面向威胁情报的小样本文档级事件抽取方法

王金芳[1]，郭渊博[2,3]，邓淼磊[1]，汤萌萌[4]，Umer Nauman[1]

（1.河南工业大学信息科学与工程学院，河南 郑州 450001；2.海南大学网络空间安全学院，海南 海口 570100；
3.保密通信全国重点实验室，四川 成都 610041；4.郑州航空工业管理学院计算机学院，河南 郑州 450001）

摘 要：文档级事件抽取是构建自动化威胁情报系统的关键技术。针对小样本条件下长文本编码困难、跨句论元分布和角色边界模糊等问题，提出融合事件–共指异构图（ECHG）建模与角色感知对比学习的新框架。通过构建包含句子、提及、实体和事件对象的异构图，统一建模多粒度语义关系，减少对共指解析工具的依赖；引入角色感知对比学习，增强事件内不同角色的区分能力。在CASIE和NetSecDoc数据集上的实验表明，该方法在多种小样本设置下均优于基线模型，F1值最高提升11.73%，具有良好的泛化能力和应用前景。

关键词：威胁情报；文档级事件抽取；小样本学习；异构图建模；对比学习

中图分类号：TP391

文献标志码：A

DOI：10.11959/j.issn.1000–436x.2025249

Few-shot document-level event extraction method for threat intelligence

WANG Jinfang[1], GUO Yuanbo[2,3], DENG Miaolei[1], TANG Mengmeng[4], Umer Nauman[1]

1. School of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China
2. School of Cyberspace Security, Hainan University, Haikou 570100, China
3. National Key Laboratory of Security Communication, Chengdu 610041, China
4. School of Computer Science, Zhengzhou University of Aeronautics, Zhengzhou 450001, China

Abstract: Document-level event extraction is a critical technology for constructing automated threat intelligence systems. To address challenges including the difficulty of encoding long texts under few-shot conditions, cross-sentence argument distribution, and ambiguous role boundaries, a novel framework integrating event-coreference heterogeneous graph modeling and role-aware contrastive learning was proposed. A heterogeneous graph containing sentences, mentions, entities, and event objects was constructed to unify the modeling of multi-granular semantic relationships, thereby reducing reliance on coreference resolution tools. Role-aware contrastive learning was also introduced to enhance the ability to distinguish between different roles within an event. Experiments conducted on the CASIE and NetSecDoc datasets demonstrate that the proposed method outperforms baseline models across various few-shot settings, with an improvement of up to 11.73% in F1 score. The framework shows strong generalization capability and promising application potential.

Keywords: threat intelligence, document-level event extraction, few-shot learning, heterogeneous graph modeling, contrastive learning

收稿日期：2025–08–18；**修回日期：**2025–12–10

通信作者：邓淼磊，dengmiaolei@haut.edu.cn

基金项目：国家自然科学基金资助项目（No.62276091）；保密通信全国重点实验室基金项目（No.6142103042401）

Foundation Items: The National Natural Science Foundation of China (No.62276091), The National Key Laboratory of Security Communication Foundation (No.6142103042401)

0　引言

随着网络攻击手段的演进，从海量的非结构化安全文本（如漏洞报告、高级持续性威胁（APT）分析）中自动抽取结构化事件信息，对于构建自动化威胁情报系统至关重要[1]。然而，网络安全文档具有专业术语密集、事件描述常跨越多个句子的特点，且高质量标注数据因敏感性问题而稀缺[2]。使得小样本条件下的文档级事件抽取（DEE, document-level event extraction）成为推动威胁情报自动化的核心挑战[3]。

事件抽取（EE, event extraction）旨在识别文本中的事件类型、触发词及参数角色，构建结构化表示[4-5]。相较于句子级事件抽取（SEE, sentence-level event extraction），DEE 需要整合整篇文档信息以识别完整事件结构，更贴近实际需求，但同时也面临长文本编码、跨句论元关联与小样本下监督信号不足等多重挑战[6-7]。现有方法大多依赖大量标注数据及外部共指解析工具，在应用于网络安全领域时存在局限[8]。首先，为处理长文本而采用的分段策略会割裂原文的语义连贯性，损害对跨句分布论元的识别能力[9]；其次，在共指解析环节多依赖外部工具，而这些工具在专业性强、句式特殊的网络安全文本上性能显著下降，易引发错误传播[10]；最后，在小样本条件下，模型难以学习到足够区分性特征以准确界定事件内部不同角色的语义边界[11]。为应对上述挑战，本文提出一种融合事件–共指异构图（ECHG）建模与角色感知对比学习的新框架。该框架通过长文档 Transformer（Longformer）编码器捕获文档级上下文，利用局部语义片段增强细粒度理解；构建包含句子、提及、实体与事件对象的异构图，统一建模多粒度语义与共指关系以减少外部工具依赖；引入角色感知对比学习机制，增强小样本下对事件参数的区分能力；最终通过多任务学习协同优化事件检测与参数抽取。

本文主要围绕3个核心问题展开研究。

1) 文档级语义与共指关系的联合建模挑战：现有方法难以在避免外部工具依赖的前提下，有效建模网络安全长文档中复杂的跨句语义与共指关系。

2) 小样本下事件角色表示区分性不足的挑战：在标注数据稀缺的情况下，模型难以学习到具有足够判别性的事件角色表示，导致角色混淆。

3) 领域特定评估基准缺乏的挑战：面向网络安全且专为评估小样本 DEE 方法设计的公开数据集不足，制约了相关研究的公平比较与有效验证。

针对上述挑战，本文的主要贡献如下。

1) 提出事件–共指异构图建模方法：设计了一种端到端的异构图模型，统一编码句子、提及、实体与事件对象，显式建模其间的语义与共指关系，显著降低了对独立共指解析工具的依赖。

2) 设计角色感知对比学习机制：引入了一种融合角色信息的弱监督对比学习策略，通过构造角色感知的对比样本，有效拉大不同角色、拉近相同角色的表示距离，增强了小样本下的角色区分能力。

3) 构建并开源网络安全小样本 DEE 数据集 NetSecDoc：提供了一个涵盖多种新型攻击类型、标注质量高、适用于小样本评估的基准数据集，为后续研究提供了重要的实验平台。

1　相关工作

1.1　文档级事件抽取

文档级事件抽取旨在从整篇文档识别结构化事件。早期方法依赖人工特征与规则模板，在网络安全领域用于提取攻击模式[12]或恶意软件特征[13]，但泛化能力有限。

近年来，深度学习方法成为主流。基于序列标注的方法，如采用双向长短期记忆网络–条件随机场（BiLSTM-CRF），难以建模长距离依赖[14]。基于图神经网络的方法通过构建异构图捕获文档结构，例如结构感知的文档级事件抽取模型 Struct-DEE 融合了句法与多粒度语义[15]，或通过事件关系图建模事件间关联[16]。基于预训练语言模型的方法利用来自 Transformer 的双向编码器表征（BERT）、鲁棒优化的 BERT 预训练方法（RoBERTa）等模型增强表示，但在长文档处理[17]或跨句关联[18]采用方面存在局限。Longformer 进行处理，但未充分考虑跨句论元关联。基于提示学习的方法将事件抽取转化为文本生成[19]，通过设计提示模板在低资源场景下表现良好[20]。

1.2　小样本事件抽取

为应对标注稀缺问题，小样本学习被引入事件抽取。通用领域的小样本方法通常基于度量学习

（如原型网络[21]）或元学习[22]。在文档级任务上，Yang等[18]首次提出小样本文档级事件参数抽取，设计*N*-Way-*D*-Doc（即从*N*个事件类别中，每类选取*D*篇文档构建支持集）采样策略适配文档级模型。Wang等[23]提出多语言提示与层次化原型模块，结合四元对比学习缓解语义歧义与标签噪声。

尽管如此，现有小样本方法在应对网络安全长文档时仍面临两大难题：一是对文档内部复杂的语义结构（如长距离共指）建模不足；二是缺乏对事件内部不同参数角色的细粒度区分机制，导致在样本极少时角色混淆严重。

1.3　上下文建模与对比学习在文档级事件抽取中的应用

有效建模文档级上下文是DEE的核心。图神经网络常被用于构建文档异构图来捕捉语义关系，例如通过引入句子、提及等节点[24]或构建事件关系图[25]。此外，也有研究基于Transformer架构进行增强，例如融入句法特征[26]。

然而，这些方法在处理跨句共指时，普遍依赖斯坦福大学核心自然语言处理工具（Stanford CoreNLP）等外部工具[27]，在网络安全专业文本上性能不稳定，易导致错误传播。尽管近期研究尝试端到端建模共指[28-29]，但对事件语义的融入仍不充分。

对比学习（CL, contrastive learning）通过拉近正例、推远负例来学习判别性表示，被用于提升小样本下的模型鲁棒性，例如用于缓解标签噪声[30]或筛选高价值样本[20]。然而，现有对比学习策略多应用于事件类型或样本整体层面，未能深入事件内部，难以针对攻击者、目标等具体角色学习具有区分性的表示。

综上所述，面向网络安全的小样本文档级事件抽取仍存在不足：缺乏能够同时处理长文档语义、跨句共指并避免依赖外部工具的统一框架，以及针对事件内部角色的细粒度对比增强机制。

2　方法

2.1　整体框架概述

如图1所示，本文方法框架由4个核心模块构成，以端到端方式协同优化。首先，信息抽取和表示模块对输入文档进行编码，并利用原型学习在小样本条件下增强语义表示。其次，事件–共指异构图模块构建并推理一个包含句子、提及、实体及事件对象的多粒度异构图，以统一建模文档级语义与共指关系。然后，角色感知弱监督对比学习模块作用于事件子图上，通过拉近相同角色、推远不同角色的表示，增强模型对事件内部结构的区分能力。最后，事件抽取模块利用前序模块学习到的增强表示，完成事件检测与参数抽取。

2.2　信息抽取、表示与原型增强

给定包含n个句子的文档D，每个句子$s_i=\{w_1,w_2,\cdots,w_t\}$由$t$个词组成。此外，句子中还标注有$k$个实体提及$\{M_i\}$和$u$个代词$\{P_i\}$，且满足$k,u<t$。

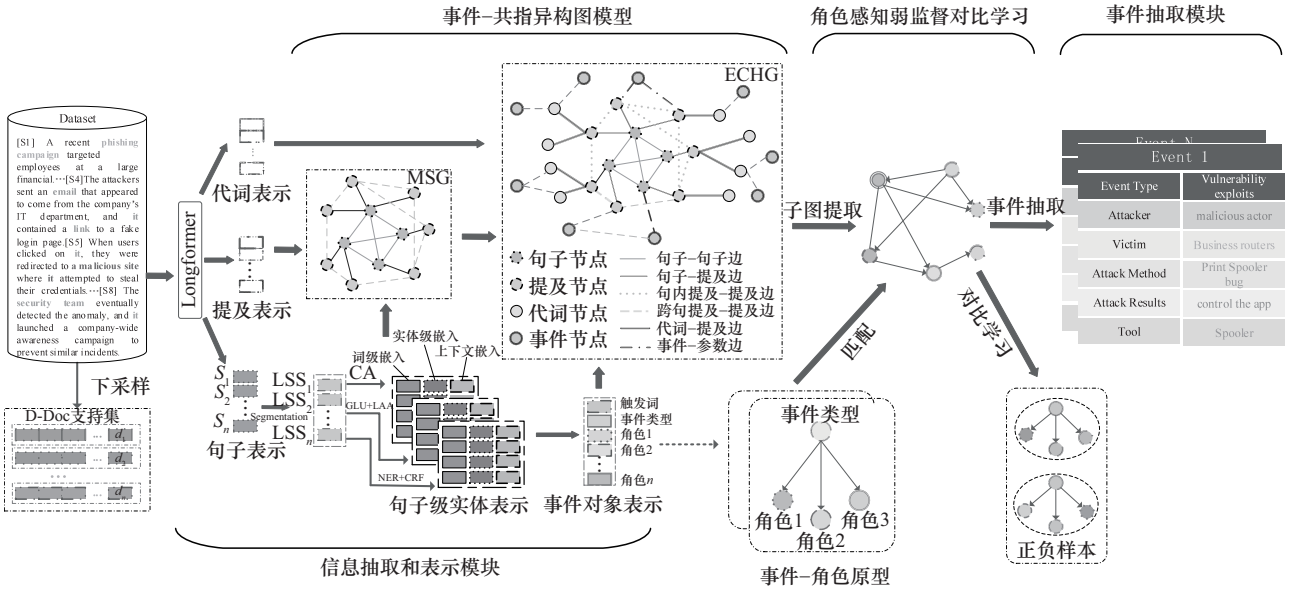

图1　本文方法框架

2.2.1　多粒度语义表示建模

采用 Longformer 模型提取每个词 w_j 的全局上下文表示 $h_j \in \mathbb{R}^d$。

$$\{h_1,h_2,\cdots,h_t\} = \text{Longformer}(\{w_1,w_2,\cdots,w_t\}) \quad (1)$$

为建模局部上下文（如代词指代）并生成局部语义片段（LSS, local semantic span），本文借鉴修辞结构理论[31]中语义单元应保持相对完整与连贯的思想，设计了一个轻量且鲁棒的自动化划分流程。如图 1 中 S5 句子，合理的 LSS 划分有助于区分不同代词的指代对象。具体而言，LSS 的生成并非依赖一个外部的、全功能的修辞结构理论（RST, rhetorical structure theory）解析器，以避免在网络安全复杂句式下引入不可靠的解析错误。本文的实现包含以下 3 个步骤：1)基元切分，利用一个轻量级、预训练的依存句法分析器识别子句边界，并结合标点符号（如逗号、分号）将长句切分为初步的基元片段；2)启发式合并，受 RST 中“核心–卫星”结构与语义连贯性原则启发，本文设计了一组简单的规则判断相邻片段是否应合并。例如，当前一片段以代词结尾，而后一片段以名词性短语开头时，倾向于合并，以将潜在的指代关系约束在同一 LSS 内处理；3)约束与回退，为保持计算效率与语义合理性，对 LSS 的长度设定上下限。若上述规则产生冲突或无法应用，则回退到基于标点的简单切分。

通过上述方法，获得了既能反映局部语义连贯性，又避免依赖重型外部解析器的 LSS 划分结果。在每个 LSS 内使用门控卷积网络（如式(2)）和局部注意力机制（如式(4)）计算得到 token 的局部增强表示 h'_i。

$$\text{GLU}(x) = (\sigma(\boldsymbol{W}_g * x)) \odot (\boldsymbol{W}_v * x) \quad (2)$$

$$\boldsymbol{Q} = \boldsymbol{W}_q^{\mathrm{T}} h_a, \boldsymbol{K} = \boldsymbol{W}_k^{\mathrm{T}} h_b, \boldsymbol{V} = \boldsymbol{W}_v^{\mathrm{T}} h_b \quad (3)$$

$$\text{LocalAtt}n(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}} + \boldsymbol{R}}{\sqrt{d_k}}\right)\boldsymbol{V} \quad (4)$$

其中，$x = \text{LSS}_j$ 是当前 LSS 的输入表示，$*$ 表示一维卷积操作，σ 是 sigmoid 函数，$\boldsymbol{W}_g$ 和 $\boldsymbol{W}_v$ 是可学习参数，$\odot$ 表示按元素乘法，$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 分别是从当前 LSS 提取的查询、键和值矩阵，$\boldsymbol{R}$ 是相对位置编码矩阵，d_k 为缩放因子。

为融入实体类型信息，本文设计了鲁棒的标签编码机制。考虑到小样本设定下任务特定标注的稀缺性，本机制所使用的实体类型（NER）标签并非来自支持集或查询集的标注，而是采用一个在通用领域语料上预训练且在训练阶段冻结的轻量 NER 模块，为输入文档生成初步的实体类型预测。本机制的核心目的是审慎地利用这些可能包含噪声的预测标签信息。该机制通过动态注意力计算 token h'_i 与标签嵌入 l_k 之间的关联权重。

$$\alpha_{i,k} = \text{LeakyReLU}([h'_i \| l_k] W^a) \quad (5)$$

$$\hat{\alpha}_{i,k} = \frac{\exp(\alpha_{i,k})}{\sum_{k'=1}^{L} \exp(\alpha_{i,k'})} \quad (6)$$

其中，L 表示标签的数量，W^a 是可训练参数。$\|$ 表示向量拼接操作。$\hat{\alpha}_{i,k}$ 为当前 token 生成的、融合了上下文感知标签语义的增强向量表示。同时，构建一个加权邻接矩阵以建模标签间的局部关联。

$$\boldsymbol{A}_{i,j} = \sum_{k=1}^{L} \hat{\alpha}_{i,k} \cdot \hat{\alpha}_{j,k} \quad (7)$$

最终，通过聚合邻居标签信息并与原始表示融合，得到每个 token 融合了筛选后标签语义的最终增强表示 $\hat{h}_i$。

$$\hat{h}_i = [h'_i \| \sum_{k=1}^{L} \hat{\alpha}_{i,k} l_k] \quad (8)$$

此设计使得模型能够在不增加小样本标注负担的前提下，利用实体类型的先验语义知识来增强上下文表示，体现了方法的实用性。

2.2.2　实体与事件实例初始化

在获得增强的 token 表示 $\hat{h}_i$ 后，本模块并行完成实体识别与事件实例初始化，为后续的图构建和事件抽取提供基础。

1) 实体抽取。实体抽取任务本身使用条件随机场（CRF），本文简单地获取句子表示 $\hat{h}_s$、提及表示 $\hat{h}_m$ 和代词表示 $\hat{h}_p$，并定义实体抽取的损失函数为

$$\hat{h}_s = g_{\max}(\{\hat{h}_i\}_{i=1}^{t})$$

$$\hat{h}_m = g_{\text{avg}}(\{\hat{h}_i\}_{i=1}^{k})$$

$$\hat{h}_p = g_{\text{avg}}(\{\hat{h}_i\}_{i=1}^{u})$$

$$L_{\text{NER}} = -\sum_{x_i \in D} \ln p_\theta(y_i|\hat{h}_i) \quad (9)$$

其中，$x_i \in D$ 表示数据集中的样本，y_i 是对应的真实实体标签，$p_\theta(\cdot|\hat{h}_i)$ 是模型基于增强表示 $\hat{h}_i$ 的预测概率。请注意，此处的 CRF 学习与 2.2.1 节中利用外部预训练 NER 标签增强表示是 2 个独立且互补的过程：前者使用少量标注进行任务自适应，后者利用外部知识进行通用语义增强。

2) 事件实例初始化与原型构建。为在小样本

条件下引导模型理解事件结构，同时进行事件实例初始化和原型构建。

事件查询与表示初始化：基于预定义的事件模式，为每个事件类型e_t及其角色r_k初始化一个可学习的事件查询向量$\boldsymbol{Q}_{t,k}$，利用该查询对句子表示进行聚合。首先，通过最大池化获得第s个句子的表示$s_s = \text{maxpooling}(\hat{h}_1,\hat{h}_2,\cdots,\hat{h}_t)$。然后，生成文档级的事件类型感知表示$S_t^e$和角色感知表示$S_{t,k}^r$作为事件节点的初始化表示。

$$S_t^e = \text{maxpooling}(\{\text{Transformer}(s_s,\boldsymbol{Q}_{t,:})\}_{s=1}^{N}) \quad (10)$$

$$S_{t,k}^e = \text{maxpooling}(\{\text{Transformer}(s_s,\boldsymbol{Q}_{t,k})\}_{s=1}^{N}) \quad (11)$$

触发词检测：使用一个分类层在token表示$\hat{h}_i$上进行触发词检测。

$$p(y_i = \text{Trigger}|\hat{h}_i) = \text{softmax}(\boldsymbol{W}_t\hat{h}_i + b_t) \quad (12)$$

其中，$\boldsymbol{W}_t$为分类层的权重矩阵，b_t为对应的偏置向量。检测到的触发词表示h_i^{trigger}将与对应的事件类型表示S_t^e融合，形成初步的事件对象表示。

原型向量构建：这是小样本学习的核心。利用支持集样本为每个事件类型和语义角色计算原型向量，作为该类别在表示空间中的“锚点”。

$$P_c = \frac{1}{|S_c|}\sum_{i\in S_c} h_i^{\text{final}} \quad (13)$$

$$P_r = \frac{1}{|S_r|} h_j^{\text{final}} \quad (14)$$

其中，S_c和S_r分别为属于事件类型c和角色r的支持集token集合，h_i^{final}是token的最终增强表示。这些原型将在后续的事件检测（作为分类参考）和参数匹配（作为角色表示先验）中起到关键作用。

2.3　事件–共指异构图建模

本文核心创新之一是构建事件–共指异构图，以端到端方式统一建模文档级语义单元及其复杂关系。ECHG的构建是一个从基础语义单元到复杂事件结构的递进过程，如图2所示。

2.3.1　异构图构建

首先，构建基础语义图（MSG），包含句子节点和实体提及节点，并通过句子–句子、句内提及–提及、跨句提及–提及句子–提及等边连接，捕获基础语义关联。

其次，为建模共指关系，引入代词节点，构建提及–代词共指子图（MSPG）。通过相似度计算函数建立共指关系。

$$t_{ij} = \frac{\exp(\text{sim}(\hat{h}_{mi},\hat{h}_{pj}))}{\sum_{j'}\exp(\text{sim}(\hat{h}_{mi},\hat{h}_{pj'})) + \varepsilon} \quad (15)$$

然后，基于阈值过滤和Top-K选择得到候选共指集合。

$$P = \{t_{ij}|i\in[1,k],j\in[1,u]\} \quad (16)$$

$$C_{\hat{h}_{pj}} = \{\hat{h}_{m1},\hat{h}_{m2},\cdots,\hat{h}_{ml_{t_{ij}=\eta}}\} \quad (17)$$

$$C'_{\hat{h}_{pj}} = \begin{cases} C\in\mathbb{R}^{l\times d_w} | \ l < K \\ C\in\mathbb{R}^{k\times d_w} | \ l \geqslant K \end{cases} \quad (18)$$

为进一步优化共指链接的权重，本文融合了实

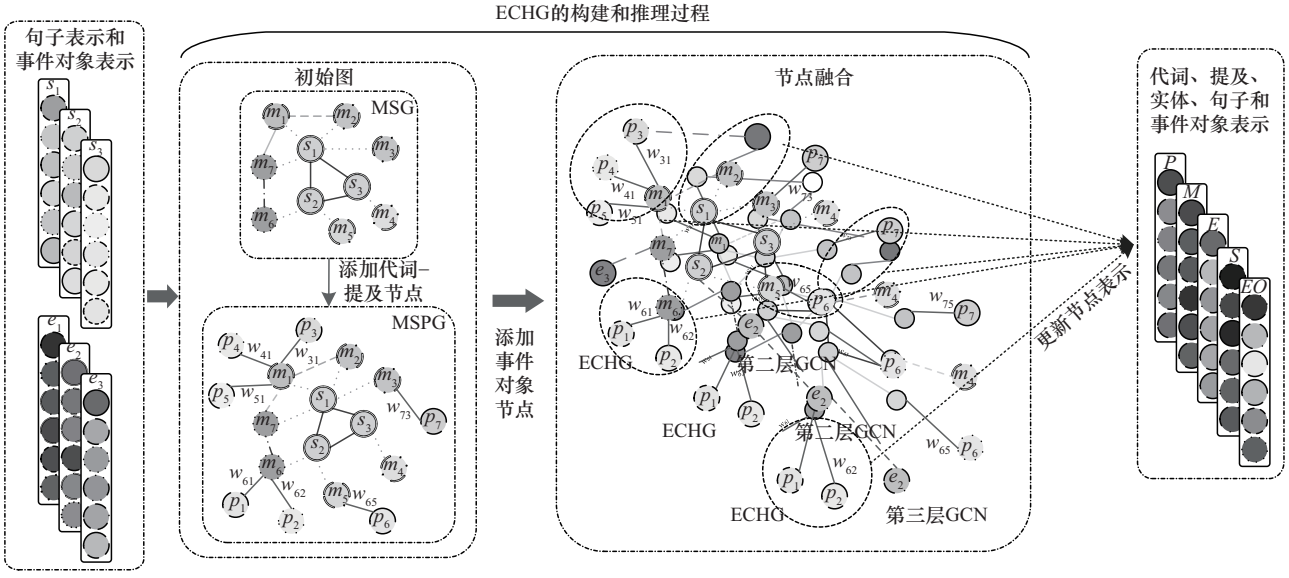

图2　事件–共指异构图模型

体共指一致性和LSS语义相似性2种语义信号，得到最终的联合权重。

1) 第一类权重(实体共指一致性)

本文为代词p_i的每一个候选提及m_i计算一个P-level权重。具体而言，首先将每个候选提及的语义相似度与其所属实体的共指频率相结合，构造一个初始得分，随后通过Softmax函数进行归一化，形成最终的概率分布。计算式为

$$(\omega_{i1}^{p},\omega_{i2}^{p},\cdots,\omega_{ij}^{p}) = \mathrm{Softmax}\left(\frac{f(m_1)\cdot t_{i1}}{l_j},\frac{f(m_2)\cdot t_{i2}}{l_j},\cdots,\frac{f(m_j)\cdot t_{ij}}{l_j}\right) \tag{19}$$

其中，$f(m_i)$表示候选提及m_i所指代的实体在整个文档中的出现频率；t_{ij}表示代词p_i与候选提及m_i之间的原始语义相似度，由式(18)计算得出；l_j是一个归一化因子，与提及m_i所在LSS的长度相关。引入该因子旨在消除上下文长度对相似度计算的偏差，确保不同来源的提及具有可比性。

2) 第二类权重(LSS相似性)

针对提及所在LSS的语义一致性，提取LSS的上下文向量作为表征，并使用余弦相似度计算提及与代词在LSS层面的相关性。

$$\hat{h}_{\mathrm{LSS}_{ij}} = \frac{1}{e}\sum_{k=1}^{e}\hat{h}_{ik} \tag{20}$$

其中，i表示第i个句子，e是当前LSS的token数量。然后，本文计算提及与代词在该LSS中的互注意力得分。

$$\omega_{ij}^{E} = \cos(\hat{h}_{mj},\hat{h}_{\mathrm{LSS}_{ij}}) \tag{21}$$

最终，提及与代词的联合权重为

$$\omega_{ij} = \alpha\cdot\omega_{ij}^{P} + \beta\cdot\omega_{ij}^{E} \tag{22}$$

其中，α、β是可调节参数。

引入事件对象节点V_e，构建完整ECHG。为每个检测到的事件触发词创建事件对象节点e_{obj}，由触发词、事件类型及其参数构成。

$$V_e = \{e_1,e_2,\cdots,e_L\} \tag{23}$$

其中，L是检测到的事件对象总数。

新增3类边以封装事件语义。

① 事件–参数边E_{ea}：连接事件与其上下文中的候选论元(提及V_m或代词V_p)。

$$E_{\mathrm{ea}} = \{(e_i,m_j),(e_i,p_k)\} \tag{24}$$

② 事件–句子边E_{es}：连接事件与其源句子。

$$E_{\mathrm{es}} = \{(e_i,s_j)\} \tag{25}$$

③ 事件–事件边E_{ee}：基于时序或因果推理连接相关事件。

$$E_{\mathrm{ee}} = \{(e_i,e_j)|\mathrm{rel}(e_i,e_j)\in\{\mathrm{before,after,cause,effect}\}\} \tag{26}$$

2.3.2　图推理与损失函数

构建完成的ECHG通过多层图卷积网络(GCN)进行推理。

对于第$l+1$层的节点表示更新式为

$$\hat{h}_i^{l+1} = \mathrm{ReLU}\left(\sum_{g\in G}\sum_{j\in N(i)}\frac{1}{|N(i)|}\boldsymbol{W}_g^{l}\hat{h}_j^{l} + B_g^{l}\right) \tag{27}$$

其中，$\hat{h}_i^{l+1}$是节点i在第$l+1$层的表示，$N(i)$表示节点i的邻居节点集合，$\boldsymbol{W}_g^{l}$和B_g^{l}分别是第l层的权重矩阵和偏置项，ReLU是激活函数。最终节点的表示通过拼接所有层的输出获得。

$$\bar{h}_i = [\hat{h}_i^{1};\hat{h}_i^{2};\cdots;\hat{h}_i^{n}] \tag{28}$$

在推理过程中，进行节点融合：将代词节点信息融合至其指代的提及节点。

$$\bar{h}_{mi} = \frac{1}{R}\sum_{r\in R}x_{ij}[\tilde{h}_{pj};\tilde{h}_{mi}] \tag{29}$$

其中，R表示与提及节点相关的代词数量，x_{ij}是联合关系权重。

接下来，将指向同一实体的提及节点融合为实体节点。

$$\bar{h}_{ei} = \frac{1}{K}\sum_{k\in K}\tilde{h}_{mk} \tag{30}$$

其中，K表示与实体节点相关提及节点数量。

句子节点的表示直接采用GCN输出。

$$\bar{h}_{si} = \bar{h}_i \tag{31}$$

ECHG的优化通过一个图结构损失L_{graph}实现，该损失包含2个部分。

1) 边预测损失

用于训练模型识别图中存在的边。对于每种类型的边集合$E_k\in E$，使用一个二分类器来预测边的存在性。

$$L_{\mathrm{edge}}^{(k)} = -\frac{1}{|E_k|}\sum_{(i,j)\in E_k}[y_{ij}^{(k)}\ln p_{ij}^{(k)} + (1-y_{ij}^{(k)})\ln(1-p_{ij}^{(k)})] \tag{32}$$

其中，$y_{ij}^{(k)}\in\{0,1\}$，表示节点i和节点j是否存在类型为k的边，$p_{ij}^{(k)} = \mathrm{sigmoid}\left(\frac{\boldsymbol{o}_i^{\mathrm{T}}\boldsymbol{o}_j}{d}\right)$为基于节点嵌入的点积计算相似度，$\boldsymbol{o}_i$和$\boldsymbol{o}_j$分别为节点$i$和节点$j$的GNN输出表示，$d$为缩放因子。总边预测损失为

$$L_{\text{edge}} = \sum_{k=1}^{K} \alpha_k \cdot L_{\text{edge}}^{(k)} \tag{33}$$

其中，K是边类型总数，α_k是可学习或预设的权重系数。

2) 节点表示一致性损失

通过对比学习使同一节点在不同图增强视图中的表示保持一致。

$$L_{\text{consistency}} = -\frac{1}{N}\sum_{i-1}^{N} \ln \frac{\exp\left(\dfrac{\text{sim}(h_i^{(1)}, h_i^{(2)})}{\tau}\right)}{\sum_{j=1}^{N} \exp\left(\dfrac{\text{sim}(h_i^{(1)}, h_i^{(2)})}{\tau}\right)} \tag{34}$$

其中，$h_i^{(1)}$和$h_i^{(2)}$表示节点i在2种图增强下的表示，$\text{sim}(\cdot,\cdot)$为余弦相似度或其他距离函数，τ为温度系数。

总图损失为二者加权和。

$$L_{\text{graph}} = \beta_1 \cdot L_{\text{edge}} + \beta_2 \cdot L_{\text{consistency}} \tag{35}$$

其中，β_1和β_2为控制两项之间相对重要性的超参数。

2.4 角色感知对比学习

为了在小样本条件下进一步增强模型对事件内部角色（如攻击者与目标）的区分能力，提出了角色感知对比学习机制。

对于每个事件对象e_i，定义其事件子图g_n，其中包含该事件的触发词节点和一组带有角色标签$\{r_k\}$的参数实体节点。

对比学习的目标是：对于子图中一个属于角色r_k的实体表示$z_i^{r_k}$，拉近它与同一事件子图内其他角色$r_l(l \neq k)$的实体表示$z_j^{r_l}$的距离，同时推远它与所有其他事件子图中实体表示的距离。

角色感知对比损失定义为

$$L_{\text{role}}^{\text{cl}}(g_n) = -\frac{1}{N}\sum_{n=1}^{N}\sum_{i\in g_n}\sum_{r_k\in \Re} \ln \frac{\exp\left(\dfrac{\text{sim}(z_i^{r_k}, z_j^{r_l})}{\tau}\right)}{\sum_{m\notin g_n} \exp\left(\dfrac{\text{sim}(z_i^{r_k}, z_m)}{\tau}\right)} \tag{36}$$

其中，τ是温度系数，$\text{sim}(\cdot)$为余弦相似度函数，R为角色集合。通过在一个训练批次内对所有事件子图计算式(36)并取平均，得到总体对比损失为

$$L_{\text{RWC}} = \frac{1}{|G_B|}\sum_{g_n\in G_B} L_{\text{role}}^{\text{cl}}(g_n) \tag{37}$$

其中，G_B表示当前训练批次B中所有事件子图的集合。为了同时优化图结构建模与语义表示学习，本文将ECHG图连接预测损失与角色感知对比损失联合优化。总体训练目标为

$$\min(\lambda_1 \cdot L_{\text{graph}} + \lambda_2 \cdot L_{\text{RWC}}) \tag{38}$$

其中，λ_1和λ_2为可调节权重参数。

2.5 事件抽取模块

在完成ECHG图建模与角色感知对比学习后，本模块利用聚合的全局语义信息进行事件检测与参数抽取，实现结构化事件识别。

2.5.1 事件检测

事件检测利用ECHG推理后增强的全局表示。对于每个候选事件实例，聚合其对应子图g_n中所有实体节点的表示得到事件级表示。

$$E_{\text{event}}^{(l)} = \text{maxpooling}(\{\bar{h}_i | v \in g_n\}) \tag{39}$$

然后，将其与句子表示$\bar{h}_{si}$以及事件级表示$\bar{h}_e$拼接后，通过分类器预测事件类型，该分类器由可训练参数矩阵$\boldsymbol{W}_{\text{type}}$和偏置项$b_{\text{type}}$组成，输出该事件实例属于各个类型的概率分布。

$$Y_{\text{type}}^{(l)} = \text{softmax}(\boldsymbol{W}_{\text{type}} \cdot \text{concat}(\bar{h}_{si}, \bar{h}_e, E_{\text{event}}^{(l)}) + b_{\text{type}}) \tag{40}$$

为了缓解类别不平衡问题，采用焦点损失（Focal Loss）优化事件检测目标。

$$L_{\text{det}} = -\sum_{l=1}^{L}(1 - P_{\text{det}})^{\gamma} \cdot \ln(P_{\text{det}}) \cdot Y_{\text{gold}} \tag{41}$$

其中，L为文档中识别出的事件实例总数，P_{det}为模型检测第l个事件实例属于其真实类型的概率值，Y_{gold}为对应的真实标签，γ为Focal Loss的调制因子。

2.5.2 参数抽取

参数抽取任务为已检测事件分配具体论元。基于ECHG提供的候选实体（含显式提及与代词解析结果），引入一组预定义的语义角色查询向量$\{r_l\}_{l=1}^{L}$，计算每个候选实体e_k在角色r_l下的匹配得分。

$$S(e_k, r_l) = \text{MLP}(\text{concat}(e_k, r_l, E_{\text{type}})) \tag{42}$$

其中，E_{type}是当前事件类型的嵌入表示，该得分融合了实体表示、角色查询和事件类型信息。

将任务建模为多标签分类，使用二元交叉熵损失。

$$L_{\text{arg}} = -\sum_{k=1}^{N}\sum_{l=1}^{K} y_{kl} \cdot \ln(p_{kl}) + (1 - y_{kl}) \cdot \ln(1 - p_{kl}) \tag{43}$$

其中，N是文档中的候选实体总数，K是预定义的语义角色数量，y_{kl}是真实标签，p_{kl}是模型预测的概率值。

同时集成角色感知对比学习目标，拉近共现角色实体、推远无关实体，增强语义一致性。在训练过程中，本文引入超参数μ_1、μ_2和μ_3作为目标函数权重，整体训练目标为多任务联合优化形式。

$$L_{\text{total}} = \mu_1 \cdot L_{\text{det}} + \mu_2 \cdot L_{\text{arg}} + \mu_3 \cdot L_{\text{RWC}} \quad (44)$$

3　实验

在CASIE和NetSecDoc数据集上进行了系统实验。通过对比主流方法和消融研究，验证了方法在事件检测和参数抽取中的有效性与鲁棒性。

3.1　实验设置

3.1.1　数据集

1) CASIE公共数据集

CASIE数据集由Satyapanich等[32]构建，包含1 000篇网络安全新闻，涵盖5类事件：Attack.Databreach、Discover.Vulnerability、Patch.Vulnerability、Attack.Phishing和Attack.Ransom，是当前主流的网络安全事件抽取基准。

2) NetSecDoc自建网络安全事件数据集

NetSecDoc包含1 000篇APT报告、通用漏洞披露CVE公告等，定义9类事件：DataBreach（145篇）、Phishing（101篇）、Ransom（109篇）、DDoSAttack（78篇）、Malware（81篇）、VulnerabilityExploitation（104篇）、SupplyChain（35篇）、VulnerabilityDiscover（189篇）以及VulnerabilityPatch（158篇）。由领域专家标注，确保质量。2个数据集采用统一标注规范。详细对比如表1所示。

表1　CASIE与NetSecDoc数据集特征对比

数据集	CASIE	NetSecDoc
事件类型	5	9
实体提及总数	4 200	6 800
代词总数	2 100	1 750
事件涉及的论元角色总数	45	62
每篇文档的平均token数量	537.5	975
平均每篇文档事件数	3.2	2.5
平均触发词长度	1.18	1.05
包含多个事件的文档占总文档的比例	41.3%	33.0%
包含多个事件类型的文档占总文档的比例	5.2%	3.5%
跨句事件的比例	41%	46.5%

3.1.2　模型配置

参照Yang等[18]的实验配置，输入长度2 048，隐藏层768维，4层Transformer编码器。GCN文档级3层，实体级2层，层间随机失活（dropout）率为0.3。使用AdamW优化器，初始学习率为3×10^{-5}，批量大小为64，测试为128，共训练100轮，选取开发集F1值最高模型。对比学习温度τ为0.9，损失权重μ_1=0.1、μ_2=1.0和μ_3=1.0。图构建阈值设为η为0.7，实验在2块NVIDIA V100（32 GB）上完成。

3.1.3　小样本采样策略

采用N-Way-D-Doc策略，随机选取N类事件和D篇文档。支持集用于学习模式，查询集用于推理，模拟低资源场景。

在CASIE与NetSecDoc上设置3-Way-1-Doc、3-Way-2-Doc和6-Way-2-Doc等组合。因文档较长、事件稀疏，输入限制为2 048 token，长文本采用滑动窗口切分。训练时屏蔽非任务类别，防止信息泄露。

选择有限N与D组合，主要考虑：N过大则满足条件文档稀少；D>2易超显存，且引入过多“None-type”（NOTA）增加噪声；同时需平衡任务复杂度与模型学习效率。

3.1.4　评估指标

事件检测采用准确率（Acc）和F1值。参数抽取在严格/部分匹配下计算精确率P、召回率R和F1值。采用宏平均（Macro-F1）和微平均（Micro-F1）评估性能，计算式分别为

$$\text{Macro-F1}=\frac{1}{N_{\text{class}}}\sum_{i=1}^{N_{\text{class}}}\text{F1}_i \quad (45)$$

$$\text{Micro-F1}=\frac{\sum\text{TP}}{\sum\text{TP}+\frac{1}{2}\sum(\text{FP}+\text{FN})} \quad (46)$$

其中，TP、FP、FN分别为真正例、假正例和假负例，Macro-F1反映各类别均衡性，Micro-F1体现整体稳定性。

3.1.5　基线模型

为验证本文方法的有效性，选取以下代表性模型作为基线。

文档端到端有向无环图模型（Doc2EDAG）[5]：通过有向无环图（DAG）建模事件结构，其贪婪解码简化版GreedyDec通过贪心策略构建论元链。

图交互追踪模型（GIT）[4]：Doc2EDAG的扩展，引入异构图神经网络与状态跟踪，增强全局上

下文建模。

原型–双向编码器表征模型（Proto-Bert）[33]：基于BERT的原型学习模型，适用于短文本少样本场景，但受限于512 token长度，难以处理长文档中的跨句依赖。

原型–长文档模型（Proto-LongFormer）[18]：结合原型学习与LongFormer，支持长文本输入，在*N*-Way-*D*-Doc设定下具备良好适应性。

基于事件提示的渐进学习模型（LAAP）[19]：利用事件提示模板学习事件类型与句子的关联关系，通过渐进式实体类型筛选策略提升文档级事件论元抽取的准确性。

动态超图模型（DSH）[34]：采用动态超图结构建模文档全局信息，通过相关性矩阵生成动态超边，有效解决事件论元分散与多事件交互的挑战。

3.2 主实验结果与分析

在CASIE与NetSecDoc数据集上，本文方法在3-Way-1-Doc、3-Way-2-Doc和6-Way-2-Doc这3种小样本设定下，与上述基线模型在事件检测和参数抽取任务中进行对比。实验结果表明，本文方法优于主流模型。Proto-LF虽支持长文本建模，但仅聚焦句子级聚合，缺乏对文档级结构、跨句共指与多粒度语义的深度融合。相比之下，本文方法通过局部语义片段增强细粒度理解，构建事件–共指异构图捕捉全局依赖，并引入角色感知对比学习提升原型区分能力，在长文本、低资源、多事件场景下展现出更强的鲁棒性与泛化性能。

如表2和表3所示，在CASIE数据集中，本文方法在3种设置下的F1值均排名第一。例如，在3-Way-2-Doc设置下F1值达56.83%，显著高于Proto-Bert（F1: 55.17%）和GIT（F1: 47.83%），同时明显优于LAAP（54.42%）和DSH（55.87%），表明其更强的跨文档建模能力。Doc2EDAG虽召回率较高（如3-Way-2-Doc下达到49.63%），但精确率低，存在过拟合。

在NetSecDoc数据集上，本文优势更加显著：在3-Way-2-Doc设置下F1值为76.51%，大幅优于Proto-LF（68.40%），且明显领先于LAAP（66.25%）和DSH（67.08%）。特别值得注意的是，在最具挑战性的6-Way-2-Doc设置下，本文方法F1

表2 CASIE数据集在3种不同设置上不同模型的事件检测实验结果对比

模型	3-Way-1-Doc			3-Way-2-Doc			6-Way-2-Doc		
	P	*R*	F1	*P*	*R*	F1	*P*	*R*	F1
Doc2EDAG	40.23%	37.51%	37.42%	42.81%	49.63%	**56.91%**	42.71%	34.32%	46.81%
GIT	41.67%	45.53%	43.91%	45.90%	52.30%	47.83%	41.19%	31.72%	**54.81%**
Proto-Bert	49.73%	42.30%	47.63%	54.01%	**54.90%**	55.17%	50.73%	37.31%	51.70%
Proto-LF	51.82%	53.91%	53.40%	58.03%	54.51%	56.21%	54.18%	42.63%	54.71%
LAAP	52.15%	50.83%	51.47%	57.83%	51.42%	54.42%	50.25%	46.83%	48.47%
DSH	53.42%	52.67%	53.04%	59.17%	52.89%	55.87%	52.67%	**48.92%**	50.73%
本文方法	**55.24%**	**54.09%**	**55.63%**	**63.61%**	53.51%	56.83%	**57.39%**	43.70%	52.08%

表3 NetSecDoc数据集在3种不同设置上不同模型的事件检测实验结果对比

模型	3-Way-1-Doc			3-Way-2-Doc			6-Way-2-Doc		
	P	*R*	F1	*P*	*R*	F1	*P*	*R*	F1
Doc2EDAG	54.90%	39.71%	59.43%	49.11%	**72.31%**	65.27%	58.21%	47.33%	50.61%
GIT	56.32%	47.52%	52.43%	64.09%	65.73%	59.91%	54.30%	**58.63%**	67.41%
Proto-Bert	64.23%	53.90%	53.12%	63.21%	70.41%	64.33%	69.80%	45.83%	57.88%
Proto-LF	66.53%	61.37%	63.53%	67.41%	65.13%	68.40%	72.50%	56.33%	64.71%
LAAP	67.85%	62.15%	64.87%	68.92%	63.85%	66.25%	70.35%	54.27%	61.42%
DSH	69.20%	63.45%	66.18%	70.15%	64.28%	67.08%	71.83%	55.89%	62.95%
本文方法	**70.60%**	**64.23%**	**68.31%**	**71.29%**	63.43%	**76.51%**	**74.50%**	57.41%	**79.17%**

值达到 79.17%，而 LAAP 和 DSH 分别仅为 61.42% 和 62.95%，优势差距超过 16 个百分点。这表明本文方法在处理长文本、多事件类型共存的情况下仍保持了较高的识别精度。

由表 4 和表 5 可见，严格匹配下，在 CASIE 数据集的 3-Way-2-Doc 设置中，本文方法 Macro-F1 达 57.01%，优于 LAAP（52.87%）和 DSH（56.25%），但在 3-Way-1-Doc 设置下，DSH 模型在 Macro-F1 指标上表现最佳（53.67%），略优于本文方法（53.73%），体现了其动态超图结构在简单小样本场景下的优势。在 NetSecDoc 数据集的严格匹配条件下，LAAP 模型在 3-Way-1-Doc 设置下 Macro-F1 指标上达到 56.25%，略优于本文方法的 56.93%，反映了事件提示模板在处理网络安全文档时的有效性。然而在更具挑战性的 6-Way-2-Doc 设置中，本文方法在 Micro-F1 上达到 61.27%，显著优于 LAAP（57.15%）和 DSH（58.42%），显示其在长文本、多类型事件中的稳定识别能力。

在部分匹配条件下，如表 6 和表 7 所示，各模型表现呈现差异化特点，在 CASIE 数据集中，LAAP 在 3-Way-1-Doc 的精确率领先（65.15%），而本文方法在 Macro-F1（60.01%）和 Micro-F1（62.93%）上保持优势；在 3-Way-2-Doc 设置下，本文方法 Macro-F1 达 63.03%，显著优于 LAAP（56.27%）和 DSH（57.45%）。在 NetSecDoc 数据集中，LAAP 和 DSH 展现出强劲竞争力：在 6-Way-2-Doc 设置下，DSH 的 Macro-F1 达 65.95%，略高于本文方法（64.82%），但本文方法在 Micro-F1 上以 69.45% 保持领先。总体而言，本文方法在复杂场景下保持竞争优势，特别是在处理跨句依赖时优势明显，而 LAAP 和 DSH 在特定场景下也表现良好。

3.3　消融实验与模块有效性分析

为验证各模块有效性，本文在 NetSecDoc 数据集上进行消融实验，如表 8 所示，移除核心模块均导致性能显著下降。其中，异构图建模模块的缺失对性能影响最大，尤其在最具挑战性的 6-Way-2-Doc 设置下，事件检测 F1 值从 79.2% 急剧下降至 56.9%，降幅达 22.3 个百分点。这一现象凸显了在极度稀缺的样本下，显式的结构化关系建模对于整合跨句信息、解耦多事件、实现全局推理具有不可

表 4　严格匹配条件下 CASIE 数据集在 3 种不同设置上不同模型的参数抽取实验结果对比

模型	3-Way-1-Doc				3-Way-2-Doc				6-Way-2-Doc			
	P	*R*	Macro-F1	Micro-F1	*P*	*R*	Macro-F1	Micro-F1	*P*	*R*	Macro-F1	Micro-F1
Doc2EDAG	38.87%	35.41%	34.15%	32.67%	39.08%	**59.13%**	33.43%	35.31%	36.91%	33.50%	33.23%	31.08%
GIT	47.65%	46.93%	44.41%	42.73%	49.20%	47.21%	45.93%	43.11%	56.21%	43.20%	43.83%	41.77%
Proto-Bert	43.33%	49.10%	45.83%	47.02%	52.55%	36.52%	46.73%	48.21%	45.88%	46.51%	46.33%	44.71%
Proto-LF	51.27%	53.09%	48.12%	50.61%	55.93%	50.84%	51.33%	49.42%	48.63%	52.60%	50.35%	**52.31%**
LAAP	55.85%	54.91%	49.25%	53.83%	59.75%	54.33%	52.87%	56.75%	56.58%	53.17%	52.33%	51.75%
DSH	53.42%	55.18%	53.67%	55.42%	60.92%	57.47%	56.25%	**58.18%**	57.83%	**55.42%**	53.67%	48.92%
本文方法	56.44%	**57.42%**	**53.73%**	**55.60%**	**61.70%**	51.43%	**57.01%**	55.37%	**59.10%**	54.12%	**54.23%**	50.68%

表 5　严格匹配条件下 NetSecDoc 数据集在 3 种不同设置上不同模型的参数抽取实验结果对比

模型	3-Way-1-Doc				3-Way-2-Doc				6-Way-2-Doc			
	P	*R*	Macro-F1	Micro-F1	*P*	*R*	Macro-F1	Micro-F1	*P*	*R*	Macro-F1	Micro-F1
Doc2EDAG	41.40%	38.21%	34.83%	36.80%	42.44%	39.13%	35.50%	37.63%	39.91%	36.07%	33.59%	35.50%
GIT	50.63%	49.55%	45.03%	47.21%	52.13%	50.52%	**58.71%**	45.12%	48.07%	46.03%	54.39%	44.08%
Proto-Bert	51.73%	**60.19%**	46.92%	47.11%	55.25%	52.73%	48.50%	51.53%	54.62%	49.03%	48.43%	47.33%
Proto-LF	54.09%	53.03%	51.80%	52.73%	58.14%	53.92%	49.33%	55.69%	59.41%	56.73%	52.08%	50.47%
LAAP	63.15%	59.45%	56.25%	**58.83%**	**65.92%**	61.85%	56.17%	58.42%	65.35%	57.27%	55.33%	57.15%
DSH	**63.42%**	59.83%	55.67%	57.92%	65.18%	62.47%	57.25%	59.88%	66.83%	**59.89%**	56.95%	58.42%
本文方法	62.53%	57.63%	**56.93%**	58.33%	64.61%	**62.53%**	53.42%	**60.88%**	**67.90%**	59.33%	**58.01%**	**61.27%**

表6 部分匹配条件下CASIE数据集在3种不同设置上不同模型的参数抽取实验结果对比

模型	3-Way-1-Doc				3-Way-2-Doc				6-Way-2-Doc			
	P	*R*	Macro-F1	Micro-F1	*P*	*R*	Macro-F1	Micro-F1	*P*	*R*	Macro-F1	Micro-F1
Doc2EDAG	40.83%	37.46%	34.01%	36.33%	41.70%	38.41%	36.72%	37.21%	38.44%	**67.24%**	33.79%	35.45%
GIT	49.42%	48.81%	44.53%	46.22%	**69.72%**	49.69%	45.57%	47.85%	47.17%	45.53%	43.41%	45.13%
Proto-Bert	47.23%	45.28%	43.09%	45.55%	51.33%	48.31%	44.52%	46.29%	50.07%	49.69%	45.47%	47.63%
Proto-LF	51.62%	49.01%	47.53%	49.13%	51.55%	52.39%	48.27%	50.82%	55.03%	53.18%	49.33%	51.35%
LAAP	**65.15%**	59.83%	55.25%	57.41%	62.97%	60.49%	56.27%	58.25%	63.35%	59.27%	55.33%	57.25%
DSH	62.42%	61.67%	56.67%	59.92%	64.18%	62.89%	57.45%	59.89%	65.83%	62.42%	60.95%	59.47%
本文方法	64.73%	**62.50%**	**60.01%**	**62.93%**	54.17%	**67.33%**	**63.03%**	**65.29%**	**70.41%**	56.13%	**61.30%**	**63.71%**

表7 部分匹配条件下NetSecDoc数据集在3种不同设置上不同模型的参数抽取实验结果对比

模型	3-Way-1-Doc				3-Way-2-Doc				6-Way-2-Doc			
	P	*R*	Macro-F1	Micro-F1	*P*	*R*	Macro-F1	Micro-F1	*P*	*R*	Macro-F1	Micro-F1
Doc2EDAG	43.25%	40.23%	36.83%	38.92%	44.61%	41.83%	**66.18%**	39.87%	41.70%	38.33%	35.51%	37.07%
GIT	50.21%	48.54%	45.23%	47.01%	52.33%	51.65%	46.33%	48.02%	54.17%	**68.01%**	48.59%	50.43%
Proto-Bert	53.47%	52.53%	48.11%	**65.37%**	57.83%	51.23%	37.32%	50.14%	54.73%	48.03%	49.27%	47.69%
Proto-LF	57.80%	59.21%	52.93%	46.33%	60.92%	52.43%	48.41%	53.69%	66.45%	52.67%	60.33%	57.41%
LAAP	72.15%	63.45%	60.25%	62.42%	71.92%	68.85%	61.17%	64.25%	78.35%	61.27%	64.33%	66.15%
DSH	69.42%	65.67%	61.67%	63.92%	**73.18%**	69.89%	62.25%	65.88%	79.03%	63.42%	**65.95%**	67.42%
本文方法	**73.71%**	**65.89%**	**63.72%**	50.60%	72.43%	**70.52%**	51.44%	**68.93%**	**79.64%**	55.31%	64.82%	**69.45%**

替代的关键作用。Longformer模块的移除显著削弱了长文档理解能力，在3-Way-2-Doc设置下F1值下降8.1个百分点。角色感知对比学习模块的去除削弱了模型对事件内部角色的判别能力，在6-Way-2-Doc设置下F1值下降11.8个百分点。阈值控制模块的移除对性能影响相对温和（平均下降2~3个百分点），但其在过滤低质量边连接、提升图结构纯净度方面的稳定作用不容忽视。

表8 不同模块配置下的事件检测消融实验研究

模块配置	3-Way-1-Doc	3-Way-2-Doc	6-Way-2-Doc
本文	68.3%	76.5%	79.2%
移除Longformer模块	63.5%	68.4%	64.7%
移除异构图建模模块	61.8%	62.3%	56.9%
移除角色感知对比学习模块	65.7%	66.2%	67.4%
移除阈值控制模块	67.3%	71.3%	72.5%

从表9可以看出，在严格匹配条件下，移除异构图模块影响最大，6-Way-2-Doc中Macro-F1下降6.3%，表明其对参数边界识别至关重要。Longformer模块的缺失在长文档中影响更明显，特别是在3-Way-2-Doc设置下Macro-F1下降了3.6个百分点。角色对比学习模块的去除降低了模型对稀有参数角色的识别能力；阈值控制模块的移除虽未引起剧烈波动，但在多个设置下仍表现出轻微下降，说明其在图结构优化方面具有一定价值。

部分匹配趋势类似，本文方法仍保持优势。如表10所示，异构图建模模块在6-Way-2-Doc设置下Macro-F1下降了6.4个百分点，说明其不仅增强了模型对边界精确性的识别，还提升了对参数提及内容的整体理解；Longformer模块的移除在长文档场景中依然影响较大，尤其是在3-Way-1-Doc和6-Way-2-Doc设置下；角色对比学习模块的去除使得模型在多事件类型共存情况下泛化能力减弱；阈值控制模块的移除影响较小，但仍表现出一定下降，说明其在图结构推理过程中起到辅助优化作用。

3.4 不同样本量下的性能对比

在NetSecDoc数据集上测试不同支持集大小（{1,2,4,8}）下的性能，如图3所示。

表9　　严格匹配条件下不同模块配置的参数抽取消融实验研究

模块配置	3-Way-1-Doc		3-Way-2-Doc		6-Way-2-Doc	
	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
本文	56.9%	58.3%	53.4%	60.9%	58.0%	61.3%
移除Longformer模块	52.1%	54.7%	49.8%	57.3%	53.2%	58.4%
移除异构图建模模块	50.2%	52.9%	47.6%	55.3%	51.7%	56.7%
移除角色对比学习模块	54.6%	56.8%	51.1%	59.1%	56.2%	59.6%
移除阈值控制模块	55.8%	57.6%	52.6%	60.2%	57.3%	60.9%

表10　　部分匹配条件下不同模块配置的参数抽取消融实验研究

模块配置	3-Way-1-Doc		3-Way-2-Doc		6-Way-2-Doc	
	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
本文	63.7%	50.6%	51.4%	68.9%	64.8%	69.5%
移除Longformer模块	59.1%	48.4%	47.3%	65.7%	60.3%	66.2%
移除异构图建模模块	57.2%	46.9%	45.5%	63.0%	58.4%	64.1%
移除角色对比学习模块	61.3%	49.6%	49.7%	67.1%	62.1%	67.8%
移除阈值控制模块	62.5%	50.2%	50.8%	68.3%	63.9%	68.9%

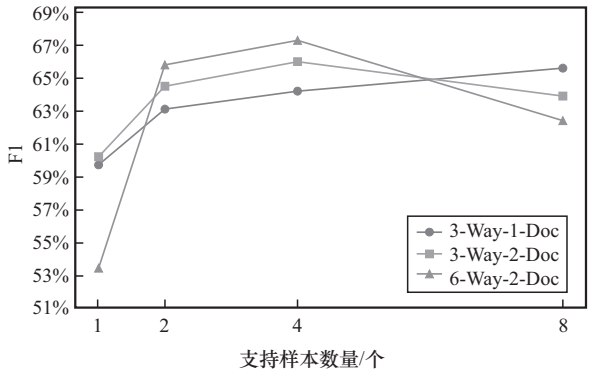

图3　不同支持集大小下的事件检测性能对比

从图3可以看出，随着支持样本数量的增加，模型整体性能呈现出上升趋势，但在某些设置下也出现了明显的性能波动。这一现象表明，样本增多未必持续提升效果：过多样本易导致原型偏向、引入语义噪声，干扰图建模与对比学习，影响泛化与稳定性。

如图4所示，参数抽取任务中2类指标均表现出相似的趋势：在4个样本时达到峰值，随后在8个样本时略有回落。这说明适量的支持样本能够有效提升模型对事件及其参数之间复杂语义关系的理解能力，而过多样本则可能引入噪声，影响最终参数抽取的稳定性。

4　案例分析

4.1　成功案例对比

为验证本文方法在真实网络安全场景下的有效

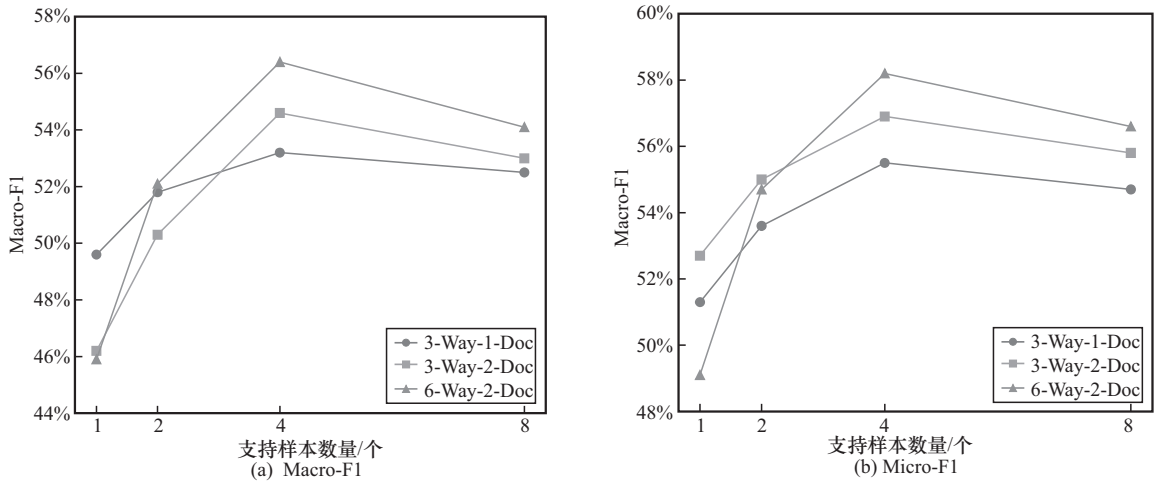

图4　不同支持集大小下的参数抽取性能对比

性，选取了一段包含多个安全漏洞参数的复杂文档（Document 1）进行深入分析。该文档描述了跨站脚本攻击（XSS）漏洞与打印后台处理程序（Print Spooler）漏洞的发现与利用情况，涉及跨句论元分布、多事件交织等典型挑战。表11展示了针对Document 1的事件抽取结果对比，通过与基线模型Proto-LF的对比，本文方法在关键参数抽取任务中展现出显著优势。

Document 1

[S2] …

[S3] Although security researchers in the industry have been looking for bugs in **Spooler** for more than a decade this year.

[S4] In a blog post, Young detailed how he discovered not one but three persistent **XSS vulnerabilities** – two of which remain unpatched – that could allow a **malicious actor** to **access and control the app**.

[S5] One of the vulnerabilities due to be discussed, tracked as **CVE-2021-1675** and issued with a CVSS score of 7.8, is a critical **Print Spooler bug** that was included in Microsoft's latest Patch Tuesday, published on **June 8**.

[S6] …

[S9] Multiple vulnerabilities in **Cisco Small Business routers** could allow any unauthenticated actor to potentially plant a backdoor in devices, allowing for persistent access to internal networks.

[S10] Netting a $20,000 bug bounty payout for their exploit, the researchers **inserted malicious JavaScript** into web pages along with text written in a language that was non-native to a target user's Edge settings.

[S11] …

1) 跨句论元关联能力：本文构建的事件–共指异构图成功建立了[S9]中“Cisco Small Business routers”与[S4-S5]中漏洞描述的语义关联。通过提及节点与事件对象的动态连接，模型克服了传统方法在长距离依赖建模上的局限。

2) 多事件区分与焦点识别：面对文档中并存的XSS漏洞与Print Spooler漏洞，本文方法通过角色感知对比学习机制，准确识别出以CVE-2021-1675为核心的主要漏洞事件，避免了Proto-LF将[S4]中显性提到的“XSS vulnerabilities”误认为主要漏洞的名称混淆问题。

3) 因果关系推理能力：本文方法成功建立了[S10]中攻击方法（inserted malicious JavaScript）与[S4]中攻击结果（access and control the app）之间的跨句因果关系，体现了深层的语义理解能力。相比之下，Proto-LF未能识别此因果关系，导致攻击结果识别缺失。

表11 Document 1事件抽取结果对比

事件类型及参数	本文方法	Proto-LF
Event Type	Vulnerability Exploits	Vulnerability Exploits
Attacker	malicious actor	malicious actor
Victim	Cisco Small Business routers	Null
Vulnerability Name	Print Spooler bug	XSS vulnerabilities
Vulnerability ID	CVE-2021-1675	CVE-2021-1675
Vulnerability Located Device	Spooler	Spooler
Vulnerability published Date	June 8	June 8
Attack Method	inserted malicious JavaScript	inserted malicious JavaScript
Attack Results	access and control the app	Null

4.2 局限性案例验证

尽管在案例1中表现优异，为全面评估方法性能，进一步分析了本文方法出现识别偏差的代表性案例（Document 2）。如表12所示，本文方法在Document 2上出现了识别偏差，具体分析如下。

Document 2

[S1] In a sophisticated **supply chain attack** discovered last quarter, **threat actors** infiltrated a trusted **software vendor's distribution network**.

[S2] By compromising the vendor's update servers, **the attackers** stealthily **injected malicious payloads** into legitimate software updates.

[S3] Thousands of enterprise clients unknowingly installed these trojanized updates over **a three-month period**.

[S4] The backdoors **established persistent remote access** to corporate networks, enabling lateral movement and data exfiltration.

[S5] **Security analysts** at multiple firms **detected anomalous network traffic** patterns but initially attributed them to routine maintenance activities.

[S6] …

表 12　Document 2 事件抽取结果对比

事件类型及参数	Ground Truth	本文方法
Event Type	Supply Chain Attack	Supply Chain Attack
Attacker	threat actors	Security analysts
Attack Target	software vendor's distribution network	software vendor's distribution network
Attack Method	injecting malicious payloads	injecting malicious payloads
Attack Results	established persistent remote access	~~detected anomalous network traffic~~
Duration	three-month period	three-month period

1) 复杂指代链解析失败：模型未能正确追踪从[S1]的“threat actors”到[S2]的“the attackers”的指代关系，特别是在[S5]中引入"Security analysts"后，模型混淆了攻击方与防御方的角色边界，错误地将安全分析员识别为攻击者。

2) 多阶段事件边界模糊：文档描述了攻击实施[S1]~[S4]、攻击检测[S5]和事后调查[S6]等阶段。本文方法未能清晰划分这些阶段，将检测响应误认为攻击行为的一部分。

3) 因果时序关系误解：模型错误地将[S5]中的检测行为识别为攻击结果，而忽略了[S4]中明确描述的“established persistent remote access”这一直接后果。这反映了对网络安全事件标准生命周期理解的不足。

5　结束语

本文提出融合事件–共指异构图和角色感知对比学习的文档级事件抽取框架。通过多粒度异构图建模事件与参数的语义依赖，不需要外部共指工具，实现端到端学习。在此基础上，引入角色感知的对比学习机制，进一步增强了事件内部不同语义角色之间的表示区分能力。实验结果表明，本文方法在CASIE与NetSecDoc数据集上均显著优于多种主流基线模型，尤其在低资源条件下展现出更强的泛化能力与鲁棒性。未来工作将聚焦于增强模型对复杂指代关系和因果时序的理解能力，探索轻量化图推理机制并融入领域知识图谱，以进一步提升事件知识抽取的自动化水平及其在威胁情报系统中的实际应用效能。

参考文献：

[1] WAGNER T D, MAHBUB K, PALOMAR E, et al. Cyber threat intelligence sharing: Survey and research directions[J]. Computers & Security, 2019, 87: 101589.

[2] 施蒂妮, 曾剑平. 利用微调大语言模型的检索增强文档级多事件抽取[J]. 小型微型计算机系统, 2025: 1-11.
SHI D N, ZENG J P. Retrieval-augmented document-level multi-event extraction with fine-tuned large language models[J]. Journal of Chinese Computer Systems, 2025: 1-11.

[3] ZHUANG L, FEI H, HU P. Syntax-based dynamic latent graph for event relation extraction[J]. Information Processing & Management, 2023, 60(5): 103469.

[4] XU R, LIU T, LI L, et al. Document-level event extraction via heterogeneous graph-based interaction model with a tracker[J]. arXiv Preprint, arXiv: 2105.14924, 2021.

[5] ZHENG S, CAO W, XU W, et al. Doc2EDAG: an end-to-end document-level framework for Chinese financial event extraction[J]. arXiv Preprint, arXiv: 1904.07535, 2019.

[6] REN X, YANG T, WANG Y, et al. Learning disentangled representation by exploiting pretrained generative models: a contrastive learning view[J]. arXiv Preprint, arXiv: 2102.10543, 2021.

[7] 何丽, 李泽龙, 宋靖靖, 等. 提示模板引导的文档级金融事件抽取方法研究[J]. 数据分析与知识发现, 2025, 9(7): 154-164.
HE L, LI Z L, SONG J J, et al. Document level financial event extraction method guided by prompt template[J]. Data Analysis and Knowledge Discovery, 2025, 9(7): 154-164.

[8] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]//International Conference on Machine Learning. Cambridge: JMLR, 2020: 1597-1607.

[9] TANG H Z, CAO Y N, ZHANG Z Y, et al. Multi-granularity heterogeneous graph for document-level relation extraction[C]//Proceedings of the ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2021: 7683-7687.

[10] PRADHAN R, KUMAR D. Event segmentation and event boundary advantage: role of attention and postencoding processing[J]. Journal of Experimental Psychology General, 2022, 151(7): 1542-1555.

[11] CHEN Z, JI W T, DING L L, et al. Document-level multi-task learning approach based on coreference-aware dynamic heterogeneous graph network for event extraction[J]. Neural Computing and Applications, 2024, 36(1): 303-321.

[12] EVANS R, ORĂSAN C. Sentence simplification for semantic role labelling and information extraction[C]//Proceedings of Natural Language Processing in a Deep Learning World. Bulgaria: Incoma Ltd., 2019: 285-294.

[13] GAO X P, DIAO Z G, WEI K L, et al. Event extraction via rules and machine learning[C]//Proceedings of the 2019 IEEE 6th International Conference on Cloud Computing and Intelligence Systems (CCIS). Piscataway: IEEE Press, 2019: 41-46.

[14] SHEN S R, QI G L, LI Z, et al. Hierarchical Chinese legal event ex-

traction via pedal attention mechanism[C]//Proceedings of the 28th International Conference on Computational Linguistics. Stroudsburg: ACL, 2020: 100-113.

[15] LIU Y H, GAO N, ZHANG Y F, et al. Enhancing document-level event extraction via structure-aware heterogeneous graph with multi-granularity subsentences[C]//Proceedings of the ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2024: 12657-12661.

[16] ZHOU J, SHUANG K, AN Z Z, et al. Improving document-level event detection with event relation graph[J]. Information Sciences, 2023, 645: 119355.

[17] ALJABARI A, DUAIBES L, JARRAR M, et al. Event-arguments extraction corpus and modeling using BERT for arabic[J]. arXiv Preprint, arXiv: 2407.21153, 2024.

[18] YANG X, LU Y, PETZOLD L. Few-shot document-level event argument extraction[J]. arXiv Preprint, arXiv: 2209.02203, 2022.

[19] XU J H, YANG C, KANG X J. LAAP: learning the argument of an entity with event prompts for document-level event extraction[J]. Neurocomputing, 2025, 613: 128584.

[20] SONG C Y, CAI F, WANG M R, et al. TaxonPrompt: taxonomy-aware curriculum prompt learning for few-shot event classification[J]. Knowledge-Based Systems, 2023, 264: 110290.

[21] CHEN Y B, XU L H, LIU K, et al. Event extraction via dynamic multi-pooling convolutional neural networks[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Stroudsburg: ACL, 2015: 167-176.

[22] NGUYEN T H, CHO K, GRISHMAN R. Joint event extraction via recurrent neural networks[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: ACL, 2016: 300-309.

[23] WANG S Y, ZHENG J M, CHEN W Y, et al. MultiPLe: multilingual prompt learning for relieving semantic confusions in few-shot event detection[C]//Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2023: 2676-2685.

[24] QIU Z T, WU J, YANG J, et al. Heterogeneous social event detection via hyperbolic graph representations[J]. IEEE Transactions on Big Data, 2025, 11(1): 115-129.

[25] LIU L, LIU M, LIU S S, et al. Event extraction as machine reading comprehension with question-context bridging[J]. Knowledge-Based Systems, 2024, 299: 112041.

[26] LI H, ZHAO X, YU L, et al. DEEDP: document-level event extraction model incorporating dependency paths[J]. Applied Sciences, 2023, 13 (5): 2846.

[27] LEE K, HE L, ZETTLEMOYER L. Higher-order coreference resolution with coarse-to-fine inference[J]. arXiv Preprint, arXiv: 1804.05392, 2018.

[28] ZHANG M Y, FANG F, LI H, et al. MHGEE: event extraction via multi-granularity heterogeneous graph[C]//International Conference on Computational Science. ICCS 2022. Berlin: Springer, 2022: 473-487.

[29] HUANG G H, MIN Z P, GE Q, et al. Towards document-level event extraction via Binary Contrastive Generation[J]. Knowledge-Based Systems, 2024, 296: 111896.

[30] PALIT S, BANERJEE B, CHAUDHURI S. Prototypical quadruplet for few-shot class incremental learning[J]. Procedia Computer Science, 2023, 222: 25-34.

[31] MANN W C, THOMPSON S A. Rhetorical structure theory: description and construction of text structures[C]//Natural Language Generation. Berlin: Springer, 1987: 85-95.

[32] SATYAPANICH T, FERRARO F, FININ T. CASIE: extracting cybersecurity event information from text[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(5): 8749-8757.

[33] SNELL J, SWERSKY K, ZEMEL R S. Prototypical networks for few-shot learning[C]//Proceedings of the Neural Information Processing Systems. Massachusetts: MIT Press, 2017: 4077-4087.

[34] REN Q, WANG W H, YU J, et al. Dynamic structure hypergraph for document-level event extraction[C]//Proceedings of the ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2025: 1-5.

[作者简介]

王金芳（1996– ），女，河南濮阳人，河南工业大学博士生，主要研究方向为信息安全、网络安全事件抽取等。

郭渊博（1975– ），男，陕西周至人，博士，海南大学教授、博士生导师，主要研究方向为网络安全、数据挖掘、机器学习和人工智能安全等。

邓淼磊（1977– ），男，河南南阳人，博士，河南工业大学教授、博士生导师，主要研究方向为信息安全、人工智能安全等。

汤萌萌（1989– ），女，河南信阳人，博士，郑州航空工业管理学院讲师，主要研究方向为信息安全、网络安全事件抽取等。

Umer Nauman（1992– ），男，巴基斯坦人，河南工业大学在站博士后，主要研究方向为网络安全、量子安全、区块链安全和密码学等。