48.3	37.8	82.8	51.9
大语言模型	LLaMA2-7B	12.1	50.7	19.5	6.5	51.5	11.5	7.8	50.8	13.5
	davinci-002	23.2	80.0	36.0	11.4	58.4	19.1	14.1	66.3	23.2
	davinci-003	33.2	74.4	45.9	12.7	54.3	20.6	17.6	59.2	27.1
	GPT-3.5-turbo	27.6	80.2	41.0	15.2	61.8	24.4	20.7	67.5	31.7
	GPT-4	27.2	94.7	42.2	16.9	64.7	26.8	20.9	73.4	32.5
	GPT-4o-mini	31.1	64.7	33.8	4.9	36.7	8.2	-	-	-
	DeepSeek	36.6	62.3	44.1	17.9	37.7	21.9	-	-	-
	DeepSeek-R1	34.7	81.9	43.3	20.3	60.2	26.2	-	-	-
	Dr. ECI	29.0	75.4	41.9	-	-	-	-	-	-
提示微调	DPJL	65.3	70.8	67.9	-	-	-	-	-	-
	GenECI	59.5	57.1	58.8	-	-	-	-	-	-
	KEPT	50.0	68.8	57.9	-	-	-	-	-	-
	DAPrompt	64.5	73.6	68.5	59.9	59.3	59.0	61.4	63.7	62.1
	HOTECI	66.1	72.3	69.1	**81.4**	40.6	55.1	**63.1**	51.2	56.5
	LKCER	67.3	72.7	69.8	20.3	60.2	26.2	-	-	-
事件图推理	CHEER	56.9	69.6	62.6	45.2	52.1	48.4	49.7	53.3	51.4
	ERGO	57.5	72.0	63.9	51.6	43.3	47.1	48.6	53.4	50.9
	iLIF	**76.8**	66.3	71.2	53.5	65.9	59.1	59.2	66.1	62.5
	PPAT	62.1	68.8	65.3	54.0	50.2	52.0	56.8	56.0	56.4
	RichGCN	49.2	63.0	55.2	39.2	45.7	42.2	42.6	51.3	46.6
	MAIGO（本文）	57.7	83.8	64.5	55.5	81.2	**63.2**	53.4	81.8	**63.6**

注：“-”表示未进行实验或原始文献未报告结果。黑体数值表示最优结果，下划线数值表示次优结果。

Table 3 Experimental Results of Causal Existence Identification on MAVEN-ERE Dataset

表 3 MAVEN-ERE数据集上的因果存在性识别实验结果 %

类型	模型	句内因果			跨句因果			总体性能		
		P	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
深度语义编码	BERT	46.0	48.8	47.4	42.1	45.2	43.6	42.4	45.5	43.9
	SENDIR	51.4	53.6	52.5	51.9	52.8	52.4	51.9	52.9	52.4
大语言模型	davinci-002	19.6	**92.9**	32.4	-	-	-	-	-	-
	davinci-003	25.0	75.1	37.5	-	-	-	-	-	-
	GPT-3.5-turbo	19.9	85.8	32.3	-	-	-	-	-	-
	GPT-4	22.5	92.4	36.2	-	-	-	-	-	-
	GPT-4o-mini	-	-	-	25.2	35.7	25.2	-	-	-
	DeepSeek	-	-	-	37.8	53.2	40.6	-	-	-
	DeepSeek-R1	-	-	-	37.5	**81.9**	39.2	-	-	-
	Dr. ECI	22.0	80.0	34.5	-	-	-	-	-	-
提示微调	GenECI	63.8	60.5	60.9	-	-	-	-	-	-
	HOTECI	59.4	68.5	64.1	64.8	45.5	53.5	62.2	53.9	57.3
事件图推理	ERGO	63.1	65.3	64.2	48.7	62.0	54.6	49.6	62.3	55.2
	iLIF	**74.4**	51.5	60.9	**67.1**	49.2	56.8	**67.7**	49.4	57.1
	MAIGO（本文）	62.3	81.1	**70.5**	51.8	78.9	**62.6**	54.4	**79.5**	**64.6**

注：“-”表示未进行实验或原始文献未报告结果。黑体数值表示最优结果，下划线数值表示次优结果。

Table 4 Experimental Results of Causal Direction Identification on ESL Dataset

表 4 ESL 数据集上的因果方向性识别实验结果 %

类型	模型	句内因果			跨句因果			总体性能		
		P	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
深度语义编码	BERT	62.4	32.6	42.8	34.4	30.7	32.4	40.7	31.3	35.4
	RoBERTa	59.7	38.0	46.4	31.3	34.2	32.7	37.3	35.5	36.4
	Longformer	59.0	40.5	48.0	35.2	30.5	32.7	41.6	33.8	37.3
	SENDIR	56.0	52.6	54.2	38.6	39.4	39.0	43.8	43.7	43.7
大语言模型	LLaMA2-7B	17.5	17.0	17.2	6.8	19.2	10.0	8.3	18.5	11.5
	GPT-4o-mini	15.0	38.8	18.8	3.1	26.5	5.3	-	-	-
	DeepSeek	25.9	40.8	30.2	14.3	29.8	16.9	-	-	-
	DeepSeek-R1	20.2	45.5	24.3	16.4	48.7	20.8	-	-	-
事件图推理	ERGO	58.8	47.6	52.6	36.1	41.2	38.5	41.5	43.3	42.4
	iLIF	**66.7**	54.5	**60.0**	41.2	44.6	42.8	**47.9**	47.8	47.8
	MAIGO（本文）	50.2	**72.8**	54.1	**42.8**	**77.1**	**50.4**	45.3	**75.4**	**55.1**

注：“-”表示未进行实验或原始文献未报告结果。黑体数值表示最优结果，下划线数值表示次优结果。

块；对于 4），改为使用事件的初始上下文表示作为推理模块的输入，以调整初始 ECG 的结构，而不进行任何迭代优化。

Table 5 Experimental Results of Causal Direction Identification on MAVEN-ERE Dataset

表 5 MAVEN-ERE数据集上的因果方向性识别实验结果 %

类型	模型	句内因果			跨句因果			总体性能		
		P	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
深度语义编码	BERT	46.0	48.8	47.4	42.1	45.2	43.6	42.4	45.5	43.9
	SENDIR	46.6	44.2	45.4	46.8	43.0	44.8	46.8	43.1	44.9
大语言模型	GPT-4o-mini	-	-	-	20.2	36.4	24.5	-	-	-
	DeepSeek	-	-	-	20.5	53.6	26.7	-	-	-
	DeepSeek-R1	-	-	-	29.0	**82.6**	39.8	-	-	-
事件图推理	ERGO	62.3	63.1	62.7	47.8	59.8	53.1	48.7	60.1	53.8
	iLIF	**73.7**	49.7	59.4	**66.3**	47.5	55.3	**66.9**	47.6	55.6
	MAIGO（本文）	58.2	**75.7**	**63.8**	51.6	78.5	**62.3**	53.2	**77.7**	**63.2**

注：“-”表示未进行实验或原始文献未报告结果。黑体数值表示最优结果，下划线数值表示次优结果。

Table 6 Experimental Results of Causal Existence Identification on MECI Dataset

表 6 MECI 数据集上的因果存在性识别实验结果 %

模型	英语			西班牙语		
	P	*R*	*F*1	*P*	*R*	*F*1
mBERT	38.4	46.0	41.9	43.9	41.5	42.7
KnowMMR	42.1	45.2	43.6	39.2	49.8	43.9
DiffusECI	**70.1**	68.3	69.2	62.9	50.2	55.8
HOTECI	66.6	67.1	66.8	60.7	60.7	60.7
GPT-3.5-turbo	24.6	79.4	37.6	7.3	**74.2**	13.3
GPT-4o-mini	22.3	75.9	33.9	8.6	69.7	12.5
Meta-MK	55.3	71.4	62.3	57.7	61.0	59.3
GIMC	63.4	54.8	58.8	**77.5**	55.7	64.8
RichGCN	53.6	71.8	61.4	54.8	67.2	60.4
MAIGO（本文）	61.1	**81.0**	**69.3**	60	72.7	**66.3**

注：黑体数值表示最优结果，下划线数值表示次优结果。

图 4 展示了在 MAIGO 上进行的消融实验结果：移除任何模块都会导致 MAIGO 的性能下降。在接下来的 3 个小节中，将详细分析这些模块/组成部分在 MAIGO 中的重要性。

3.3.1 多粒度语境感知事件嵌入的贡献（问题 2）

根据图 4 所示的消融实验结果，移除注意力机制与掩码注意力机制（即禁用多粒度语境感知事件嵌入）会导致模型性能下降约 1 个百分点。为深入验证增强事件上下文嵌入对因果关系识别性能的提升作用，从各测试集中选取 2 个样本，通过降维可视化方法对其嵌入表示进行分析。其结果如图 5 所示，

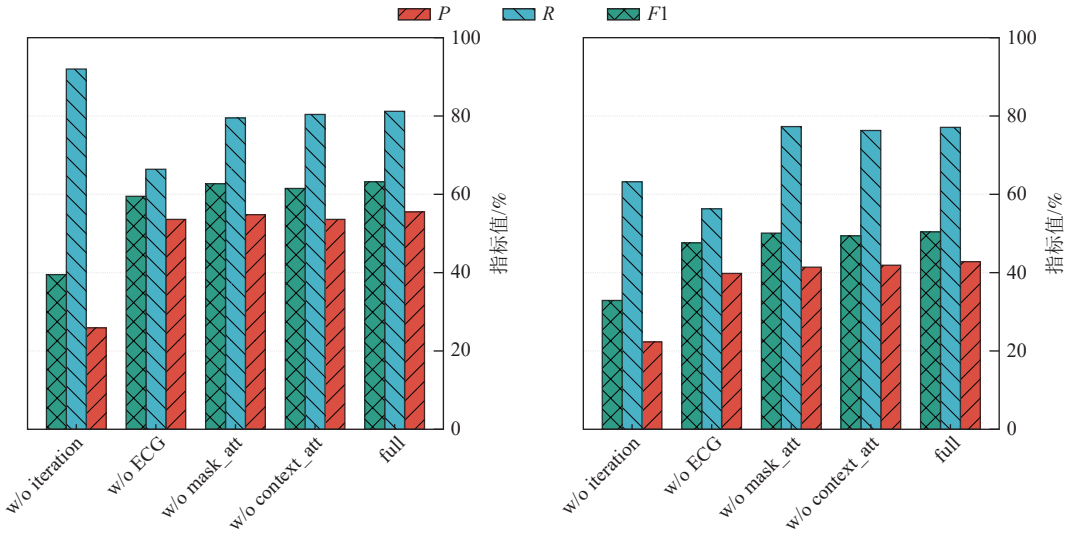


Fig. 4 Ablation experiments results on ESL dataset

图 4 在 ESL 数据集上的消融实验结果

图 5(a)(c)为使用预训练语言模型直接获取事件提及的嵌入, 图 5(b)(d)为采用 MAIGO 增强嵌入后的可视化。虚线箭头表示因果事件对, 实线箭头表示事件对在降维空间中的方向向量。可视化结果表明, 增强后的事件嵌入有效缩小了存在因果关系的事件提及之间的距离, 同时相对扩大了无关事件提及间的距离, 这一特性对事件因果关系识别任务具有显著促进作用。

3.3.2 基于 ECG 的结构信息融合贡献(问题 3)

根据图 4, 直接移除 ECG, 仅依赖事件上下文表示和推理模块会导致整体性能下降近 4 个百分点。这表明基于 AMR 初始化的 ECG 在 MAIGO 中作为关键的先验结构知识, 这也与初始 ECG 的高召回率一致。对于 ECG 的初始化方法, 在 ESL 数据集上进行了对比实验, 使用随机初始化和全连接图初始化作为基于 AMR 初始化的对比, 结果如表 7 所示。该结果表明, 基于 AMR 的初始化在所有性能指标上全面优于其他 2 种初始化方法。

Table 7 Final Performance Comparison of Different Graph Initialization Methods

表 7 不同图初始化方法的最终性能对比 %

初始化方法	存在性			方向性		
	P	*R*	*F*1	*P*	*R*	*F*1
AMR	**55.5**	**81.2**	**63.2**	**42.8**	**77.1**	**50.4**
随机	48.3	59.1	56.4	39.8	55.4	47.1
完全图	50	80.5	58.1	41.2	76.7	48.6

注: 黑体数值表示最优结果。

为了进一步验证使用 ECG 融入事件结构信息的作用, 首先将参数μ设置为 1, 即不将事件的图节点嵌入融合到其增强表示中。然后, 在图 6 中记录了训练过程中的损失以及训练集上评估指标值的变化。可以观察到, 移除基于 ECG 的事件图级嵌入显著降低了训练效果。

3.3.3 迭代图优化的贡献(问题 4)

图 4 显示, 如果省略迭代优化, 由于 ECG 与事件表征之间缺乏交互, 模型无法学习因果推理和判断的模式(边的新增和删除)。虽然召回率略有提升, 但精确度和 *F*1 分数显著下降。

除了消融研究, 实验中还根据推导的Φ值计算了测试集上的平均迭代次数, 以及迭代过程中 MAIGO 性能指标的变化(在存在性识别模式下), 如图 7 所示, 在迭代过程中, *F*1 分数最初上升, 随后稳定, 这表明早期迭代的有效性。随后, ECG 及其嵌入的更新达到相对平衡。尽管可能仍会发生少量的边添加或删除, 推理结果的整体质量趋于稳定。

3.4 参数敏感性分析(问题 5)

为了研究 MAIGO 模型中关键超参数对模型性能的影响, 在 ESL 数据集上对这些超参数进行了全面的比较分析, 其结果图 8 所示。

分析的主要超参数包括: 1)控制 ECG 初始化期间事件连接边构建的参数 K_1 和 K_2; 2)因果推理器中添加和删除边的阈值θ; 3)用于融合嵌入表示的参数μ。在评估某一超参数时, 其他超参数被固定为其最优值。从图 8 中可以看出, 与 K_1 相比, K_2 对模型性能



(a) 简单嵌入方法-案例1　(b) MAIGO的嵌入方法-案例1

(c) 简单嵌入方法-案例2　(d) MAIGO的嵌入方法-案例2

Fig. 5　Multi-granularity context-aware enhanced event embedding methods versus original embedding methods

图 5　多粒度语境感知增强的事件嵌入方法与原始嵌入方法对比

的影响较小，尽管K_2的不同值仍会对性能产生一定影响。相比之下，K_1在决定模型性能方面起着更重要的作用。通常，增大K_1和K_2会提高召回率，但同时会导致精确度下降。因此，要获得最优的$F1$分数，需要在这些指标之间进行权衡。

关于θ和μ，模型性能随着它们的值增加呈现出先提升后下降的趋势。具体来说，过高的θ会阻碍ECG的有效更新，而过低的θ可能会削弱ECG基于AMR初始化的“先验知识”。与θ相比，μ对模型性能的影响较不明显。值得注意的是，当$\mu>0.5$时，模型表现出更好的性能，这表明事件上下文嵌入对于模型的性能更加重要。

3.5　案例分析

为了展示MAIGO的有效性和局限性，从2个实验数据集中各选取了1个测试样本，分析其迭代优化过程，如图9所示。

图9(a)展示了MAIGO的1个成功识别实例。初始ECG包含1个标注的跨句因果关系和所有7个句内因果关系(来自标记的因果关系)，但也包含8个错误的因果关系。通过迭代，MAIGO识别出额外

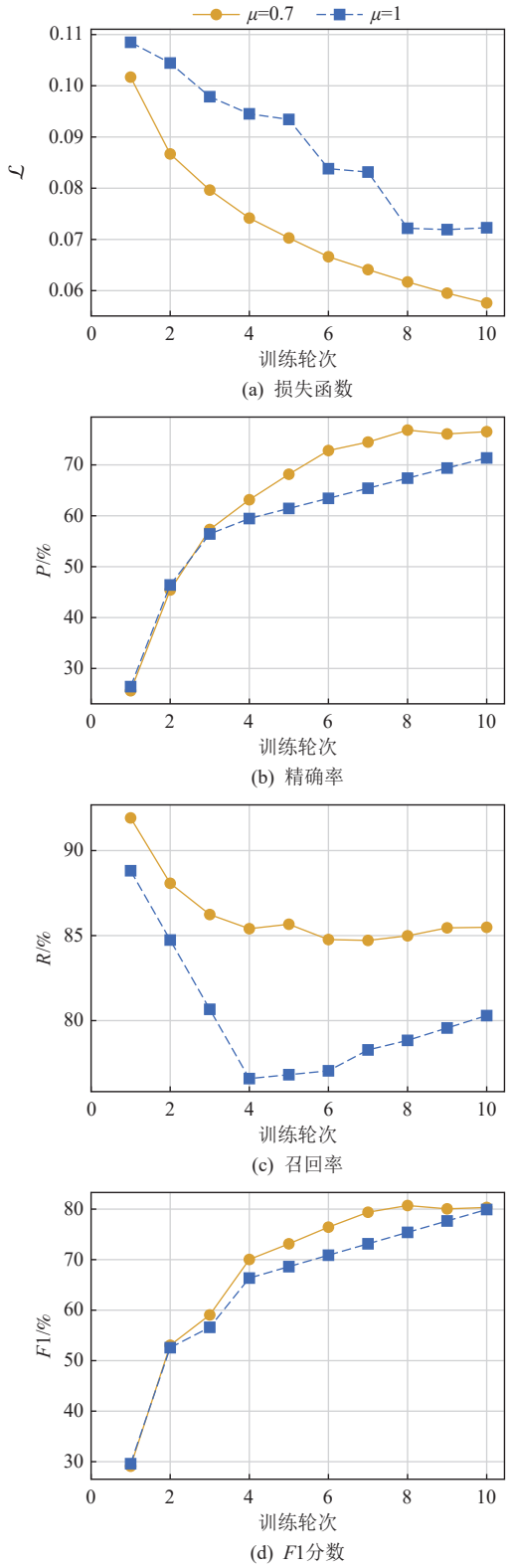


Fig. 6 Performance metrics of MAIGO vary with training epochs on ESL training set

图 6 MAIGO 在 ESL 训练集上的性能指标随训练轮次的变化

的 1 个跨句因果关系，且未引入新的错误因果关系。尽管有 1 个正确的句内因果关系被移除，但错误的

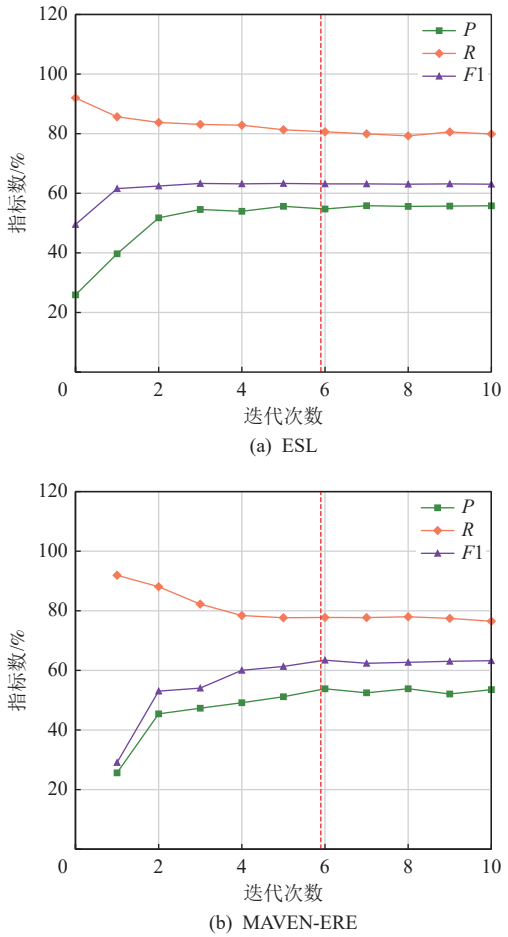


Fig. 7 Test performance metrics of MAIGO vary during the iterative optimization process

图 7 MAIGO 在迭代优化过程中的测试性能指标变化

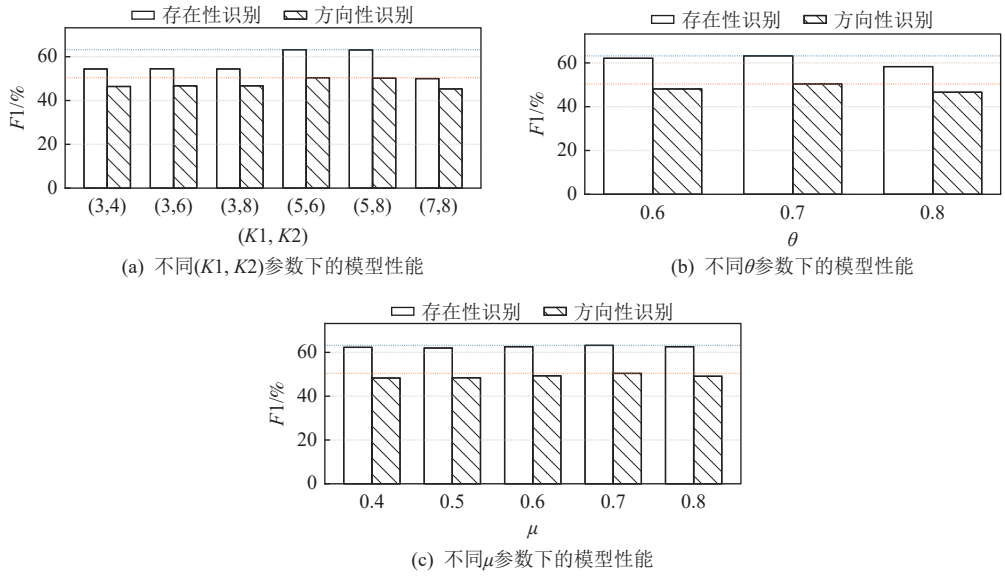
句内因果关系从 8 个减少到 4 个，整体提升了 ECG 的性能。

图 9(b)则展示了 1 个失败的识别过程。在优化过程中，错误的句内因果关系从 3 个增加到 5 个，且丢失了 1 个正确的跨句因果关系。然而，最终 ECG 的性能仅略低于初始 ECG，这凸显了基于 AMR 初始化的识别能力以及优化过程中 ECG 的整体稳定性。

上述分析表明，AMR 初始化有效保留了真实的因果联系，而迭代优化减少了错误的因果关系。

3.6 模型局限性分析

在 ESL 数据集上，部分基线模型的句内因果关系识别性能优于 MAIGO。这是因为 MAIGO 中，模型的注意力资源更多地被分配去关联跨句的上下文信息，这可能使其对句内及其局部且关键的因果信号的专注度相对弱于专门为句内任务优化的模型。此外，在 ECG 初始化期间，AMR 可能引入过多的虚假链接，而这些边在训练过程中未能被有效消除，导致了较低的精确率。



(a) 不同(*K*1, *K*2)参数下的模型性能

(b) 不同θ参数下的模型性能

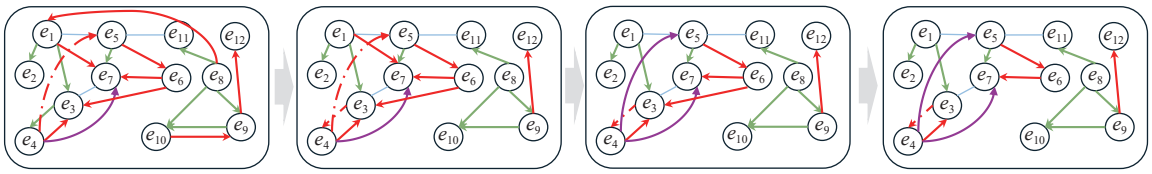
(c) 不同μ参数下的模型性能

Fig. 8 Experimental results of the hyperparameter sensitive analysis on MAIGO

图 8 MAIGO 的超参数敏感性分析实验结果

Test Document ID：7_6ecbplus

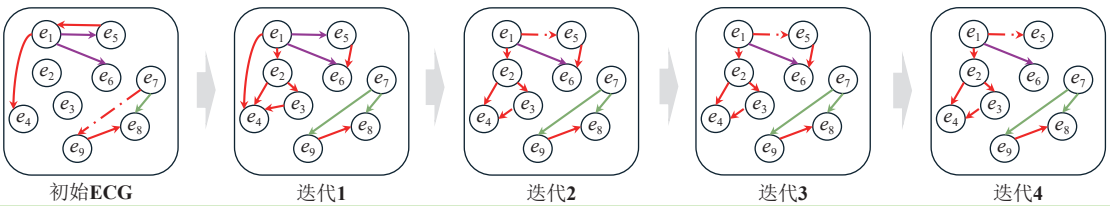
[CLS] Klitschko stops[e1] Thompson in regulation win[e2] July 7 , 2012. **[SEP]** Wladimir Klitschko retained[e3] his IBF , WBO , WBA & IBO heavyweight world titles with a second professional victory[e4] over Tony Thompson on Saturday. **[SEP]** Klitschko , 36 , stopped[e5] the 40 - year - old Thompson with two seconds remaining[e6] of the sixth round in Switzerland to retain[e7] his collection of titles⋯**[SEP]** Klitschko dumped[e8] Thompson onto the canvas with a booming right[e9] end towards the end[e12] of the fifth round of the fight[e10] in Bern , and from then on it was simply a matter of time before referee Sam Williams stopped[e11] matters.**[SEP]**



Test Document ID：Battle_of_Malacca_(1641)

[CLS] The first serious attempt was the siege[e1] of Malacca in 1606 by the third VOC fleet from Holland with eleven ships, led [e2] by Admiral Cornelis Matelief de Jonge that lead to [e3] the naval battle [e4] of Cape Rachado. **[SEP]** Although the Dutch were routed, the Portuguese fleet of Martim Afonso de Castro, the Viceroy of Portuguese India, suffered heavier casualties [e5] and the battle rallied [e6] the forces of the Sultanate of Johor in an alliance with the Dutch. **[SEP]** ⋯ This combined [e7] Dutch-Johor effort effectively destroyed [e8] the last bastion of Portuguese power, removing [e9] their influence in the Malay archipelago. **[SEP]** ⋯

(a) 成功识别实例



(b) 失败识别实例

共指　正确的句内因果关系　正确的句间因果关系　虚假因果关系　未识别的因果关系

Fig. 9 Case study for MAIGO in initial and iterative optimization processes

图 9 MAIGO 初始化和迭代优化过程的案例研究

4 结　论

为学习高质量的事件因果表征，有效融合结构信息与文本语义信息以提升文档级事件因果关系识别的性能，本文提出了基于多粒度语境感知与迭代图优化的模型 MAIGO。MAIGO 采用句子级、文档级与图级注意力机制，从多粒度捕捉事件与其上下

文间的语义交互来提升事件嵌入质量。此外,模型利用 AMR 构建初始 ECG,并通过掩码注意力机制与迭代优化算法,联合优化事件嵌入与 ECG 拓扑结构,从而实现了图结构语义与文本上下文语义的深度融合与相互增强。在多个基准数据集上的对比实验与深入分析验证了 MAIGO 模型的有效性及其整体框架设计的优势。

然而,本研究仍存在一些局限性。首先,模型性能依赖于 AMR 解析器的质量,因此缺乏成熟解析工具的语言或领域可能限制模型的适用性。其次,由于图结构表示与文本序列表示之间存在模态差异,当前融合方法的表达能力仍有提升空间。最后,模型包含多个组件,其联合优化过程较为复杂,对训练稳定性与效率提出了挑战。未来的研究工作可从以下方向展开:一是深入研究跨模态(图与文本)表征的对齐与融合机制;二是探索集成开源 LLM 或大规模预训练图神经网络作为基础编码器,以期获得更强大、更通用的初始事件与结构表征。

作者贡献声明:曾泽凡提出了方法思路和实验方案并撰写论文;司悦航进行了实验验证并修改论文;成清指导方法设计和实验验证;刘忠统筹指导整个研究过程和修改论文。曾泽凡和司悦航为本文的共同第一作者。

参 考 文 献

[1] Cheng Qing, Zeng Zefan, Hu Xingchen, et al. A survey of event causality identification: Taxonomy, challenges, assessment, and prospects[J]. ACM Computing Surveys, 2025, 58(3): 1−36

[2] Berant J, Srikumar V, Chen P C, et al. Modeling biological processes for reading comprehension[C]//Proc of the 19th Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2014: 1499−1510

[3] Ma Yubo, Wang Zehao, Li Mukai, et al. MMEKG: Multi-modal event knowledge graph towards universal representation across modalities[C]//Proc of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. Stroudsburg, PA: ACL, 2022: 231−239

[4] Romeo S, Da San Martino G, Belinkov Y, et al. Language processing and learning models for community question answering in arabic[J]. Information Processing & Management, 2019, 56(2): 274−290

[5] Khoo C, Chan S, Niu Yun, et al. A method for extracting causal knowledge from textual databases[J]. Singapore Journal of Library & Information Management, 1999, 28: 48−63

[6] Khoo C S G, Kornfilt J, Oddy R N, et al. Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing[J]. Literary and Linguistic Computing, 1998, 13(4): 177−186

[7] Mirza P. Extracting temporal and causal relations between events[C]//Proc of the 52nd ACL Student Research Workshop. Stroudsburg, PA: ACL, 2014: 10−17

[8] Mirza P, Tonelli S. An analysis of causality between events and its relation to temporal information[C]//Proc of the 25th Int Conf on Computational Linguistics: Technical Papers. Stroudsburg, PA: ACL, 2014: 2097−2106

[9] Caselli T, Vossen P. The Event StoryLine Corpus: A new benchmark for causal and temporal relation extraction[C]//Proc of the 1st Events and Stories in the News Workshop. Stroudsburg, PA: ACL, 2017: 77−86

[10] Do Q, Chan Y S, Roth D. Minimally supervised event causality identification[C]//Proc of the 16th Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2011: 294−303

[11] Zhao Xiaoyan, Deng Yang, Yang Min, et al. A comprehensive survey on relation extraction: Recent advances and new frontiers[J]. ACM Computing Surveys, 2024, 56(11): 1−36

[12] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//Proc of the 17th Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2019: 4171−4186

[13] Hu Zhilei, Li Zixuan, Jin Xiaolong, et al. Semantic structure enhanced event causality identification[C]//Proc of the 61st Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2023: 10901−10913

[14] Liu Jian, Chen Yubo, Zhao Jun. Knowledge enhanced event causality identification with mention masking generalizations[C]//Proc of the 29th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2021: 3608−3614

[15] Zhang Hang, Ke Wenjun, Zhang Jianwei, et al. Prompt-based event relation identification with constrained prefix attention mechanism[J]. Knowledge-Based Systems, 2023, 281: 111072

[16] Xiang Wei, Zhan Chuanhong, Wang Bang. DAPrompt: Deterministic assumption prompt learning for event causality identification[J]. arXiv preprint, arXiv: 2307.09813, 2023

[17] Man H, Nguyen C V, Ngo N T, et al. Hierarchical selection of important context for generative event causality identification with optimal transports[C]//Proc of the 30th Joint Int Conf on Computational Linguistics, Language Resources and Evaluation. Stroudsburg, PA: ACL, 2024: 8122−8132

[18] Chen Meiqi, Cao Yixin, Deng Kunquan, et al. ERGO: Event relational graph transformer for document-level event causality identification[C]//Proc of the 29th Int Conf on Computational Linguistics. Stroudsburg, PA: ACL, 2022: 2118−2128

[19] Tran P M, Nguyen T H. Graph convolutional networks for event causality identification with rich document-level structures[C]//Proc of the 19th Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2021: 3480−3490

[20] Liu Cheng, Xiang Wei, Wang Bang. Identifying while learning for

document event causality identification[C]//Proc of the 62nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2024: 3815−3827

[21] Cheng Zhangtao, Huang Haoran, Xue He, et al. An event causality identification model based on prompt learning and hypergraph[J]. Computer Science, 2025, 52(9): 303−312 (in Chinese)
(程章桃, 黄浩燃, 薛荷, 等. 基于提示学习与超图的事件因果关系识别模型[J]. 计算机科学, 2025, 52(9): 303−312)

[22] Pan Lei, Yuan Hongxiao, Zhong Zhun, et al. Event causality identification based on graph network construction with large language models[J/OL]. Journal of Southwest Jiaotong University, 2025[2026-02-10]. https://link.cnki.net/urlid/51.1277.u.20250408.1541.004 (in Chinese)
(潘磊, 袁鸿霄, 钟准, 等. 基于大模型构建图网络的事件因果关系识别[J/OL]. 西南交通大学学报, 2025[2026-02-10]. https://link.cnki.net/urlid/51.1277.u.20250408.1541.004)

[23] Gao Jinglong, Ding Xiao, Qin Bing, et al. Is ChatGPT a good causal reasoner? A comprehensive evaluation[C]//Proc of the 28th Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2023: 11111−11126

[24] Ishii H, Ma Qiang, Yoshikawa M. Incremental construction of causal network from news articles[J]. Journal of Information Processing, 2012, 20(1): 207−215

[25] Huang Peixin, Zhao Xiang, Hu Minghao, et al. Distill, fuse, pre-train: Towards effective event causality identification with commonsense-aware pre-trained model[C]//Proc of the 30th Joint Int Conf on Computational Linguistics, Language Resources and Evaluation. Stroudsburg, PA: ACL, 2024: 5029−5040

[26] Pu Ruili, Li Yang, Wang Suge, et al. Enhancing event causality identification with event causal label and event pair interaction graph[C]//Proc of the 61st Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2023: 10314−10322

[27] Liu Jintao, Zhang Zequan, Guo Zhi, et al. KEPT: Knowledge enhanced prompt tuning for event causality identification[J]. Knowledge-Based Systems, 2023, 259: 110064

[28] Shen Shirong, Zhou Heng, Wu Tongtong, et al. Event causality identification via derivative prompt joint learning[C]//Proc of the 29th Int Conf on Computational Linguistics. Stroudsburg, PA: ACL, 2022: 2288−2299

[29] Zuo Xinyu, Cao Pengfei, Chen Yubo, et al. LearnDA: Learnable knowledge-guided data augmentation for event causality identification[C]//Proc of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2021: 3558−3571

[30] Zuo Xinyu, Chen Yubo, Liu Kang, et al. KnowDis: Knowledge enhanced data augmentation for event causality detection via distant supervision[C]//Proc of the 28th Int Conf on Computational Linguistics. Stroudsburg, PA: ACL, 2020: 1544−1550

[31] Zuo Xinyu, Cao Pengfei, Chen Yubo, et al. Improving event causality identification via self-supervised representation learning on external causal statement[C]//Proc of the 59th Annual meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2021: 2162−2172

[32] Mikolov T, Chen Kai, Corrado G, et al. Efficient estimation of word representations in vector space[C]//Proc of the 1st Workshop at Int Conf on Learning Representations. San Diego, CA: OpenReview. net, 2013: 1−10

[33] Zhang Shu, Zheng Dequan, Hu Xinchen, et al. Bidirectional long short-term memory networks for relation classification[C]//Proc of the 29th Pacific Asia Conf on Language, Information and Computation. Stroudsburg, PA: ACL, 2015: 73−78

[34] Zhuang Liu, Wayne L, Ya Shi, et al. A robustly optimized BERT pre-training approach with post-training[C]//Proc of the 20th Chinese National Conf on Computational Linguistics. Stroudsburg, PA: ACL, 2021: 1218−1227

[35] Li Haoran, Gao Qiang, Wu Hongmei, et al. Advancing event causality identification via heuristic semantic dependency inquiry network[C]//Proc of the 21st Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2024: 1467−1478

[36] Kadowaki K, Iida R, Torisawa K, et al. Event causality recognition exploiting multiple annotators' judgments and background knowledge[C]//Proc of the 9th Conf on Empirical Methods in Natural Language Processing and the 9th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2019: 5816−5822

[37] Cui Shiyao, Sheng Jiawei, Cong Xin, et al. Event causality extraction with event argument correlations[C]//Proc of the 29th Int Conf on Computational Linguistics. Stroudsburg, PA: ACL, 2022: 2300−2312

[38] Mu Feiteng, Li Wenjie. Enhancing event causality identification with counterfactual reasoning[C]//Proc of the 61st Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2023: 967−975

[39] Wang Xiaoyu, Chen Yulin, Ding Ning, et al. MAVEN-ERE: A unified large-scale dataset for event coreference, temporal, causal, and subevent relation extraction[C]//Proc of the 19th Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2022: 926−941

[40] Lai V D, Veyseh A P B, Nguyen M V, et al. MECI: A multilingual dataset for event causality identification[C]//Proc of the 29th Int Conf on Computational Linguistics. Stroudsburg, PA: ACL, 2022: 2346−2356

[41] Mirza P, Sprugnoli R, Tonelli S, et al. Annotating causality in the TempEval-3 corpus[C]//Proc of the 52nd Workshop on Computational Approaches to Causality in Language. Stroudsburg, PA: ACL, 2014: 10−19

[42] Gao Lei, Choubey P K, Huang Ruihong. Modeling document-level causal structures for event causal relation identification[C]//Proc of the 17th Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2019: 1808−1817

[43] Beltagy I, Peters M E, Cohan A. Longformer: The long-document transformer[J]. arXiv preprint, arXiv: 2004.05150, 2020

[44] Wu Sifan, Zhao Ruihui, Zheng Yefeng, et al. Identify event causality with knowledge and analogy[C]//Proc of the 37th AAAI Conf on Artificial Intelligence. Palo Alto, CA, AAAI, 2023: 13745−13753

[45] Yuan Changsen, Huang Heyan, Cao Yixin, et al. Discriminative reasoning with sparse event representation for document-level event-event relation extraction[C]//Proc of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA: ACL, 2023: 16222−16234

[46] Man H, Dernoncourt F, Nguyen T H. Mastering context-to-label representation transformation for event causality identification with diffusion models[C]//Proc of the 28th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2024, 38: 18760−18768

[47] Cai R, Yu S, Zhang J, et al. Dr. ECI: Infusing large language models with causal knowledge for decomposed reasoning in event causality identification[C]//Proc of the 31st Int Conf on Computational Linguistics. Stroudsburg, PA: ACL, 2025: 9346−9375

[48] Man H, Nguyen M, Nguyen T. Event causality identification via generation of important context words[C]//Proc of the 11th Joint Conf on Lexical and Computational Semantics. Stroudsburg, PA: ACL, 2022: 323−330

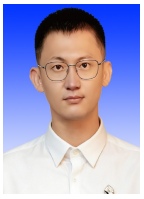
[49] Su Ya, Zhang Hu, Zhang Guangjun, et al. Enhancing event causality identification with LLM knowledge and concept-level event relations[C]//Proc of the 31st Int Conf on Computational Linguistics. Stroudsburg, PA: ACL, 2025: 7403−7414

[50] Chen Meiqi, Cao Yixin, Zhang Yan, et al. CHEER: Centrality-aware high-order event reasoning network for document-level event causality identification[C]//Proc of the 61st Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2023: 10804−10816

[51] Liu Zhenyu, Hu Baotian, Xu Zhenran, et al. PPAT: Progressive graph pairwise attention network for event causality identification[C]//Proc of the 32nd Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2023: 5238−5247

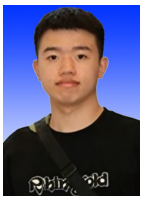
[52] Wang Shen, Cai Fei, Zhang Mengxi. Multilingual event causality identification via meta-learning with knowledge[C]//Proc of the 1st Int Conf on Generative Artificial Intelligence and Information Security. Piscataway, NJ: IEEE, 2024: 18−23

[53] He Zhichao, Cao Pengfei, Jin Zhuoran, et al. Zero-shot cross-lingual document-level event causality identification with heterogeneous graph contrastive transfer learning[C]//Proc of the 30th Joint Int Conf on Computational Linguistics, Language Resources and Evaluation. Stroudsburg, PA: ACL, 2024: 17833−17850



Zeng Zefan, born in 1993. PhD. His main research interests include causal learning and strategic early warning.

曾泽凡, 1993 年生。博士。主要研究方向为因果学习、战略预警。



Si Yuehang, born in 2000. PhD candidate. His main research interests include knowledge graph reasoning and large language models.

司悦航, 2000 年生。博士研究生。主要研究方向为知识图谱推理、大语言模型。



Cheng Qing, born in 1986. PhD, associate professor. His main research interests include big data, machine learning, and knowledge graph.

成　清, 1986 年生。博士，副教授。主要研究方向为大数据、机器学习、知识图谱。



Liu Zhong, born in 1968. PhD, professor. His main research interests include operational research and intelligent systems engineering.

刘　忠, 1968 年生。博士，教授。主要研究方向为运筹规划、智能系统工程。