



第 3 章

语句级事件抽取技术

3.1 语句级事件抽取概述

语句级事件抽取(sentence-level event extraction)是信息抽取领域的核心任务之一,旨在从单个自然语言语句中识别事件触发词,判定事件类型,抽取与事件相关的论元并判断其充当的角色,最终将非结构化文本转化为结构化的事件信息^[1]。语句级事件抽取任务仅关注当前语句内的上下文信息,其核心目标是在语句内完成事件要素的精准解析和抽取,不涉及跨句信息整合,与篇章级事件抽取(需跨句子对上下文关系建模)形成互补。

语句级事件抽取包含事件检测与事件论元抽取。前者判断输入文本中是否存在符合要求的事件以及事件对应的类型,后者关注获取充当给定事件的角色论元。对于同时完成上述两个子任务的研究,称为事件识别及其论元抽取。

本章将围绕事件抽取的 3 个核心任务展开递进式讨论,系统呈现近年来语句级事件抽取的前沿解决方案。在 3.2 节中,针对事件检测、事件论元抽取、事件识别及其论元抽取 3 个核心任务,分别从通用技术演进路径切入,并结合深度学习技术(如 CNN、RNN 等)探讨模型优化策略。3.3 节聚焦依存关系增强的事件检测,结合句法与语义依存结构的深度融合,展现句法结构信息如何提升触发词识别精度。3.4 节针对中文金融领域的特殊挑战,详细介绍句法语义依存图与缺失补全规则,解决特定领域事件识别与成分缺失问题。在 3.5 节以开放事件抽取为目标,介绍一种双向多通道层次图注意力网络,展示如何通过多特征融合与图结构优化突破封闭域限制,实现更具泛化性的开放事件抽取。

本章组织结构的设计以“任务分解、技术深化”为逻辑。其一,从任务流程出发,以事件抽取任务的固有流程为脉络,遵循“事件检测→论元抽取→事件识别及要素抽取”的技术逻辑。其二,从技术层面出发,从基础工作总结到前沿技术解析,再到垂直领域与开放场景的技术延伸,既通过不同的深度学习模型的对比展现技术演进规律,又针对依存关系利用不足、中文缺失结构复杂、开放事件泛化困难等问题提出相应的解决方案。

3.2 现有语句级事件抽取技术

3.2.1 事件检测技术

事件检测作为事件抽取的基石,旨在识别文本中包含的事件,核心是识别触发词并判定其类型。该任务面临两大挑战:触发词多义性导致类型误判,以及事件类型长尾分布造成低频事件识别困难。



早期研究聚焦于特征工程,随后逐步从依赖词法、实体类型等浅层特征向深层语义表示过渡。Li 等(2015)^[2]通过抽象语义表示(AMR)技术融合词法、句法特征生成深层语义向量,Nguyen 等(2015)^[3]引入卷积神经网络(CNN)自动学习隐藏特征,突破传统特征工程的误差瓶颈,并通过对 Skip-Grams 建模进一步捕捉非连续语义结构^[4]。针对早期模型仅支持单字符检测的局限,Ghaeini 等(2016)^[5]利用前向后向循环神经网络(FBRNN)实现单词/短语级的事件结块检测;Liu 等(2017)^[6]首次通过注意力机制显式融入论元信息,强化触发词与论元的语义关联。随着技术的发展,模型开始对上下文语义进行深度挖掘。例如,Liu 等(2018)^[7]从论元及上下文可以被多次使用的角度设计了触发词检测的动态记忆网络,充分挖掘上下文的语义信息。

然而,先前的工作存在着未充分利用句法依存结构信息(如仅将依存关系作为辅助特征,未显式对依存边的语义差异建模)以及复杂事件场景适应性有限(在多事件重叠语句及触发词模糊场景中,低频事件漏检与多义触发词误判问题仍较突出)等瓶颈。针对这些问题,本节聚焦四大前沿的技术路线,解析其核心思想和技术突破,探讨如何解决事件检测中的这些问题。

1. 基于句法关系的边增强图卷积网络

基于句法关系的边增强图卷积网络(Edge-Enhanced Graph Convolution Networks, EE-GCN)^[8]的核心动机在于:传统基于图卷积网络(GCN)的事件检测模型虽利用了句法依存结构,但多将依存边视为无类型的连接,忽略了依存标签(如“nsubj”“dobj”)所蕴含的语义信息。例如,“nsubj”边表达了触发词的主语角色,这种方向性语义对事件类型判定具有关键作用。此外,同一依存标签在不同语境下可能传递不同信号(如“nmod”既可能表示时间,也可能表示地点),传统方法采用静态标签嵌入无法捕捉这种动态语义差异。

因此,该方法的核心思想是通过动态更新依存边的表示,显式融合句法结构与依存标签信息,同时解决传统 GCN 对依存标签利用不足和静态标签表示的局限性。EE-GCN 的整体架构包含输入层、边增强图卷积层和分类层,其核心模块在于边感知节点更新模块与节点感知边更新模块的交互设计。接下来结合核心步骤阐述其实现逻辑。

(1) 输入处理与图结构构建。

① 节点特征初始化。输入句子,经词嵌入和实体类型嵌入拼接后,通过 Bi-LSTM_o 为上下文语义编码,生成初始节点表示。

② 边特征建模。构建三维邻接张量 \mathbf{E} ,其中 $\mathbf{E}_{i,j,r}$ 表示节点 w_i 与 w_j 之间的依赖关系。若节点 i 和节点 j 之间存在依赖边且标签为 r ,则 $\mathbf{E}_{i,j,r}$ 初始化为 r 的嵌入向量;否则设为全零向量。例如,依赖标签“nsubj”对应的嵌入向量将用于初始化触发词与其主语之间的边。

(2) 边增强图卷积层。通过迭代更新节点与边的表示,实现语义信息的双向增强。

① 边感知节点更新(EANU)。对于第 l 层,节点表示通过聚合邻居节点的特征更新。具体地,对邻接张量的每一通道(对应一种依赖类型)执行图卷积操作,如式(3.1)所示。

$$\mathbf{H}_i^l = \sigma(\mathbf{E}_{:,i,:}^{l-1} \mathbf{H}_i^{l-1} \mathbf{W}) \quad (3.1)$$

其中, \mathbf{W} 为可学习参数, σ 为 ReLU 激活函数。通过平均池化整合多通道信息,得到更新后的节点表示 \mathbf{H}^l 。

② 节点感知边更新(NAEU)。动态调整边表示以反映上下文语义。对于边 (i, j) ,其

更新如式(3.2)所示。

$$\mathbf{E}_{i,j,:}^l = \mathbf{W}_u (\mathbf{E}_{:,i}^{l-1} \oplus \mathbf{h}_i^l \oplus \mathbf{h}_j^l) \quad (3.2)$$

其中, \mathbf{h}_i^l 和 \mathbf{h}_j^l 为节点 i 和 j 的更新后表示, \mathbf{W}_u 为变换矩阵。该模块通过拼接节点与边的特征,生成上下文相关的动态边表示,解决静态标签表示的局限性问题。最终节点表示经全连接层和 Softmax 函数后输出事件类型概率分布。

2. 知识保持型增量式社会事件检测模型

EE-GCN 聚焦于动态句法关系的建模,而知识保持型增量式社会事件检测模型 (Knowledge-Preserving incremental social event detection via heterogeneous GNNs, KPGNN)^[9] 主要聚焦于如何在动态数据流中持续检测新事件并保留已有知识。现有一些传统方法(如增量聚类、社区检测)忽略社交消息中的多模态信息(如文本、用户、时间)及其结构化关联,导致语义与上下文信息未被充分利用。同时,增量场景下新事件类型不断涌现,但传统分类模型依赖固定类别标签,难以灵活扩展。此外,随着社交流数据持续涌入,模型需高效处理动态更新的图结构,避免内存与计算资源过载。

基于此,该模型的核心动机在于:通过异构图对社交媒体中的多模态元素(用户、实体、关键词等)建模,结合对比学习与增量学习框架,实现动态事件检测与知识积累的协同优化。

KPGNN 的架构围绕异构社交图的构建与动态维护展开,核心步骤如下:

(1) 异构社交图建模。

① 多模态节点构建。将社交消息中的文本、用户、实体、时间等元素抽象为异构节点,并通过边连接相关节点(如消息与其包含的词汇、提及的用户之间的关联)。

② 同质化消息图转换。将异构图转换为同质的消息图(仅包含消息节点),通过矩阵计算生成邻接矩阵,其中消息节点之间边的权重由共享的异构元素(如共同词汇、用户)决定。例如,若两消息共享多个词汇,其边的权重增大。

(2) 增量式图神经网络编码。新消息块到达时,将对应节点与边动态插入消息图;维护阶段定期移除过时消息节点,保持图的轻量化。随后采用多层图注意力网络(GAT),通过聚合邻居节点的语义信息,生成消息节点的嵌入表示。

3. 基于潜在图与依存图网络融合的事件检测模型

在复杂文本中,多个事件常共存于同一句子中,而传统依存树仅反映句法结构,无法自动建立触发词与语义关键节点(如“职位”“机构”)的关联,导致歧义触发词分类困难。同时,依存树中远距离相关词需通过多跳路径连接,句法解析错误易沿路径累积,影响事件分类。此外,传统图网络仅依赖依存树,忽视潜在语义结构,难以适应多事件复杂场景。针对这一系列问题,研究者提出基于潜在图与依存图网络融合的事件检测模型(Syntactic Graph and Induction Latent Graph, SGILG)^[10]。该模型的核心思想是通过潜在图补充依存树的语义缺失,利用多视角图结构融合提升事件检测能力。

具体来说,SGILG 架构包含四大模块:

(1) 双图结构构建。首先,利用 Stanford CoreNLP 句法工具,生成句法依存树(依存图),构建邻接矩阵,捕捉句法依赖关系。随后,通过自注意力机制生成语义关联矩阵,自动学习触发词与上下文的关键连接(如“离职”与“董事长”的语义关联),从而构成潜在图。

(2) 图卷积网络编码。分别对依存图和潜在图进行图卷积操作,聚合邻居节点信息,生



成句法感知与语义感知的特征表示。

(3) 正则化约束。通过异质性正则项强制潜在图与依存图学习差异化特征,避免信息冗余。同时,利用正交正则项约束潜在图的注意力矩阵正交化,增强语义分布的均匀性。

(4) 特征融合与分类。模型拼接双图特征,通过线性层与 Softmax 分类器预测事件类型,支持多事件并行检测。

4. 基于无触发词双注意力的中文事件检测模型

传统事件检测依赖触发词识别,然而,这种基于触发词的中文事件检测方法存在两个局限。其一,触发词存在多义性。同一触发词在不同上下文中可能表示不同事件类型。例如,“释放”既可表示“攻击事件”(释放催泪弹),也可表示“释放-假释事件”(当庭释放)。仅依赖触发词本身无法准确分类,需结合上下文消除歧义。其二,触发词不匹配。中文缺乏自然分隔符,分词后触发词可能被拆分或与其他词重叠。例如,“并购”中的“并”和“购”分别触发不同事件,但分词后难以准确匹配标注的触发词,导致识别错误。

基于此,研究者提出了一种基于无触发词双重注意力的中文事件检测模型(Chinese Event Detection Without Triggers based on Dual Attention, EDWTDA)^[11]。该方法的核心思想是跳过触发词识别步骤,通过双重注意力机制直接确定事件类型。核心模块为双重注意力机制,具体如下。

(1) 局部注意力机制。其目的是捕获句子中的关键词,模拟隐藏触发词,缓解词匹配问题。主要通过 Bi-LSTM 提取句子特征,结合事件类型嵌入向量计算词级注意力权重,权重高的词被视为隐式触发词。

(2) 全局注意力机制。其目的是利用文档上下文解决多义词歧义问题。主要通过结合句子特征、事件类型嵌入和文档嵌入,计算全局注意力权重,捕捉跨句子的语义关联。

随后,针对传统多标签分类任务复杂度高且样本分布不均衡的问题。该模型将事件检测转化为二进制分类任务,每个句子对每个事件类型单独预测其是否存在。例如,句子“地震发生”对“地震”事件类型输出 1,对其他类型输出 0。最后,引入焦点损失,动态调整易分类样本和难分类样本的权重,缓解长尾分布问题。

本节梳理了事件检测领域的几大创新方法,涵盖动态关系建模、增量学习、双图协同与无触发词检测。这些方法突破了传统范式的局限性,为复杂场景下的事件检测提供了新思路。

3.2.2 事件论元抽取技术

3.2.1 节梳理了事件检测技术的演化路线与创新方法。研究者通过深度融合句法结构、动态语义编码及多模态图网络,显著提升了触发词识别与事件分类的精度。然而,事件抽取的完整语义解析不仅需要精准定位事件类型,还需进一步提取事件的核心参与者(论元)及其语义角色。这一任务即事件论元抽取(Event Argument Extraction, EAE),其目标是从触发词所在的局部上下文中识别出充当事件参与者、时间、地点等角色的论元,并为其标注预定义的角色(如施事者、受事者、工具等),从而构建结构化的事件框架。

与事件检测不同,论元抽取面临更复杂的语义解析挑战:

(1) 事件中存在隐式的论元实例(指由指代词替代的论元),需通过上下文推理或句法

分析进行实际论元的识别。

(2) 同一实体在同一事件或不同事件中充当多重角色,需对实体与多重角色之间的动态语义关联建模。

(3) 在开放域场景下,论元角色缺乏预定义本体,需动态推断语义关联。

针对这些问题,早期研究多依赖规则或统计模型。随着深度学习技术的发展,论元抽取逐步转向结构增强与语义深化的双轨范式。基于动态记忆网络(DMN)^[7]的模型通过多次迭代注意力机制,强化论元与触发词的上下文关联;而 Nguyen 等(2018)^[12]考虑利用句法依存中的图结构信息,有效地更新实体的卷积信息,从而设计了以论元为中心的图卷积池化网络。此外,也有研究者引入外部知识或深层语义推理,如使用预训练语言模型(BERT、RoBERTa 等),通过掩码语言建模策略,隐式地学习角色的逻辑分布,或者结合问答框架将论元抽取转换为对特定语义槽的填充任务^[13~16],有效捕捉上下文中触发词与论元间的长距离语义关联,并基于预设问题模板识别事件中的隐式论元。此外,部分研究探索了基于强化学习的策略,通过模拟论元间的交互关系优化角色分配策略^[17]。

然而,现有方法仍存在角色概念割裂(如“卖家”与“买家”共享“人物”属性但未被联合建模)、多模态信息利用不足(句法、语义结构分离)、低资源泛化瓶颈(标注数据稀缺)等局限。针对这些问题,本节将聚焦 3 种创新方案,为复杂场景下的论元抽取提供解决方案。

1. 分层模块化的事件论元抽取模型

先前的工作通常对每个候选论元进行独立的角色分类,忽略了不同角色间的概念关联性。例如,“Seller”(卖家)和“Buyer”(买家)同属“Person”(人物)或“Org”(组织)等高级概念,共享语义特征,这种关联性可以为角色分类提供有效的归纳偏置,但传统方法未能充分利用它,导致模型难以捕捉角色间的隐含关联。基于此,Wang 等(2019)^[18]提出了分层模块化的事件论元抽取(Hierarchical Modular Event Argument Extraction, HMEAE)模型。该方法通过构建层级模块化网络,将论元角色的概念层级结构转化为神经网络的可学习模块,以增强论元角色分类的语义推理能力。具体而言,该模型将事件论元抽取任务分解为 3 个模块,分别处理文本编码、语义聚焦与角色分类。下面简要介绍这 3 个模块。

(1) 实例编码器。该模块首先基于两种不同的模型变体,分别使用 CNN 和 BERT 编码器,将句子中的每个词转换为隐藏向量;然后以触发词和候选论元的位置为锚点,将句子划分为 3 个区域,分别为触发词前的上下文片段、触发词与候选论元之间的中间片段以及候选论元的后续上下文片段。随后,对每个区域进行最大池化,提取各区域的显著特征,最终拼接成一个全局的实例嵌入。这一操作能有效捕捉触发词与论元的局部关联信息,同时保留全局语义。

(2) 分层模块注意力模块。该模块基于概念层次结构,聚焦于与目标论元角色相关的语义区域。具体来说,首先为每个高级概念(如“Person”“Time”)设计一个独立模块。每个模块包含可训练的参数向量,表示高级概念的抽象语义(如“Time”模块的参数向量用于捕捉“2023 年”“昨天”等时间相关特征)。随后,通过多层感知机计算每个词与每个高级概念的关联分数。例如,“Time”模块会对“2023 年”“昨天”等时间词赋予高分。在此之后,针对具体论元角色(如“Seller”),整合其所属的多个高级概念(如“Person”和“Org”),将多个概念的注意力分数取平均值,生成综合注意力分数,并根据综合注意力分数对词向量加权求和,得到一个聚焦于目标角色的嵌入表示。例如,“Seller”的嵌入表示会融合“Person”与



“Org”相关的上下文特征。

(3) 论元角色分类器。得到目标角色的嵌入表示之后,该模型将实例编码器生成的全局嵌入(实例嵌入)与分层注意力聚焦的角色嵌入拼接,形成最终特征向量。通过 Softmax 层计算每个角色的概率。例如,模型可能判定“Steve Jobs”在“出售”事件中充分“Seller”角色,因其语义特征与“Person”“Org”高度匹配。

2. 基于句法与语义结构的图变换网络

在事件论元抽取任务上,先前的方法虽然能够利用句法结构(如依存树)对词语间依赖关系建模,但存在两个关键缺陷:

(1) 语义结构缺失。句法结构仅反映语法关系,无法捕捉词语间的语义关联(如“家族成员”与“女婿”的语义相似性)。

(2) 长距离依赖处理不足。当论元候选与触发词距离较远时,仅依赖句法路径难以建立有效关联(例如触发词“杀害”与候选实体“侯赛因·莫勒姆”通过中间实体“巴希尔家族成员”间接关联)。

基于此,Veyseh 等(2020)^[19]提出了一个基于句法与语义结构的图变换网络(Graph Transformer Networks with Syntactic and Semantic Structures, SemSynGTN)。该方法同时对句法与语义结构建模,融合双结构语义,通过图变换网络动态整合多跳路径信息,最后引入互信息约束,迫使模型仅保留与任务相关的关键信息。

具体而言,该模型分为四大任务模块,以下是各模块的简要描述:

(1) 语句编码模块。在该模块中,模型用一个组合嵌入表示对每个单词进行编码,该嵌入表示由词嵌入(取词的首个子词在 BERT 中最后一层的隐藏向量)与相对距离嵌入(候选论元和触发词到当前词的位置距离,随机初始化并在训练中更新)拼接,将所有词的嵌入表示拼接为词序列后输入 Bi-LSTM 捕获语句序列语义,生成词级隐藏表示,作为后续结构生成的基础。

(2) 结构生成模块。该模块生成两种关键结构。①句法结构。基于通用依存解析生成二元邻接矩阵,计算候选论元与触发词在依存树中的路径长度,生成有针对性的句法权重矩阵;②语义结构。基于自注意力机制,计算词对间的语义关联分数,并通过触发词与候选论元的隐藏向量,动态调整注意力权重,增强任务相关性。

(3) 结构组合模块。将句法、语义等初始结构分多个通道输入图变换网络(GTN),通过加权求和生成中间结构;对中间结构进行矩阵相乘,捕捉多跳路径(如“触发词→句法邻居→语义邻居→候选论元”),生成最终结构;将每个通道的最终结构作为邻接矩阵输入 GCN,通过多层图卷积逐层聚合邻域信息。

(4) 信息瓶颈正则化模块。设计一个判别器区分真实样本(GTN 输出与原始句子表示的联合分布)和噪声样本(GTN 输出与其他句子随机组合的边际分布),通过最小化判别器的对抗损失来近似估计互信息,并将其作为正则化项加入总损失函数,迫使 GTN 仅保留与事件论元抽取任务相关的关键信息,同时丢弃输入句子中的冗余或无关节节。

3. 基于双重问答的半监督事件论元抽取框架

在低资源场景下,存在事件论元抽取的数据稀疏与模型泛化两个难题。传统的方法主要针对两个核心问题。其一,数据缺陷。传统的监督方法依赖大量标注数据,但实际场景中

多数事件类型标注样本极少(如 ACE2005 中 60%事件类型样本不足 100 个),并且数据增强方法需依赖外部资源(如知识库),构建成本高且覆盖不全。其二,模型缺陷。先前的方法缺乏角色语义的显式利用,而且要为不同角色设计独立分类器或标签,难以适应稀疏数据。

基于此,研究者提出了基于双重问答的半监督事件论元抽取框架(Semi-Supervised Event Argument Extraction via Dual Question Answering, DualQA)^[16],通过问答范式实现参数共享与语义利用,并利用未标注数据缓解标注不足问题。具体来说,DualQA 将论元抽取任务分解为两个互逆的问答子任务。

(1) 事件论元识别(EAR)。回答“什么充当角色”,通过问题生成引导模型从文本中提取对应角色的论元。

(2) 事件角色识别(ERR)。回答“角色是什么”,根据论元候选反推其角色标签。EAR 与 ERR 模型共享编码器和注意力模块,通过半监督训练相互验证伪标签,减少错误传播。例如,EAR 抽取的论元由 ERR 验证其角色,反之亦然,仅保留一致结果作为新标注数据加入训练集,逐步提升模型性能,直至未标注数据耗尽或模型收敛。

以下是各模块的具体描述。

(1) 问题生成与实例编码模块。首先,EAR 根据事件类型和角色生成问题,ERR 根据论元候选生成问题。在生成两个互逆的问题后,将事件上下文(句子、触发词、事件类型)拼接为结构化输入,使用预训练 BERT 模型对上下文和生成的问题分别编码,得到词语的向量表示。

(2) 流注意力机制模块。构建双向注意力交互(Context-to-Question、Question-to-Context),计算每个上下文词与问题词的相关性,突出与问题相关的内容。例如,“伊拉克军队”与问题中的“攻击者”关联性更高。同时,反向计算问题词对上下文的重要性,识别关键上下文片段(如触发词附近的动词),将双向注意力结果与原始编码向量结合,生成“问题感知”的上下文表示,编码论元与角色的语义关联。

(3) 论元分类(EAR)与角色分类(ERR)模块。构建论元分类器与角色分类器。前者基于流注意力输出的上下文表示,预测论元的起始和结束位置;后者对候选论元的上下文表示进行卷积操作,提取局部特征,预测候选论元属于各角色的概率。

(4) 半监督双重训练策略模块。在联合训练阶段,该策略交替优化 EAR 和 ERR 模型,通过最大化两者对标注数据的对数似然概率,促进任务间的协同学习。随后,针对未标注数据,EAR 根据事件模式生成候选论元,ERR 验证其角色一致性;ERR 预测候选论元的角色后,EAR 反向验证论元是否存在,保留双向验证一致的高置信度结果作为新增标注数据。通过迭代训练与数据增强,逐步扩充高质量标注集,直至模型收敛或未标注数据耗尽,最终在低资源场景下显著提升模型准确性。

本节梳理了语句级事件论元抽取的挑战与技术路线,针对事件中存在的隐式论元实例、多角色重叠及数据不足等复杂场景下的事件论元抽取探讨了新思路。

3.2.3 事件识别及其论元抽取技术

事件识别及其论元抽取作为事件抽取的核心任务,旨在从文本中检测出事件并抽取对应论元。根据任务实现范式,现有研究主要分为流水线模式与联合抽取模式两类技术路线。在早期研究中,流水线模式主导事件抽取流程——先通过特征工程整合词法、句法和实体信



息,如 Li 等(2015)^[2]利用 AMR 技术生成深层语义表示,再独立进行触发词检测与论元抽取。然而,这种分阶段处理技术受限于错误传播问题,尽管 CNN 模型缓解了特征工程的误差累积,如 Nguyen 等(2015)^[3]通过 CNN 自动学习隐藏特征,但流水线模式下论元抽取仍依赖前序触发词检测的准确性,导致模型全局效果不佳。

随着深度学习技术的发展,联合抽取模式逐渐兴起。Nguyen 等(2016)^[20]首次将 RNN 引入联合抽取,通过共享语义表示同时处理触发词检测与论元分类,开了端到端建模的先河。这一思路解决了流水线模式的模块割裂问题。此后,研究者进一步发现论元间关联的重要性:Sha 等(2016)^[21]通过正则化方法平衡论元模式,Liu 等(2017)^[6]利用注意力机制显式融入论元信息优化事件检测,这些探索为联合模型的发展奠定了基础。

在模型架构创新方面,句法依存信息与神经网络的结合成为关键突破点。Sha 等(2018)^[22]将句法依存结构嵌入 Bi-LSTM,通过对序列与依存双结构的编码提升论元角色识别效果;Nguyen 等(2018)^[12]借助 GCN 设计论元中心的池化网络,利用句法图结构显式对触发词与论元的依赖关系建模,推动事件抽取从序列建模向图结构的升级。生成对抗技术与注意力机制的引入进一步丰富了模型能力:Hong 等(2018)^[23]通过生成对抗网络(GAN)生成虚假特征并过滤噪声;Liu 等(2018)^[7]设计了动态记忆网络以多次利用论元上下文信息;Yan 等(2019)^[24]在 GCN 中融入多跳注意力,扩展了语义关联范围,逐步形成了以深度学习为核心的技术体系。

基于此,下面从 CNN、RNN、GAN、GCN 四大技术路线出发,解析深度学习在事件识别与论元抽取中的核心创新与技术路径。

1. 基于 CNN 的事件抽取方法

基于 CNN 的事件抽取模型主要通过局部特征提取和池化操作,捕获事件触发词与论元的语义关联,典型模型如下:

(1) DMCNN(动态多池卷积神经网络)^[25]。该方法提供了动态多池化机制,根据触发词的位置动态调整池化窗口,以同时捕获词汇级特征(如词嵌入)和句子级特征(如事件类型)。此外,在论元分类阶段,通过不同池化层分别处理触发词左侧、右侧及自身位置的上下文,生成多粒度特征表示,最后通过全连接层将论元角色分类。

(2) NC-CNN(非连续卷积网络)^[4]。引入非连续 n -gram 卷积,解决传统 CNN 只能捕捉连续局部特征的缺陷。具体来说,模型将候选触发词的上下文视为所有可能的非连续 n -gram 组合(即从上下文中任意选取 n 个不连续的词),通过位置嵌入编码触发词与上下文词的相对距离,并利用动态规划计算所有非连续 n -gram 的卷积得分,再通过最大池化筛选出对事件检测最关键的非连续语义关联(如“爆炸-导致-伤亡”中的远距离依赖)。

(3) HMEAE(分层模块化论元抽取)^[18]。通过模块化设计,分层处理论元角色的概念相关性(如“攻击者”与“受害者”的语义对立)。每个模块独立学习特定论元角色的表示,通过门控机制动态融合模块输出,预测所有论元。

2. 基于 RNN 的事件抽取方法

RNN 模型擅长对序列依赖建模,为事件触发词与论元的联合建模提供了一定的技术支持,典型模型如下:

(1) JRNN(联合递归神经网络)^[26]。该方法通过 Bi-GRU 对文本进行全局上下文编

码,生成融合前后文信息的隐藏状态表示,并采用联合解码框架同步预测事件触发词与论元角色,同时引入记忆机制显式对两者之间的依赖关系建模,例如记录已检测触发词类型及实体在不同事件中的角色关联,从而直接优化触发词检测与论元分类的联合目标函数,避免分阶段处理的误差传递问题。

(2) dbRNN(依赖桥 RNN)^[22]。该方法在 Bi-LSTM 的隐藏层中引入依赖桥(一种可训练的句法连接机制),将句法依存关系(如主谓、动宾关系)通过可学习的权重动态融合到当前词的编码过程中,同时利用可学习的参数和非线性变换,显式对论元间交互(如“攻击者-武器-受害者”三元组)建模,增强了事件触发词与论元的语义关联。

(3) BDLSTM-TNN(双向动态多池 LSTM 张量网络)^[27]。该方法首先通过 Bi-LSTM 生成每个词的上下文感知表示;随后根据触发词位置动态选择池化区域,提取关键特征;最后通过三维张量矩阵运算求出每对候选论元之间的语义关联分数,并分析候选论元间的成对关系(如“攻击者-受害者”的对抗关系),实现事件触发词检测与论元分类的联合优化。

3. 基于 GAN 的事件抽取方法

随着生成对抗技术的不断成熟,该技术逐渐被用于解决事件抽取问题,GAN 是基于神经网络的生成模型,主要通过生成器与判别器的对抗训练增强模型鲁棒性,典型模型如下:

(1) SELF(自调节学习)^[28]。该方法利用生成器基于 Bi-LSTM 生成混淆特征(如错误触发词),同时利用判别器结合真实标注优化分类边界,区分真实与虚假样本,从而减少噪声干扰。

(2) AEM(对抗神经事件模型)^[29]。该方法用狄利克雷分布建模事件潜在结构,利用生成器从隐变量生成事件论元分布,通过判别器区分生成事件与真实事件的差异,优化生成过程,通过对抗训练,模型能显式学习事件的隐含结构(如触发词与论元的跨句子关联)。此外,该方法允许文档包含多事件混合分布,借助神经网络非线性建模能力适配长文本中复杂的多事件场景。

4. 基于 GCN 的事件抽取方法

GCN 能够依靠图谱理论在拓扑图上完成卷积操作,在事件抽取中能够显式利用句法或语义图结构对事件关系建模,典型模型如下:

(1) JMEE(多事件联合抽取框架)^[30]。该方法构建句法依赖图,添加捷径弧(shortcut arc)连接远距离节点,利用注意力机制对邻居节点信息进行动态加权,通过图注意力卷积联合抽取多个事件,解决事件重叠和依赖问题。

(2) MOGANED(多阶图注意力事件检测)^[24]。该方法利用一阶和高阶句法图对候选触发词的多阶表示建模,通过多阶句法图捕捉触发词与上下文的复杂关系,同时利用图注意力机制筛选重要邻接节点。

(3) EE-GCN(边增强图卷积网络)^[8]。该方法显式地对句法关系标签(如主谓、动宾、修饰等)建模,通过关系感知的图卷积操作,增强节点(词语)间的语义关联,从而更精准地捕捉事件触发词与论元的依赖关系。

由此可见,基于深度学习的事件抽取技术通过 CNN、RNN、GAN 和 GCN 等架构,在触发词检测、论元角色分类及事件依赖建模方面取得了显著进展。这些方法通过端到端语义编码、动态特征融合和图结构推理,在一定程度上解决了多义触发词消歧、长距离依赖捕获

及事件交互等传统难题。

3.2.4 小结

针对本节介绍的语句级事件抽取方法,图 3.1 从采用的特征、技术和解决的问题 3 方面对本节大部分文献进行了对比和总结。

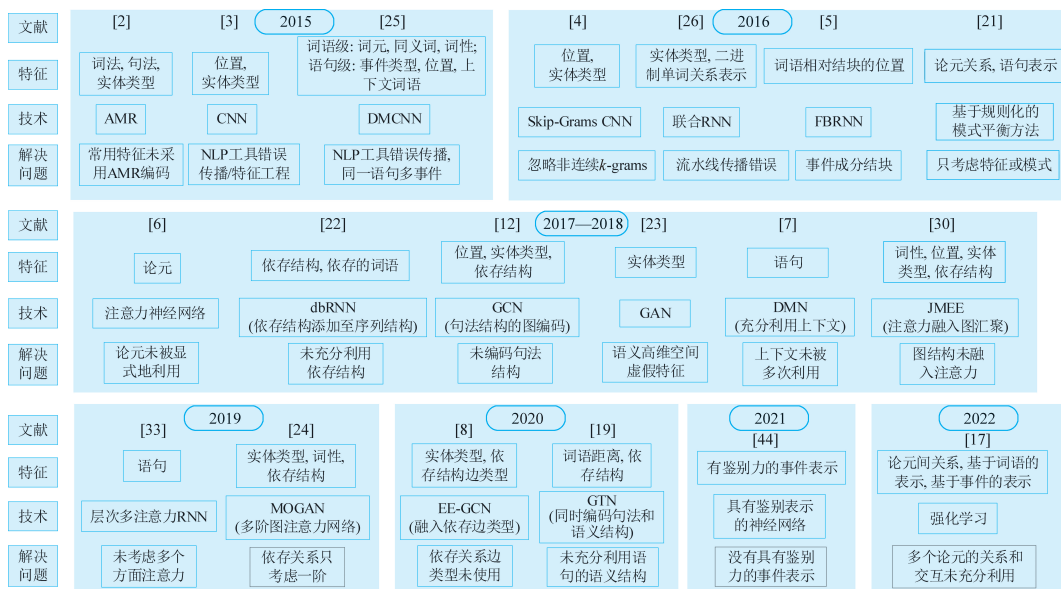


图 3.1 语句级事件抽取的主要发展历程

(1) 特征。常用的特征主要包括词法、句法、词语位置和实体类型等,在 2015—2020 年都有学者采用,只是配合使用的模型有所不同。早期,词法、词语位置和实体类型等浅显、易发现的特征被广泛使用,配合不同的模型均取得了不错的效果。随后,借助句法依存结构及相关信息变得普遍,包括使用图神经网络模型直接对该结构编码,将该结构添加至序列结构中,利用论元之间的依存关系或者论元与触发词之间的依存关系以及依存关系的边类型,等等,即使部分研究没有直接使用依存结构信息,但其采用的模型还是以该结构为基础而发展的。因此,从目前看,句法依存信息是事件抽取的一个核心特征。

(2) 技术:模型在不断地优化。在不同年份,基于大体相同的模型框架的研究很少,说明在模型相同的情况下,针对封闭域语句级事件抽取,稍微改变输入特征不足以较好地提升抽取效果。当然,在模型优化的过程中,也加入了深层的有效特征信息,而不再采用常见的特征。

(3) 解决的问题。在已经细化的语句级事件抽取任务下,要解决的问题主要是现有模型的局限以及某些线索未被考虑或者未被充分利用。随着时间的推移,相关研究在解决问题时逐渐偏向挖掘深层的线索(如依存关系类型、论元彼此的关系、更具区分能力的事件表示)以及采用先进的模型(如图神经网络、图转移网络、强化学习等)。

综上所述,可以看出句法依存信息已成为连接模型创新的核心载体。然而,现有方法仍面临依存结构动态建模能力不足的问题。此外,复杂场景的事件抽取也成为重心,领域适配问题(领域特定缺失成分处理能力薄弱)与开放场景泛化问题(如无预设本体的事件抽取)也

成为新的技术前沿。为此,3.3节将围绕解决相关问题的技术方法展开专门讨论,探讨复杂场景下语义解析的创新解决方案。

3.3 依存关系增强的事件检测技术

3.3.1 研究动机

3.2.1节梳理了事件检测技术的演进脉络,从中可以看出,无论是早期基于特征工程的统计模型,还是近期融合句法依存与图神经网络的深度学习方法,句法依存信息始终起到重要的作用。然而,传统方法多将依存句法视为静态连接路径,或仅依赖依存类型调整注意力权重,未能充分捕捉依存关系之间的结构化交互与方向性语义差异。此外,句法依存作为事件要素关联的核心载体,其隐含的角色信息对事件类型判定至关重要。因此,本节聚焦一个依存关系增强的事件检测方法^[31],解析其如何通过双向图结构建模与层级调整策略,突破传统方法对依存信息的浅层利用。该方法主要针对先前研究存在的3个核心局限:

(1) 依存关系结构建模的缺失。传统方法仅利用依存树中的词语节点关联,而忽略依存边之间的结构关系(如共享同一词语的多条依存边)。这种简化导致依存关系间的语义交互(如“主语—动词—宾语”链式结构)未被充分编码。

(2) 方向性语义的模糊性。依存边的方向性(如依存关系“nsubj”指向触发词,“dobj”由触发词指向宾语)蕴含关键角色信息,但现有模型(如GAT、GCN)多将边权重计算静态化,未结合方向动态调整语义汇聚。

(3) 核心论元的弱化。事件类型判定高度依赖触发词与核心论元(如施事者、受事者)的协同作用,但传统图结构未显式提升这些论元的层级,导致其语义贡献被次要成分稀释。

具体来说,图3.2(a)展示了语句“a bomb will go off”的词语依存树。其中,“go”和“off”为官方标注的触发词,“det”“nsubj”“aux”“compound:prt”为Stanford CoreNLP工具解析的依存关系(类型)。现有研究只考虑了词语依存图(TDG)中的词语依存信息,忽略了图3.3(b)中由依存关系构建的结构图,简称依存关系图(DRG)。其中,图3.2(a)中的每条依存边对应图3.2(b)中的一个节点,称为依存关系节点;如果图3.2(a)中的两条依存边共享同一个词语节点,则图3.2(b)中的依存关系节点之间构建一条连边。基于图3.2(b)所示的依存关系图可以较好地捕获依存关系之间的结构信息,如共享词语结构。

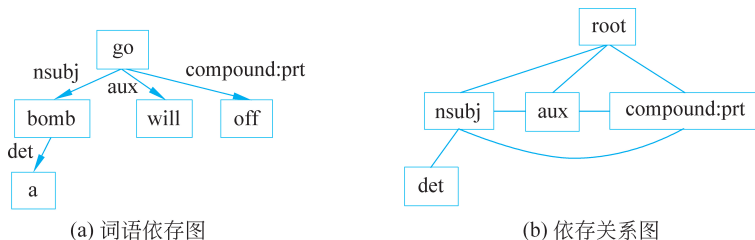


图 3.2 ACE2005 语料中“Attack”类型的示例

首先,构建依存关系图具有如下优势:

(1) 词语嵌入表示、依存关系嵌入表示可分别在不同结构的两个图上交互式地迭代更

新。例如,DRG 中的节点嵌入表示影响 TDG 中节点信息汇聚时边的权重,TDG 中的节点嵌入表示调控 DRG 中边的权重。同时,在 DRG 中依存关系嵌入表示可借助图神经网络(GNN)机制更好地编码。尽管 Cui 等(2020)在 TDG 中同时更新了词语嵌入表示和依存关系嵌入表示^[8],但依存关系嵌入表示只是简单地与词语嵌入表示拼接,然后乘以权重矩阵进行更新,这样就忽略了依存关系的结构信息。

(2) 缓解 GNN 堆叠多层导致性能急剧下降的问题(过度平滑问题)。例如,在图 3.2(a)中,“nsubj”边的信息需要经过“nsubj→bomb→go→off→compound:prt”才能传递给边“compound:prt”,中间需要借助词语作为桥梁。而依存关系图可以直接传递,很大程度减少了 GNN 堆叠的层数。

其次,在对图 3.2(a)所示的词语依存图进行编码汇聚时,对于依存边的权重计算,现有研究要么不考虑依存边的类型或者方向,要么只采用当前节点和其邻居节点的词语语义嵌入表示,而该模型的词语依存图是双向的。如图 3.3(a)所示,在对词语节点“go”的嵌入表示进行更新时,“bomb”与“go”之间的依存关系应为“nsubj”,因为“bomb”作为名词性主语的角色依存于“go”,因此,可以利用词语“bomb”和依存关系“nsubj”的嵌入表示计算“bomb”指向“go”的有向依存边的权重。此外,如图 3.3(b)所示,当对词语节点“bomb”进行更新时,“go”与“bomb”之间的依存关系不应为“nsubj”,因为“go”并不依存于“bomb”,其在词语依存树中的依存角色为 root(其与父节点之间的依存关系),这能较好地揭示“go”在语句中所充当的角色的语义,因此,利用词语“go”和依存关系 root 的嵌入表示可以计算“go”指向“bomb”的有向依存边的权重。

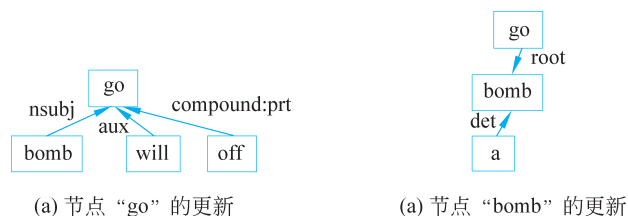


图 3.3 图 3.2(a)中 TDG 节点嵌入表示更新示例

最后,事件类型的识别不能仅仅依靠触发词,事件论元也有助于识别事件类型。如图 3.2(a)所示,触发词“go”并不能很好地判断事件类型为“Attack”,需要结合“bomb”共同决策。这是因为语句“we go up there”的触发词也为“go”,但事件类型却为“Transport”。然而,这些论元主要分布于事件的施事者和受事者中,且它们的层级低于触发词。为了强化这些核心论元,同时弱化其他价值不大的论元,应提升这些核心论元的层级。因此,上述词语依存图和依存关系图的最终形式如图 3.4 所示。

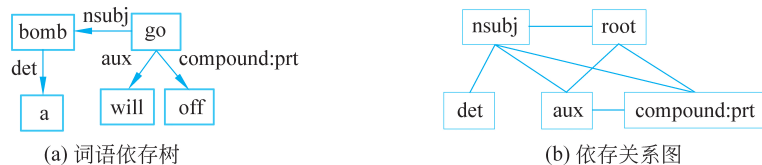


图 3.4 调整核心论元层级后的两种图结构

3.3.2 DSE-ED 模型

为了对上述思想建模,本书作者提出了一种依存关系结构加强的事件检测(Dependency Structure-Enhanced Event Detection, DSE-ED)模型,其框架如图 3.5 所示。该模型的实现过程围绕其三大核心贡献展开——**依存关系图构建**、**双向注意力机制设计**、**核心论元层级调整**。以下结合具体步骤详细阐述其实现逻辑。

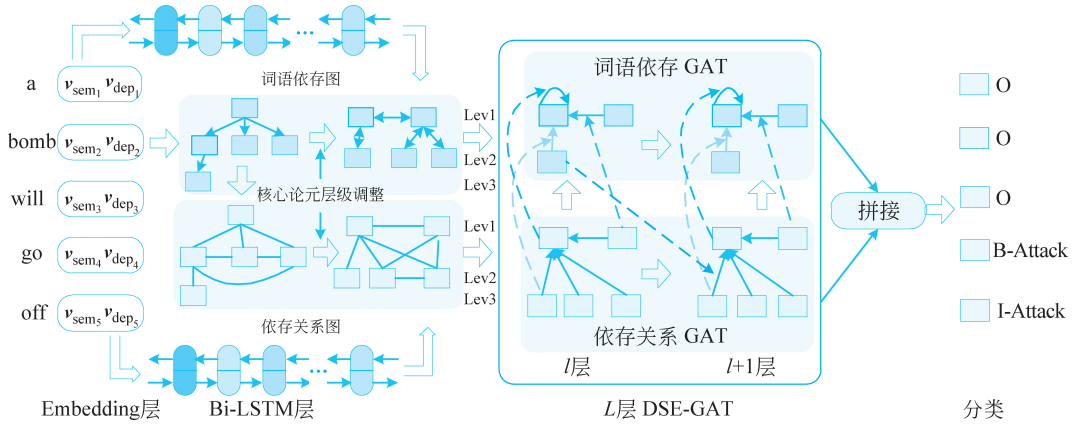


图 3.5 依存关系结构加强的事件检测模型框架

1. 输入处理与图结构构建

(1) Embedding 层。给定语句 $s = \{w_1, w_2, \dots, w_n\}$, 该模型采用已训练的词向量模型对输入文档中的词语进行语义向量化。其中, n 为语句长度, w_i 为语句中的第 i 个词语, 假设其对应的词向量为 v_i 。在词语依存树中, 每个节点与其父节点之间的依存关系作为该节点的依存关系, 虚拟的根节点 (root) 删除。依存关系的向量表示在随机初始化生成的向量表中查找, 记为 \mathbf{dep}_i 。

(2) Bi-LSTM 层。对于每个词语的语义嵌入表示 v_i , 首先利用一个前向 LSTM 对 w_1 至 w_i 的上下文信息 \vec{h}_{sem_i} 进行编码, 然后再采用一个后向 LSTM 对 w_n 至 w_i 的上下文信息 $\overleftarrow{h}_{\text{sem}_i}$ 进行编码, 具体如式 (3.3) 和式 (3.4) 所示:

$$\vec{h}_{\text{sem}_i} = \overrightarrow{\text{LSTM}}(v_1, v_2, \dots, v_i) \quad (3.3)$$

$$\overleftarrow{h}_{\text{sem}_i} = \overleftarrow{\text{LSTM}}(v_i, v_{i+1}, \dots, v_n) \quad (3.4)$$

最后将 \vec{h}_{sem_i} 和 $\overleftarrow{h}_{\text{sem}_i}$ 进行拼接, 作为词语 w_i 经过 Bi-LSTM 层后的语义嵌入表示, 记为 $\mathbf{h}_{\text{sem}_i} = \vec{h}_{\text{sem}_i} \parallel \overleftarrow{h}_{\text{sem}_i}$ 。其中, \parallel 表示拼接操作。同样, 依存关系的初始向量也经过一个 Bi-LSTM 层, 使得依存关系能够捕获一定的序列语义, \mathbf{dep}_i 输出的嵌入表示记为 $\mathbf{h}_{\text{dep}_i} = \vec{h}_{\text{dep}_i} \parallel \overleftarrow{h}_{\text{dep}_i}$ 。

(3) 词语依存图生成。对输入语句进行句法依存解析, 生成 TDG。将 TDG 中的每条依存边 (如 “nsubj” “dobj”) 转化为 **双向边**, 形成有向图结构 [图 3.2(a) → 图 3.4(a)]。例如, 依存关系 “nsubj” 从触发词 “go” 指向主语 “bomb”, 而反向边则隐含 “bomb” 作为主语依存于 “go” 的语义。



(4) 依存关系图构建。将 TDG 中的每条依存边升格为独立节点(称为依存关系节点),若多条依存边共享同一词语节点,则在 DRG 中建立这些依存关系节点间的连边[图 3.2(b)]。例如,若“nsubj”和“dobj”均连接触发词“go”,则这两个依存关系节点在 DRG 中直接相连,捕获依存关系间的交互语义。

传统方法仅将依存边作为连接词语的路径,忽略其结构化语义。因此,通过 DRG 显式地对依存边之间的关联(如共享触发词的多条依存边)进行编码,解决传统方法对依存关系结构建模的缺失。同时,DRG 允许依存关系节点直接交互,避免传统 GNN 需通过词语节点绕路传递信息的问题,缓解多层堆叠导致的过度平滑[对比图 3.2(b)与图 3.4(b)]。

2. 核心论元层级调整

(1) 层级提升规则。根据依存关系类型(如“nsubj”“dobj”),识别核心论元节点(如主语“bomb”、宾语“off”),将其在图中的层级从默认的较低层级(如层级 2)提升至与父节点同级(层级 1)。例如,在图 3.4(a)中,主语“bomb”的层级提升后,其嵌入表示在注意力汇聚时优先影响触发词“go”。

(2) 并列结构处理。针对并列依存关系(如“conj”),提升相关节点层级,确保并列成分(如“bomb”与“gun”在并列主语中)的语义权重均衡。

核心论元(如主语、宾语)的语义贡献在传统图结构中被稀释。因此,通过显式调整层级,确保核心论元在注意力计算中占据更高权重,直接增强其对事件类型判定的贡献。在嵌套句或长距离依赖场景中,层级调整策略可更加精准地定位关键论元,减轻噪声干扰。

3. 二元图注意力网络

(1) 依存关系图注意力网络。对于原始图注意力网络 $GAT^{[32]}$,其模型输入层中节点嵌入表示为 $\mathbf{H} = \{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n\}$, $\mathbf{h}_i \in \mathbb{R}^D$, n 为节点数量, D 为每个节点词语的词向量长度;本层的输出为 $\mathbf{H}' = \{\mathbf{h}'_1, \mathbf{h}'_2, \dots, \mathbf{h}'_n\}$, $\mathbf{h}'_i \in \mathbb{R}^{D'}$, D' 为节点嵌入表示的向量维度。节点 i 和节点 j 之间边的权重(注意力系数)如式(3.5)所示:

$$e_{ij} = a(\mathbf{W}\mathbf{h}_i, \mathbf{W}\mathbf{h}_j) \quad (3.5)$$

其中, $\mathbf{W} \in \mathbb{R}^{D' \times D}$ 为变换矩阵, a 为一个共享的注意力机制: $\mathbb{R}^{D'} \times \mathbb{R}^{D'} \rightarrow \mathbb{R}$, 是一个单层的前馈神经网络。采用 Softmax 标准化的表示如式(3.6)所示:

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\boldsymbol{\phi}^T[\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_j]))}{\sum_{k \in V_i} \exp(\text{LeakyReLU}(\boldsymbol{\phi}^T[\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_k]))} \quad (3.6)$$

其中, V_i 为与 i 点直接相连的节点集合, $\boldsymbol{\phi}$ 为权重向量, LeakyReLU 为激活函数, $(\cdot)^T$ 为转置操作。节点 i 当前层的输出如式(3.7)所示:

$$\mathbf{h}'_i = \sigma \left(\frac{1}{M} \sum_{m=1}^M \sum_{j \in V_i} \alpha_{ij}^m \mathbf{W}^m \mathbf{h}_j \right) \quad (3.7)$$

其中, M 为注意力的头数, σ 为非线性函数。

由于依存关系图中的连边表示共享相同词语,不同方向含义相同,于是该模型联合当前依存节点的嵌入表示与对应词语的嵌入表示共同调控依存关系图中边的权重。基于原始图注意力方法,依存关系图注意力网络中添加节点层级后第 l 层的注意力系数和依存关系的嵌入表示分别如式(3.8)和式(3.9)所示:

$$\gamma_{ij}^{(l)} = \frac{\exp(\text{LeakyReLU}(\boldsymbol{\varphi}^T \mathbf{W}^{(l)} \mathbf{h}_{\text{sem}_j}^{(l-1)} (\mathbf{h}_{\text{dep}_i}^{(l-1)})^T))}{\sum_{k \in V_i} \exp(\text{LeakyReLU}(\boldsymbol{\varphi}^T \mathbf{W}^{(l)} \mathbf{h}_{\text{sem}_k}^{(l-1)} (\mathbf{h}_{\text{dep}_i}^{(l-1)})^T))} \quad (3.8)$$

$$\mathbf{h}_{\text{dep}_i}^{(l)} = \sigma \left(\frac{1}{M} \sum_{m=1}^M \sum_{j \in V_i} u_j \gamma_{ij}^{m(l)} \mathbf{W}^{m(l)} \mathbf{h}_{\text{dep}_j}^{(l-1)} \right) \quad (3.9)$$

其中, $\mathbf{W}^{(l)} \in \mathbb{R}^{F \times D}$ 为权重矩阵, $\mathbf{h}_{\text{sem}_j}^{(l-1)} \in \mathbb{R}^D$ 为词语关系图中对应词语节点 $l-1$ 层输出的最新嵌入表示, $\mathbf{h}_{\text{dep}_i}^{(l-1)} \in \mathbb{R}^F$, F 和 D 分别表示依存关系和词语的嵌入表示维度, $u_j = \exp(-\text{lev}_j)$, lev_j 为自然数, 表示节点 j 的层级, u_j 为其标准化后的层级值, 范围为 $(0, 1]$ 。

以往的研究只关注词语依存图中词语的依存信息(词语间的依存关系及其类型), 忽略了依存关系本身的结构信息, 依存关系图注意力需要捕获依存关系之间的结构, 可以获取更多的语义表示。

(2) 词语依存图注意力网络。为了更好地体现不同依存关系的重要程度, 该模型采用当前词语节点的嵌入表示和其邻居节点的依存关系嵌入表示联合调控词语依存图中依存边的权重, 区分边的不同方向。第 l 层的注意力系数如式(3.10)所示:

$$\beta_{ij}^{(l)} = \frac{\exp(\text{LeakyReLU}(\mathbf{W}^{(l)} \mathbf{h}_{\text{dep}_j}^{(l)} (\mathbf{h}_{\text{sem}_i}^{(l-1)})^T))}{\sum_{k \in V_i} \exp(\text{LeakyReLU}(\mathbf{W}^{(l)} \mathbf{h}_{\text{dep}_k}^{(l)} (\mathbf{h}_{\text{sem}_i}^{(l-1)})^T))} \quad (3.10)$$

其中, $\mathbf{W}^{(l)} \in \mathbb{R}^{D \times F}$ 为权重矩阵, $\mathbf{h}_{\text{dep}_j}^{(l)} \in \mathbb{R}^F$ 为依存关系图中对应依存节点 l 层的最新嵌入表示, $\mathbf{h}_{\text{sem}_i}^{(l-1)} \in \mathbb{R}^D$ 。对应的词语嵌入表示如式(3.11)所示:

$$\mathbf{h}_{\text{sem}_i}^{(l)} = \sigma \left(\frac{1}{M} \sum_{m=1}^M \sum_{j \in V_i} u_j \beta_{ij}^{m(l)} \mathbf{W}^{m(l)} \mathbf{h}_{\text{sem}_j}^{(l-1)} \right) \quad (3.11)$$

上述策略不仅融入了最新的依存关系嵌入表示, 动态地区分了不同的依存关系, 同时还利用了不同方向依存边的结构语义。以往的研究只利用了词语的语义信息, 或加上静态的依存关系信息, 且不区分不同方向的依存边的结构语义, 它们都依靠模型学习依存边的权重, 没有显式地反映不同方向依存边所充当的角色的语义差别。

此后, 该模型将上述两个图结构经 L 层 GAT 产生的嵌入表示进行拼接, 如式(3.12)所示, 得到词语 w_i 的最终嵌入表示。

$$\mathbf{h}_i^{(L)} = \mathbf{h}_{\text{sem}_i}^{(L)} \parallel \mathbf{h}_{\text{dep}_i}^{(L)} \quad (3.12)$$

最后, 该模型将词语的最终嵌入表示输入全连接层, 并采用 Softmax 函数计算其在所有事件类型标签上的概率分布, 计算每个词语的事件类型概率, 并采用加权交叉熵损失优化模型。

3.3.3 实验

1. 数据集与评测标准

DSE-ED 模型选择广泛使用的 ACE2005 英文和中文语料库作为实验数据集。ACE2005 标注了 8 个事件类型和 33 个事件子类型, 英文/中文语料包含 599/633 个文档和 5357/3333 个事件。由于 ACE2005 没有划分训练集和测试集, 且样本数量较少, 该模型使用 StratifiedKFold 函数对整个语料库进行随机划分并作交叉验证。

关于评测标准, 与前期研究^[8, 12, 24-25]一样, 采取官方的评分策略 Precision、Recall 和 F1 对事件类型标签进行打分, 以此来评价该模型。由于“O”类型的词语较多, 计算整体的



F1 值会导致真实论元角色的识别效果虚高,因此评测时滤掉了“O”类型的预测结果。

2. 参数设置和基线

该模型选择 Adam 优化器,学习率、Dropout 值、Alpha 值、迭代次数和批大小 (batch size) 分别为 $1e-4$ 、0.2、2、40 和 1。词语语义嵌入表示维度和依存关系嵌入表示维度分别设为 768 和 20,GAT 的层数和注意力头数分别设为 2 和 3;Bi-LSTM 和全连接网络的层数设为 2,对应的隐藏层嵌入表示维度分别设为 100 和 50。训练集和测试集的批大小均为 1,交叉验证的折数设为 4。该模型的实验环境为 Python 3.7、PyTorch 1.10.0、Intel Xeon W-1290P CPU 和 NVIDIA Quadro RTX 4000 显卡。

与前期大部分研究^[8, 24-25, 33-34]一样,为了综合评价该模型,本实验对比了 ACE2005 英文语料库上大量的先进基线,可以将它们分为 3 类:基于序列的方法、基于图神经网络的方法和其他神经网络方法。

基于序列的方法如下:

- DMCNN^[25],建立动态池化卷积模型。
- JRNN^[26],采用双向循环神经网络。
- ANN-AugAtt^[6],利用标注的论元信息和注意力。
- dbRNN^[22],在 Bi-LSTM 中增加了依存信息。
- HBTNGMA^[35],设计层级和偏差标签网络。

基于图神经网络的方法如下:

- GCN-ED^[12],首次使用 GCN 检测事件。
- JMEE^[30],采用注意力机制强化 GCN 模型。
- MOGANED^[24],将词语依存图拆分成 1 阶、2 阶和 3 阶 3 个结构。
- GateGCN^[36],设计门控 GCN 过滤词语依存图中的噪声信息。
- EE-GCN^[8],借助依存关系类型,利用边增强的 GCN 模型进行事件检测。
- SA-GRCN^[37],通过剪除词语依存图中无关边并且添加潜在有用边,开发自注意力图残留卷积网络。
- AGGED^[38],通过自适应图生成技术动态构建任务特定的图结构,并利用多通道选择策略自动筛选信息通道。
- DualGAT^[39],构建包含句法和语义关系的双关系图,结合增强的关系图注意力网络对图结构进行编码,并引入上下文信息优化注意力权重。

其他神经网络方法如下:

- EKD^[40],利用丰富的开放域触发词知识来提高事件检测效果。
- S²-JDN^[41],借助词语-事件的共指频率等统计信息提高事件检测效果。
- Saliency ED^[42],借助触发词的显著性实现事件检测。
- HPNet^[43],构造分层的策略网络,联合检测事件。
- DNR^[44],设计对比学习策略和最大范围策略,使得模型可以更好地区分触发词附近的词语。
- RLIL^[17],利用基于对话的强化学习和增量学习技术实现事件检测。
- GPTEDOT^[45],利用预训练模型 GPT-2 实现事件检测。

- Seq2Seq^[46], 将事件检测转换为图解析任务。

由于上述基线并没有全部在 ACE2005 中文语料库上进行测评, 所以本实验只选择了其中的 DMCNN 和 HBTNGMA 等部分基线, 另外还使用了其他一些基线, 包括 HNN^[47]、NPN^[48]、TLNN^[49]、CAEE^[50] 和 EDM^[51]。

该模型在 ACE2005 英文和中文语料库上的实验效果如表 3.1 和表 3.2 所示。表 3.1 中所有基线的效果取自对应的原始文献。表 3.2 中的 * 表示结果取自 Ding 等 (2019) 的研究^[49], 其他结果取自对应的原始文献。表 3.2 中基线名称后加括号的 char 表示字符级, word 表示文档级。

表 3.1 在 ACE2005 英文语料库上的实验效果

基 线	F1 值/%	基 线	F1 值/%
DMCNN	69.1	AGGED	79.9
JRNN	69.3	DualGAT	82.7
ANN-AugAtt	71.7	EKD	78.6
dbRNN	71.9	S ² -JDN	79.5
HBTNGMA	73.3	Saliency ED	75.8
GCN-ED	73.1	HPNet	77.8
JMEE	73.7	DNR	81.8
MOGANED	75.7	RLIL	80.7
GateGCN	77.6	GPTEDOT	79.2
EE-GCN	77.6	Seq2Seq	82.8
SA-GRCN	78.0	DSE-ED	86.3

表 3.2 在 ACE2005 中文语料库上的实验效果

基 线	F1 值/%	基 线	F1 值/%
DMCNN(Char)*	57.8	NPN*	64.2
DMCNN(Word)*	60.2	TLNN*	67.8
HBTNGMA(Char)*	45.5	CAEE	73.1
HBTNGMA(Word)*	53.5	EDM	76.2
HNN*	63.0	DSE-ED	79.6

由表 3.1 和表 3.2 可知, 该模型的效果高于各基线, 在 ACE2005 英文和中文语料库上 F1 值分别提升了 3.5%~17.2% 和 3.4%~34.1%, 主要原因如下。

(1) 词语依存结构。对比基于图神经网络的方法和基于序列的方法可以发现, 前者的效果整体优于后者, 这说明词语依存结构蕴含着较为丰富的信息, 充分挖掘词语依存结构信息可提升事件检测效果。其中, MOGANED 和 GateGCN 模型分别通过拆分词语依存结构和过滤词语依存结构中的噪声数据取得了不错的效果。



(2) 依存关系。在采用图神经网络方法的基线中,由于 EE-GCN 和 SA-GRCN 对依存关系信息进行了编码,使得它们在该系列的方法中获取了较好的性能。其中,SA-GRCN 在 EE-GCN 的基础上动态调整了句法依存结构,所以在第 2 类对比模型中取得了最好的效果。DSE-ED 模型采用了词语依存结构和依存关系信息捕获依存关系之间的结构信息,使得其在 ACE2005 上 F1 值达到 86.3%。

(3) 外部知识或不同策略。近几年开发的神经网络方法借助了外部知识或开辟了新的检测策略,进一步推进了事件检测的发展。其中,EKD、S²-JDN 和 Saliency ED 等基线分别借助了开放域触发词知识、词语-事件的频率统计信息以及触发词显著性等信息;HPNet、DNR 和 RLIL 分别使用了分层策略网络、对比学习策略和最大范围策略、强化学习和增量学习等技术手段提升检测效果;GPTEDOT 基于预训练模型实现事件检测,Seq2Seq 将事件检测转化为图解析的模式。尽管它们采用了不同的知识或策略,然而都没有充分利用词语依存信息,所以效果还是低于 DSE-ED 模型。综上所述,本节介绍的 DSE-ED 模型对于事件检测具有优秀的性能,主要得益于以下 3 方面:①两种图结构(词语依存图和依存关系图)及其交互更新策略;②注意力系数的计算策略;③核心论元的调整。

总的来说,DSE-ED 模型的核心思想在于通过二元图注意力机制的协同构建与交互更新,深度融合词语依存结构与依存关系的语义信息。将句法依存图中的依存边升格为独立节点,构建依存关系图(DRG),直接对依存关系间的结构关联建模,同时保留传统词语依存图(TDG)的双向边以区分方向性语义。在模型更新过程中,两类图结构通过动态交互策略实现语义互补:DRG 的依存关系节点嵌入用于调控 TDG 中边的注意力权重,而 TDG 的词语节点嵌入则用于引导 DRG 的节点特征更新。这种双向交互机制不仅显式区分了依存方向性语义,还通过层级调整提升了核心论元的语义权重,最终在双图协同作用下提高了触发词检测与事件类型判定的精准性,缓解了传统模型对依存结构利用不足以及信息传递冗余的问题。

3.3.4 分析

为了分析和确定影响事件检测的具体因素,我们在 ACE2005 英文语料库上对消融、注意力系数计算策略、核心论元调整策略以及图中词语的层级等方面做了一定的实验验证。

1. 消融

为了验证二元图的价值,我们设计了基于 DSE-ED 模型的图消融实验。表 3.3 给出了相应的消融效果,其中的 w/o 表示消融。其中,TDG 和 DRG 分别表示词语依存图和依存关系图。消融相关图指不利用图结构更新节点的嵌入表示。注意,图的消融同时会使得注意力系数计算策略和调整层级等措施失效。

表 3.3 DSE-ED 模型的消融效果

方 案	F1 值/%
DSE-ED	86.3
w/o TDG	74.1
w/o DRG	77.5
w/o TDG & DRG	71.2

由表 3.3 可知,单独消融(第 2 行和第 3 行)分别导致事件检测效果出现 12.2%和 8.8% 的下降。由此可验证,两个图结构均蕴含着丰富的语义信息,对事件检测具有重要意义。对比二者的下降幅度,移除 TDG 的下降幅度更大。这主要是因为:TDG 中词语节点的嵌入表示取自预训练模型,本身已经具有较强的语义信息,而 DRG 中的节点嵌入表示由随机初始化生成,其表达的语义需要在模型训练过程中逐步学习。尽管同时移除 TDG 和 DRG(第 4 行)的检测效果只有 71.2%,但这还是超过了表 3.1 中的部分基线,这是因为这种方法使用了依存关系的信息,并将其输入 Bi-LSTM 层,获取了依存关系之间的部分结构语义。

2. 注意力系数计算策略

为了检验 DSE-ED 模型的注意力系数计算策略对事件检测效果的影响情况,还另外设计了 4 种注意力系数计算策略:DSE-ED(Non-Interaction)指采用本节的两种注意力系数计算策略,但两个图中交互的嵌入表示采用静态的(即词语依存图中的依存关系固定为 Bi-LSTM 层输出的依存关系嵌入表示,依存关系图中的词语嵌入表示固定为 Bi-LSTM 层输出的词语嵌入表示);DSE-ED(Sem)表示词语依存图采用原始 GAT 的注意力系数计算策略(即词语的嵌入表示调控边的注意力系数),依存关系图采用本节的注意力系数计算策略;DSE-ED(Dep)指依存关系图采用原始 GAT 注意力系数计算策略,词语依存图采用 DSE-ED 模型的注意力系数计算策略;DSE-ED(Original)指两个图均采用原始 GAT 的注意力系数计算策略。表 3.4 是 DSE-ED 不同注意力系数计算策略的效果。

表 3.4 DSE-ED 模型不同注意力系数计算策略的效果

注意力系数计算策略	F1 值/%
DSE-ED	86.3
DSE-ED(Non-Interaction)	83.5
DSE-ED(Sem)	82.6
DSE-ED(Dep)	83.0
DSE-ED(Original)	81.3

通过对比表 3.4 中第 1 行和第 2 行可知,本节设计的嵌入表示交互策略是有效的。当词语依存图和依存关系图分别采用对应 Bi-LSTM 层输出的静态嵌入表示时(即不采用另一个图更新的最新嵌入表示)时,抽取效果下降了 2.8%。第 3 行和第 4 行分别是两个图单向交互的情况(即只使用该模型设计的一种注意力系数计算策略)。其中,第 3 行是依存关系图交互词语嵌入表示的情况,第 4 行是词语依存图交互依存关系嵌入表示的情况。对比第 1 行、第 3 行、第 4 行可知,本节设计的两种注意力系数计算策略是有效的,且在词语依存图中交互依存关系嵌入表示比在依存关系图中交互词语嵌入表示发挥的作用更大。这进一步体现了本节构建的依存关系图的价值,即可以在依存关系图结构上更新以得到较好的依存关系嵌入表示。第 3 行的效果不如第 4 行,主要还是因为依存关系嵌入表示是随机初始化产生的,需要逐步学习语义,从而影响了其调控词语依存图中节点嵌入表示汇聚时边的权重。



3. 核心论元调整策略

为了检验 DSE-ED 模型的核心论元调整策略对事件检测效果的影响情况,本节设计了如表 3.5 所示的验证方案。其中,Sub、Obj 和 Parallel 分别表示只调整主语、宾语和并列依存关系的情况,All 表示同时调整这 3 个核心论元,No Adjustment 表示不做任何调整。

表 3.5 DSE-ED 模型不同核心论元调整策略的效果

核心论元调整策略	F1 值/%
DSE-ED(All)	86.3
DSE-ED(Sub)	85.2
DSE-ED(Obj)	82.1
DSE-ED(Parallel)	83.2
DSE-ED(No Adjustment)	78.2

从表 3.5 可见,调整 3 个核心论元对提升事件检测效果是有益的,其 F1 值优于其他方案 1.1%~8.1%。从效果对比上看,调整的重要性由高到低依次是主语、并列依存关系、宾语,说明有利于识别事件的核心论元多处于主语中;通过调整并列依存关系,提升处于其他并列分句中的核心论元也能提升事件检测的效果。

4. 层级

表 3.6 展示了 DSE-ED 模型中节点是否采用层级的效果。其中,DSE-ED(TDG + DRG)表示同时在 TDG 和 DRG 中采用层级,DSE-ED(TDG)和 DSE-ED(DRG)分别表示只有 TDG 和 DRG 中采用层级,DSE-ED(No Level)表示在 TDG 和 DRG 中均不采用层级。

表 3.6 DSE-ED 模型中节点是否采用层级的效果

层级方案	F1 值/%
DSE-ED(TDG+DRG)	86.3
DSE-ED(TDG)	85.2
DSE-ED(DRG)	85.5
DSE-ED(No Level)	82.2

对比第 1 行和第 4 行可知,层级的利用能够带来事件检测效果的提升。单独考虑在一个图中采用层级(第 2、3 行),会导致识别效果下降 1.1%和 0.8%,下降幅度比较接近,说明在两个图中采用层级发挥的作用相当。

3.3.5 小结

本节详细介绍了一个依存关系结构加强的事件检测模型(DSE-ED),该模型聚焦于依存关系增强的事件检测,针对传统方法对句法依存结构利用不足的问题,通过构建词语依存图与依存关系图的双向交互机制,对依存边的方向性语义与结构化关联进行显式编码,并通过核心论元层级调整策略强化关键论元对事件类型的决策权重,同时利用二元图注意力网