

## 结论

### 1. 本文工作总结

作为信息抽取的关键技术，事件抽取任务在智能问答、事实核查、推荐系统等多样化自然语言场景扮演着至关重要的支撑作用，且以事件抽取技术为基础构建的事件知识图谱同样提供了广阔的应用前景。因此，事件抽取是提升文本结构化知识构建与应用的自动化和智能化水平的必然要求。目前，得益于深度学习技术的蓬勃发展，现有方法利用特定神经网络架构研究事件抽取的不同阶段任务，展现了显著的性能进步。但是，面临**数据不平衡依赖**、**实体类型过度依赖**和**跨事件依赖**等已有模型架构存在的共性数据挑战，这些方法缺乏通用的解决方案。为此，本文围绕事件抽取中的共性数据依赖进行了深入研究。接下来，将总结本文的研究成果和相应创新点如下：

(1) 针对事件检测中存在的数据不平衡依赖，首先研究了其影响事件检测性能的方式，相应引入能够有效消解其导致的性能下降的句子级别识别信息。在此基础上，提出了基于分类器自适应知识蒸馏的事件检测方法。该方法在事件识别增强网络中构建句子级别识别输入信息，通过分类器自适应的知识蒸馏技术增强事件检测网络自动捕获句子级别识别信息的能力，实现事件检测性能的提升。实验评估表明，提出的方法可以有效消解数据不平衡依赖问题对于不同事件检测模型的性能影响，并展现了自动适应不同程度数据不平衡和跨任务扩展的能力。

**创新点：**本文首次**主动**消解了不同事件检测模型系统面临的共性数据不平衡依赖，基于其对性能影响的方式引入了句子级别识别信息进行应对；提出了分类器自适应知识蒸馏方法，实现句子级别识别信息的自动获取，通用有效地完成了不同模型系统和数据不平衡程度的依赖消解。

(2) 针对事件要素抽取中存在的实体类型过度依赖，首先定义了该问题，并验证了多种不同的事件要素抽取架构均存在对于实体类型的过度依赖，进而影响事件要素类型的抽取表现。在此基础上，提出了基于多角度对比学习的实体类型过度依赖消解方法。该方法采用两种改进的监督对比学习技术，分别从正样本和负样本两个角度实现对实体类型过度依赖的消解，且通过循环训练策略进一步提升多角度消解的效率。实验评估表明，提出的方法可以有效消解不同事件要素抽取架构的实体类型过度依赖，并在基于预训练语言模型的架构上表现出当前最优性能。

**创新点：**本文首次探究了实体提及的类型信息对要素类型建模的**负面影响**，根据

实体类型和要素类型的数据特点验证了该负面影响在不同要素抽取架构中存在的普遍性；提出了两种改进的监督对比学习模块和相辅的循环训练策略，实现了从不同角度降低建模过程对于实体类型的依赖程度，通用有效地消解了实体类型过度依赖对不同要素抽取架构的不利影响。

(3) 针对不同文本级别事件要素抽取中均存在的跨事件依赖，提出了分离-融合抽取范式，将跨事件信息获取和事件要素抽取先分开建模再进行融合，从而利用到不同抽取范式的优势。在此基础上，提出了一个基于分离-融合抽取范式的多词元链接模型，通过采用两个分隔的多词元链接构建和进一步的两阶段融合，实现了建模跨事件信息和保持简单链接推理的高效兼容。实验评估表明，本文提出的模型性能在句子级别和文档级别数据集上均超过了当前最优模型，且在包含多事件信息的文本中表现出更加显著的性能提升。

**创新点：**本文首次提出了能够**高效统一**地利用不同事件要素抽取方式各自优势的分离-融合新范式；提出了相应的多词元链接模型，在有效避免现有通用文本级别事件要素抽取方法各自局限性的同时，实现了跨事件依赖关系的高效构建与利用。

## 2. 下一步研究展望

本文深入研究了事件抽取任务中不同模型架构和文本级别存在的数据不平衡依赖、实体类型过度依赖和跨事件依赖等共有数据特性，提供了有效的解决方法，为文本信息的智能化应用发展提供了有利支撑。然而，结合本文研究内容和当下人工智能的发展潮流，依然存在以下需要进一步探索的研究问题：

(1) 研究基于多词元链接的端到端事件抽取。当前主流的端到端事件抽取方法主要基于序列生成模型，在训练和推理阶段依赖枚举部分或全部事件类型提示模版进行事件信息的获取，存在计算效率不高的局限。而基于多词元链接的方法能够实现文本、事件类型和要素类型的并行化编码和链接构建，在保证计算效率的同时，提供了良好的端到端事件抽取研究潜力。然而，相比于事件要素抽取，事件抽取任务的建模提升了链接推理的复杂性，并进一步增加了跨事件信息利用的难度。本文将在下一步工作中研究如何实现端到端事件抽取任务的有效拆解和相适应的融合机制。

(2) 研究基于大语言模型增强的事件抽取。现有监督模型较为依赖对于特定词或词对信息的映射学习，缺乏事件的复杂推理能力和抽象知识建模能力。最近，大语言模型技术通过指令微调和学习反馈，具备了对未见或罕见事件模式的良好泛化推理能力。因此，本文后续将研究如何利用大语言模型构建抽象层级更高的事件概念知识，

辅助增强传统事件抽取模型性能的进一步提升。

(3) 研究基于部分标注场景的事件抽取。由于事件抽取数据的标注需要专业的语言学知识和经验,无法保证每个事件数据的完整信息均被有效识别。此外,其标注信息的细粒度特点决定了相应的训练语料构建成本相当高昂。因此,无论是从客观存在的事件信息不完整还是语料标注的成本考虑,研究基于部分标注场景的事件抽取方法具备了广阔的应用前景和价值意义。本文将在接下来的工作中研究如何基于对比学习、伪标签生成和样本置信度评估等技术,构建适应部分标注场景的事件抽取方法。