

# 基于线索挖掘的篇章间事件关系抽取方法

胡志磊<sup>1,2,3</sup>, 李紫宣<sup>1,2,3</sup>, 隗继耀<sup>1,2,3</sup>, 张瑾<sup>1,2,3</sup>, 靳小龙<sup>1,2,3</sup>

1. 中国科学院计算技术研究所智能算法安全全国重点实验室, 北京 100190;
2. 中国科学院网络数据科学与技术重点实验室, 北京 100190;
3. 中国科学院大学计算机科学与技术学院, 北京 100049

## 摘要

篇章间事件关系抽取旨在识别篇章主题事件之间的关系。其中, 篇章主题事件指的是篇章描述的主要内容, 一般假定一个篇章内只有一个篇章主题事件。现有方法仍然面临着以下挑战。①事件关联线索难以捕捉: 事件之间缺乏共享的上下文, 难以直接获得表明事件关系的相关线索。②事件关系判断缺乏依据: 当两篇文档在内容上没有明显重叠时, 缺乏明确的证据来判断事件关系。为了解决上述挑战, 提出了一个知识增强的线索挖掘模型 KACM, 用于篇章间事件关系抽取。KACM 不仅可在篇章内部信息中挖掘相应的关联线索, 还利用外部知识来辅助关系判断。在 DTER 数据集上的实验结果表明, KACM 可以有效地在因果、时序和共指关系上进行篇章间事件关系抽取。

## 关键词

篇章间事件关系; 篇章主题事件; 事件关系抽取; 知识增强; 线索挖掘

中图分类号: TP391

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2026011

## *A cue mining-based method for inter-document event relation extraction*

Hu Zhilei<sup>1,2,3</sup>, Li Zixuan<sup>1,2,3</sup>, Wei Jiyao<sup>1,2,3</sup>, Zhang Jin<sup>1,2,3</sup>, Jin Xiaolong<sup>1,2,3</sup>

1. State Key Laboratory of AI Safety, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100090, China
2. CAS Key Laboratory of Network Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100090, China
3. School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China

## *Abstract*

Inter-document event relation extraction aims to identify relations between document-thematic events. Specifically, a document-thematic event refers to the central event conveyed by a document, with the common assumption that each document contains only one document-thematic event. Existing methods face the following challenges: (1) difficulty in capturing event association cues. There is a lack of shared context between events, making it challenging to directly obtain relevant cues that indicate their relations; (2) a lack of sufficient evidence for event relation judgment. When two documents show no obvious content overlap, clear evidence to determine their event relation is often unavailable. To address these challenges, this paper proposes a knowledge-augmented cue mining model for inter-document event

relation extraction, termed KACM. The KACM not only mines association cues from within the document content but also leverages external knowledge to assist in relation inference. Experimental results on the DTER dataset demonstrate that KACM can effectively perform inter-document event relation extraction in terms of causal, temporal, and coreference relations.

### Key words

inter-document event relation, document-thematic event, event relation extraction, knowledge augmented, cue mining

## 0 引言

篇章间事件关系抽取旨在识别篇章主题事件之间的关系，在自然语言处理领域中具有重要的意义。在该任务中，给定两个篇章，模型需要判断两个篇章表示的主题事件之间是否存在具体类别的事件关系。具体来说，篇章主题事件指的是篇章中描述的主要内容，一般假定一个篇章内只有一个篇章主题事件。另外，篇章标题被视为篇章主题事件的触发要素，篇章主题事件相关的论元分散在文档中。在实际应用中，篇章间重要的事件关系主要包括事件因果、时序和共指关系。事件时序关系反映的是事件发生的先后顺序，有助于更好地理解事件的发展脉络。事件因果关系反映了事件之间的因果联系，可以将原本分散的事件有机地连接起来。事件共指关系指的是多个篇章主题事件表示同一件事情，可以将相同的事件信息进行合并，避免冗余。利用篇章主题事件之间的因果、时序和共指关系，可以构建出详细的事件脉络，从而支持多种实际应用，如事件预测<sup>[1]</sup>、阅读理解<sup>[2]</sup>以及问答系统<sup>[3]</sup>。

本文主要研究如何识别篇章主题事件之间的多种关系，包括因果、时序和共指关系。在判断时序关系时，一方面需要明确事件的发生时间，另一方面需要挖掘事件之间的关联线索，根据文本表述方式来

确定事件发生的先后顺序。另外，识别篇章主题事件之间的因果关系也十分具有挑战性。因为篇章内容之间是相互独立的，篇章主题事件之间不存在显式的因果连接词，如“导致”“引起”等。在这种情况下，事件之间的因果关系是隐式存在的，需要模型能够充分地理解两个篇章主题事件的语义，并找到它们之间的关联线索。在识别共指关系时，同样需要结合篇章内容来进行细致的判断。

现有工作主要是对篇章内事件之间的关系进行识别。与篇章主题事件不同，在这些工作中，事件是篇章中的单个词语。一些方法<sup>[4-5]</sup>构建了事件交互图，建模了事件以及文本中的相关信息。一些方法<sup>[6-7]</sup>专注于提升事件的语义表示，对文本中的信息进行筛选，保留信息密度较高的内容。还有一些方法<sup>[8-9]</sup>建模了事件关系的传递性，采用迭代的方式来识别出更多的事件关系。但是上述方法都依赖于事件所在的上下文，需要依靠上下文信息来判断事件之间是否存在关系。而在判断篇章主题事件之间的关系时，每篇文档被视为一个独立的事件，事件之间不再共享相同的上下文。在这种情况下，篇章主题事件的信息分散在各自的文档内容中，上述方法难以有效地识别事件之间的关联关系。

在进行篇章间事件关系抽取时，现有方法仍然面临以下两个关键挑战。

- 事件关联线索难以捕捉。每个篇章主题事件对应一篇独立的文档，拥有各自

的标题和正文。这使得在判断两个篇章主题事件之间的关系时，由于缺乏共享的上下文，难以直接获得表明事件关系的相关线索。在这种情况下，事件之间的关联线索往往隐藏在文档内容中，需要深入理解文档内容，挖掘反映事件关系的隐含线索。

- 事件关系判断缺乏依据。现有方法通常需要对整个篇章的内容进行理解，之后再判断篇章主题事件之间的关系。然而，当给定的两篇文档在内容上没有明显重叠、各自描述的是不同的事件时，由于缺乏直接或明确的证据，判断它们之间的关系就变得非常困难。

为了应对上述挑战，本文提出了一个知识增强的线索挖掘模型（knowledge-augmented cue mining model, KACM）用于篇章间事件关系抽取。针对第一个挑战，KACM 分别理解每篇文档的具体内容，并对篇章内容之间的关系进行建模。具体来说，因为篇章内容比较长，涉及的信息较多，如果直接融合全部信息，会引入非常多的噪声，同时关键信息也隐藏在众多噪声信息中无法识别，所以本文对篇章内容进行分块处理，将文档切分为多个块，分别对每一块的信息进行编码。之后，模型使用注意力机制来判断篇章内容每个部分之间是否存在关联关系。这里的一个基本思想是，如果在两个篇章中，一个篇章的关键信息和另一个篇章的关键信息存在一定的关系，那么认为两个篇章主题事件之间也存在着对应的关系。针对第二个挑战，KACM 不再局限于篇章自身的内容，而是关注外部的知识。可以借鉴人类的思考过程，人类在判断两个篇章主题事件之间的关系时，一方面会阅读两个篇章并理解篇章内的信息，另一方面会利用大脑中存储的常识对两个篇章主题事件之间的关系进行判断。与人类的思考过程进行

类比，模型除了要在篇章内部信息中找到相应的关联线索，也可以利用外部知识来进行关系判断。具体来说，本文对两个篇章标题中的关键词进行提取，之后利用得到的关键词在外部知识库 ConceptNet<sup>[10]</sup>中检索相关的知识，并利用图注意力网络对这些知识进行建模，从而辅助识别篇章主题事件之间的关系。

总体而言，本文的主要贡献如下。

- 本文提出了一种知识增强的线索挖掘模型 KACM，用于篇章间事件关系抽取。模型利用注意力机制来捕捉篇章之间的重要关联线索，可以有效地支持篇章间事件关系的判断。

- 本文利用外部知识来辅助篇章间事件关系抽取，在外部知识库中检索相关知识，并对这些知识进行建模与利用。

- 在篇章间事件关系抽取数据集 DTER 上的实验结果表明，KACM 通过关联线索挖掘以及外部知识建模，可以有效地识别篇章间的事件关系。

## 1 相关工作

### 1.1 篇章主题事件抽取

为了从整个篇章层面抽取出与篇章主题相关的主题事件，研究者提出了篇章主题事件抽取任务。篇章主题事件抽取的目标是在整个篇章中抽取出与主题事件相关的论元，重点关注主题事件的准确性和完整性。该任务旨在识别篇章中最关键的事件，从而概括文章的核心内容，减轻人工分析事件的工作量。在篇章主题事件抽取任务中，面临的核心问题是事件论元分散问题。早期的方法主要是设计相关的线索选择策略，在文档中筛选与事件相关的内容，过滤与事件无关的内容。Patwardhan

等<sup>[11]</sup>利用概率框架对事件相关的内容进行建模。Huang等<sup>[12]</sup>设计了一种多层次的事件抽取架构,利用事件论元角色特定的句子分类器,识别不直接提及事件但包含角色填充词的句子。随后,Huang等<sup>[13]</sup>又提出了一个自底向上的框架来确定事件论元候选词,并且使用额外的分类器来去除句子中的虚假候选词。

随着深度学习技术的发展,之后的方法主要采用神经网络的方法。Chen等<sup>[14]</sup>构建了图模型来对事件进行建模,将句子级事件抽取器提取的每个候选角色作为图中的节点,将候选角色的多种关系作为图上的边,最后通过一个增强的图注意力网络来对图结构进行建模。该方法能够结合局部线索和全局上下文,增强候选角色的语义表示,并有效过滤噪声。Du等<sup>[15]</sup>提出了一个基于预训练的GRIT模型,通过生成式方法建模文档级上下文,捕捉角色填充实体之间的共指关系和跨角色依赖关系。GRIT能够跨句子边界进行实体抽取<sup>[16]</sup>,捕捉长距离的上下文信息,同时能够识别同一实体的不同提及,避免重复抽取。除此之外,Zhou等<sup>[17]</sup>利用文档中的冗余信息,通过实体共指图模块生成全面且共指感知的实体表示。之后,该方法构建了实体摘要图,用来合并同一论元的多个抽取结果,从而提升模型的鲁棒性和性能。

上述方法都属于监督学习的方法,对语料的标注是十分昂贵和费时的。为了应对数据稀缺的场景,研究者还提出了半监督学习的方法。Wang等<sup>[18]</sup>提出了一种从在线新闻中提取5W1H(who、what、whom、when、where和how)事件语义信息的方法。作者设计了一种基于标题重要性的主题句提取算法,该算法能够高效地从新闻中提取关键信息。该方法将基于规则的动词驱动方法和支持向量机

(support vector machine, SVM)分类器结合,可以有效地提取事件类型。在此基础上,作者进一步通过关键事件识别、事件语义元素抽取和事件本体填充3个步骤,实现了从新闻中抽取5W1H语义元素并构建事件知识库的目标<sup>[19]</sup>。在这之后,Hamborg等<sup>[20]</sup>提出了一个名为Giveme5W1H的系统,用于从新闻文章中提取主题事件的核心信息。该系统首先对新闻文章进行分词、词性标注、句法分析、命名实体识别和共指消解,并将时间、地点和实体等信息规范化为标准形式(如TIMEX3时间实例和地理坐标)。之后,系统通过独立的提取链分别回答5W1H问题,并利用句法和语义规则从文章中提取候选短语。Kar等<sup>[21]</sup>进一步提出了一个名为ArgFuse的弱监督框架。ArgFuse包含多个筛选模块,采用主动学习策略,能够在低资源环境中高效工作。在缺乏大量标注数据的情况下,该方法通过弱监督技术实现了文档级论元聚合,降低了对标注数据的依赖。

## 1.2 事件关系抽取

事件关系抽取的目标是识别事件之间的关联关系,现有的方法主要关注句内事件抽取以及篇章内事件关系抽取,其中事件指的是触发事件的单词或词组。在句内事件关系抽取任务中,需要处理的文本是单个句子,而且事件出现在同一个句子里。Liu等<sup>[22]</sup>对文本中的事件上下文模式进行建模,同时使用外部知识来增强事件的表示,最终使用门控机制将这两部分信息融合起来。Zuo等<sup>[23]</sup>提出了一种知识指导的数据增强方法,生成了更多的事件关系数据,并设计了一个循环框架不断提高生成数据的质量,在一定程度上解决了训练数据匮乏的问题。而在篇章内事件关系抽取

任务中,需要处理的文本是整个篇章,事件分散在篇章内的不同位置。Gao等<sup>[24]</sup>提出了一种基于整数线性规划的方法,用于全面识别文档中的事件因果关系,包括句内(intra-sentence)和跨句(cross-sentence)的因果关系。该方法建模了句法关系约束、篇章关系约束和事件共指关系约束。Liu等<sup>[25]</sup>提出了一个PPAT模型,首先预测句子内的事件因果关系,然后基于这些结果推断句子间的事件因果关系,这种策略考虑了句子间因果关系对句子内因果关系的依赖性。另外,该方法还构建了一个跨句事件关系图,其中每个节点表示一个事件对,若两个事件对共享相同的事件,则通过有向边将对应节点连接起来。此外,句子内的事件对节点只与句子内的事件对节点相连,而句子间的事件对节点可以通过有向边从句子内的事件对节点聚合信息。该模型通过逐步推理策略和图注意力机制在篇章内事件因果关系抽取任务中取得了性能提升。上述这些工作都是对单个篇章内事件之间的关系进行分析,目

前还没有工作研究篇章间的事件关系抽取。为了分析篇章主题事件之间的关联关系,本文对篇章间事件关系抽取任务展开研究和探索。

## 2 KACM模型

本文提出的基于知识增强的线索挖掘模型KACM如图1所示。模型的输入部分是两篇文档,其中每篇文档均包含文档标题和文档正文。这两篇文档将输入以下两个组件当中,分别是关联线索挖掘器和外部知识编码器。关联线索挖掘器对文档内容进行编码,获得其相应的上下文表示,之后捕捉两篇文档内容之间的关联线索。外部知识编码器根据篇章主题事件在外部知识库ConceptNet中检索相应的知识,并通过图注意力神经网络对外部知识进行编码,进而辅助识别篇章主题事件之间的关系。随后,模型利用前馈神经网络将关联线索信息以及外部知识信息映射为低维

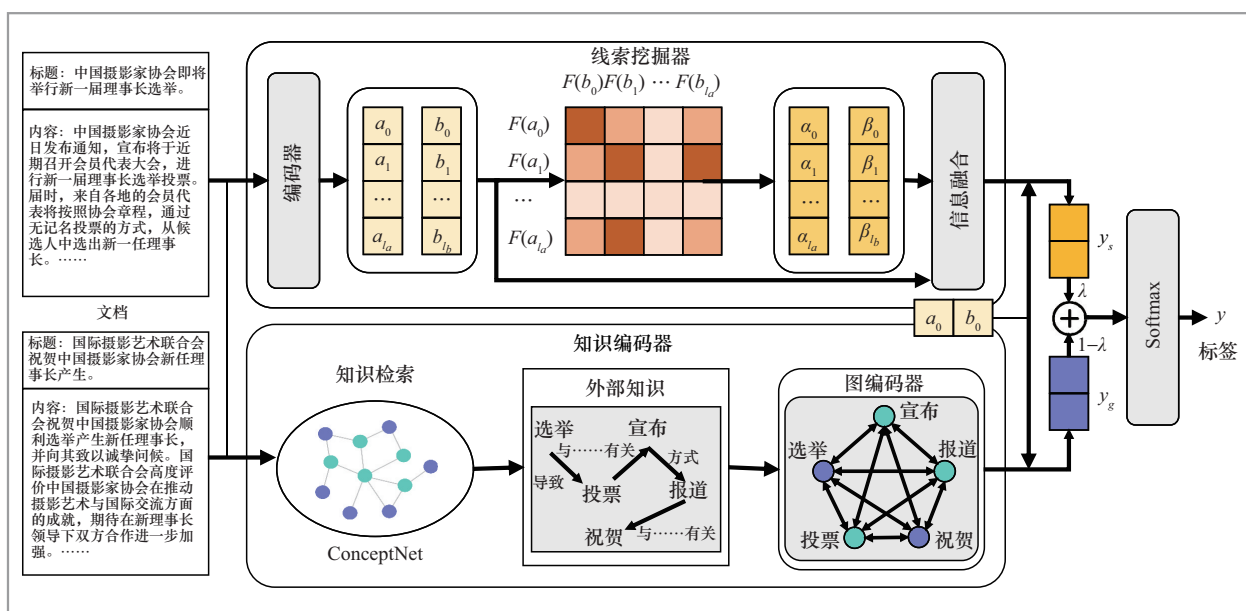


图1 KACM模型示意图

度的向量表示，并将向量表示融合后进行分类。最终，模型给出两个篇章主题事件之间存在对应关系类型的概率。接下来，本文将对模型的两个关键组件进行详细介绍。

## 2.1 关联线索挖掘器

在判断两个篇章主题事件之间的关系时，需要对文档标题和正文内容进行充分的理解，并挖掘两个事件之间的关联线索。事实上，篇章主题事件之间的关联线索会通过不同的内容体现。例如，事件关系可能通过两篇文档标题的语义得到，或者根据一篇文档的标题以及另一篇文档的正文内容，又或者根据两篇文档的正文内容来进行判断。为了全面地捕捉事件之间的关联线索，本文设计了关联线索挖掘器。关联线索挖掘器包含篇章文本编码和关联线索挖掘两个部分。

### 2.1.1 篇章文本编码

关联线索挖掘器的输入是需要判断事件关系的两篇文档，其中文档记为  $D = [T, S]$ ， $T$  表示文档的标题， $S$  表示文档的正文内容。本文使用预训练语言模型（如 RoBERTa<sup>[26]</sup>）作为文本编码器对文档进行编码。文档的标题内容一般较简短，因此直接使用文本编码器进行编码。在标题内容前后分别添加两个特殊标记 [CLS] 和 [SEP]，然后将其输入编码器中，获得编码后的向量表示。与之前的工作类似<sup>[27]</sup>，模型使用 [CLS] 标记对应的向量表示来代表文档标题的语义。对于文档的正文内容，其长度一般比较长，超过了预训练语言模型允许的输入序列长度。为了有效地处理长文本，本文采用滑动窗口机制对其进行处理。该机制将长文本划分为多个较小的片

段，并逐步滑动窗口以覆盖全文，从而实现对整个文本的处理。具体而言，模型以句子为单位进行处理，如果当前窗口无法容纳正在处理的句子，则将其置于后续的窗口中。如果单个句子的长度超过窗口限制，再将其切分为更小的片段。滑动窗口的步长等于窗口大小，窗口之间没有重叠。另外，在每个窗口的起始和结束位置，分别用特殊标记 [CLS] 和 [SEP] 进行标记。对于每个窗口的内容，采用文本编码器对其进行编码，得到相应的向量表示：

$$\{h_{[\text{CLS}]}, h_1, h_2, \dots, h_{L_w}, h_{[\text{SEP}]}\} = \text{Encoder}(\{[\text{CLS}], x_1, x_2, \dots, x_{L_w}, [\text{SEP}]\}) \quad (1)$$

其中， $L_w$  是窗口内 token 的数量， $h_i$  是 token  $x_i$  的向量表示。类似地，模型使用每个窗口 [CLS] 标记对应的向量表示来代表整个窗口的语义。在对篇章文本进行编码后，可以得到文档标题和文档各部分内容的向量表示。

### 2.1.2 关联线索挖掘

为了充分地捕捉两篇文档之间体现事件关系的线索，模型对篇章标题-篇章标题、篇章标题-篇章正文、篇章正文-篇章正文之间的语义进行分析。关联线索识别和挖掘可以帮助模型更好地判断篇章主题事件之间的关系。具体来说，因为模型的输入是两篇文档 A 和 B，因此使用  $\mathbf{a} = \{a_0, a_1, \dots, a_{l_a}\}$  和  $\mathbf{b} = \{b_0, b_1, \dots, b_{l_b}\}$  分别表示两个文档的语义表示。其中， $l_a$  和  $l_b$  分别是两篇文档各自的窗口数目， $a_0$  和  $b_0$  分别是两篇文档各自标题的表示。首先，模型计算文档 A 中每个位置的语义表示和文档 B 中每个位置语义表示之间的关联权重：

$$e_{ij} = \mathbf{F}(\mathbf{a}_i)^T \mathbf{F}(\mathbf{b}_j) \quad (2)$$

其中,  $F$  是一个双层前馈神经网络。之后, 将  $b$  中与  $a_i$  具有关联的语义信息进行聚合, 得到新的语义表示  $\alpha_i$ ; 将  $a$  中与  $b_j$  具有关联的语义信息进行聚合, 得到新的语义表示  $\beta_j$ :

$$\alpha_i = \sum_{j=1}^{l_b} \frac{\exp(e_{ij})}{\sum_{k=1}^{l_b} \exp(e_{ik})} b_j \quad (3)$$

$$\beta_j = \sum_{i=1}^{l_a} \frac{\exp(e_{ij})}{\sum_{k=1}^{l_a} \exp(e_{kj})} a_i \quad (4)$$

然后, 对每篇文档的原始表示和聚合语义信息之后的新表示进行融合, 得到相应的融合结果:

$$v_{1,i} = G([a_i, \alpha_i]) \quad \forall i \in [1, \dots, l_a] \quad (5)$$

$$v_{2,j} = G([b_j, \beta_j]) \quad \forall j \in [1, \dots, l_b] \quad (6)$$

其中,  $G$  是一个新的双层前馈神经网络,  $[\cdot, \cdot]$  表示拼接操作。在得到两组融合向量  $\{v_{1,i}\}_{i=1}^{l_a}$  和  $\{v_{2,j}\}_{j=1}^{l_b}$  后, 对其进行聚合得到聚合后的向量  $v_1$  和  $v_2$ :

$$v_1 = \sum_{i=1}^{l_a} v_{1,i} \quad (7)$$

$$v_2 = \sum_{j=1}^{l_b} v_{2,j} \quad (8)$$

同时, 因为文档标题是文档内容的高度概括, 对关系的判断非常重要, 所以再次引入文档标题的信息。将文档标题的表示  $a_0$  和  $b_0$ , 以及  $v_1$  和  $v_2$  拼接后输入最终的表示聚合器  $H_s$  中, 得到相应的特征表示:

$$y_s = H_s([a_0, b_0, v_1, v_2]) \quad (9)$$

其中,  $H_s$  是一个双层前馈神经网络。得到的特征表示  $y_s$  用于最后的分类。

## 2.2 外部知识编码器

除了充分地理解文档的语义并挖掘其关联线索, 还可以通过外部知识库中大量的知识来辅助识别篇章主题事件之间的关系。事件之间的关系并不总是显式地表述出来, 很多情况下是隐式的。这个时候就需要结合外部知识的信息来进行判断。利用外部知识, 即使文档中并没有明确说明一个事件导致另一个事件发生, 模型依然能够依靠外部知识做出正确的判断。为了充分利用外部知识库中的信息, 本文设计了外部知识编码器来对相关知识进行建模。外部知识编码器主要包括知识检索和知识编码两个阶段。

### 2.2.1 知识检索

知识检索的主要目的是在外部知识库中检索与篇章主题事件有关的知识。具体来说, 本文选用广泛使用的 ConceptNet 知识库。ConceptNet 中将知识表示为图结构, 图中的每个节点对应一个概念, 图上的每条边表示两个概念之间的语义关系。在篇章间事件关系抽取任务中, 输入是整篇文档, 而知识库中存储的是具体的概念, 为了在知识库中检索相关的知识, 需要提取出篇章主题事件的核心概念词。具体来说, 本文采用依存分析 (dependency parsing) 技术对篇章标题进行分析, 构建相应的依存句法树, 并提取其中的核心概念词。依存句法树的示意图如图 2 所示。如果依存句法树的根节点是动词, 则将其作为事件的核心概念词。如果句子中不存在动词, 则将最具有代表性的名词作为事件的核心概念词。在得到两个篇章主题事件的核心概念词之后, 模型在知识库中检索与之相关的知识。

为了体现事件之间的关联关系, 本文

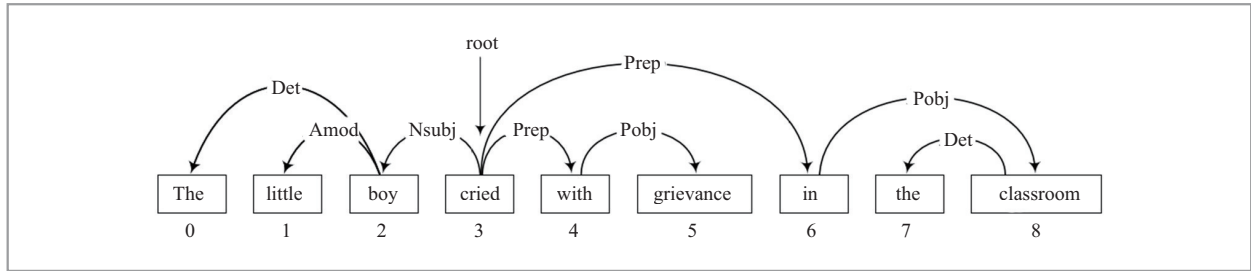


图2 依存句法树示意图

在知识库中检索两个事件之间的多跳路径。多跳路径可以作为事件之间的推理路径，其中的知识说明了两个事件之间是如何建立起关联的。这类知识对于事件关系的识别非常有帮助，尤其是在事件之间缺乏关联线索时，可以发挥关键的作用。此外，两个事件之间的推理路径越短，说明事件的相关性越强，因此模型利用两个事件之间的最短路径作为推理路径。本文将外部知识库 ConceptNet 表示为图，在图上获得两个事件之间的最短路径。当事件之间存在多条推理路径时，模型随机选择一条路径，从而避免信息冗余。随后，最短路径上的所有概念节点都将作为事件相关的知识。

### 2.2.2 知识编码

在检索到事件相关的知识后，模型对这些知识进行编码。为了充分地获取潜在的有用信息，模型对推理路径上的所有知识进行编码。具体来说，所有检索到的知识概念都会被视为图中的节点，并在此基础上构建全连接图。之后，模型使用预训练语言模型对知识概念进行编码，并将编码得到的表示作为图中节点的初始化表示，从而准确地建模知识的语义信息。接下来本文使用图注意力网络（graph attention networks, GAT）<sup>[28]</sup>对图中的节点进行进

一步编码。图注意力网络是一种用于编码图结构数据的神经网络，它利用自注意力机制来捕捉节点之间的依赖关系。图注意力网络可以捕捉节点之间复杂的关联关系，从而为图中的每个节点学习到更加丰富的特征表示。图注意力网络通过注意力机制为图中每个邻居节点分配不同的权重，并根据这些权重将邻居节点的信息聚合到中心节点上。具体地，模型计算节点  $i$  和节点  $j$  之间的注意力得分  $e_{ij}$ ：

$$e_{ij} = \text{LeaklyReLU}(\mathbf{W}_a[\mathbf{W}_e \mathbf{h}_i; \mathbf{W}_e \mathbf{h}_j]) \quad (10)$$

其中，LeaklyReLU 是一个非线性激活函数，引入非线性特性来提高模型的表达能力； $\mathbf{W}_a$  和  $\mathbf{W}_e$  是可学习的参数矩阵； $\mathbf{h}_i$  和  $\mathbf{h}_j$  分别是节点  $i$  和节点  $j$  的表示。之后，使用 Softmax 函数对节点  $i$  邻居节点的系数进行归一化：

$$\alpha_{ij} = \frac{\exp(e_{i,j})}{\sum_{k \in N_i} \exp(e_{i,k})} \quad (11)$$

其中， $N_i$  表示节点  $i$  在图上的邻居节点的集合。随后，对邻居节点的信息进行聚合，得到节点  $i$  更新后的表示。具体地，第  $l \in [0, L-1]$  层的消息传递过程如下所示：

$$\mathbf{h}_i^{l+1} = \sigma \left( \sum_{j \in N_i} \alpha_{ij} \mathbf{W}^l \mathbf{h}_j^l + \mathbf{W}_0^l \mathbf{h}_i^l \right) \quad (12)$$

其中， $\mathbf{h}_i^l$  和  $\mathbf{h}_j^l$  分别是节点  $i$  和节点  $j$  在第  $l$

层的表示； $W'$ 和 $W'_0$ 分别是用于聚合邻居节点信息和节点自身信息的权重矩阵； $\sigma$ 是激活函数（如ReLU）； $h_i^0$ 和 $h_j^0$ 是节点的初始化表示。通过图注意力网络聚合知识信息之后，模型得到两个事件对应节点更新后的表示 $u_1$ 和 $u_2$ 。

同样地，由于文档标题信息的重要性，与关联线索挖掘器中的做法类似，模型再次引入这些信息。将文档标题的表示 $a_0$ 和 $b_0$ 以及融合了外部知识的事件表示 $u_1$ 和 $u_2$ 拼接后输入最终的表示聚合器 $H_g$ ，得到融合之后的特征表示：

$$y_g = H_g([a_0, b_0, u_1, u_2]) \quad (13)$$

其中， $H_g$ 同样是一个双层前馈神经网络。得到的特征表示 $y_g$ 将用于最后的分类。

## 2.3 模型预测

利用从关联线索挖掘器和外部知识编码器中得到的向量表示进行分类。模型将得到的特征表示 $y_s$ 和 $y_g$ 进行组合，作为最终的特征表示，并将其输入Softmax层中进行分类：

$$p = \text{Softmax}(\lambda \times y_s + (1 - \lambda) \times y_g) \quad (14)$$

其中， $p$ 表示两个篇章主题事件之间存在对应关系类型的概率， $\lambda$ 是组件平衡因子，用于在模型中调节关联线索挖掘器和外部知识编码器之间的权重。

接下来对模型参数进行学习，采用交叉熵损失作为损失函数：

$$L(\theta) = -\frac{1}{|V|} \sum_{i \in V} y_i \log p_i \quad (15)$$

其中， $|V|$ 是事件对的数目， $\theta$ 是模型学习到的参数。模型的训练目标是 minimized 上述损失函数。

## 3 实验验证

### 3.1 数据集

目前针对篇章间事件关系抽取任务还没有可用的公开数据集，因此本文构建了适合该任务的DTER数据集。DTER数据集的构建分为文档选择、样本组织、标签标注3个步骤。

首先进行文档选择。新闻文档表达的内容多种多样，多数文档的篇章主题事件之间不存在关联关系，因此需要对众多的新闻文档进行聚类，将表达内容相近的文档放在同一个话题下。在同一个话题下的文档之间更有可能存在事件关系，避免了篇章间事件关系过于稀疏的问题。基于以上原因，本文选择公开数据集MidEast-TE<sup>[29]</sup>中的新闻文档作为原始文档。MidEast-TE是一个大规模的时序知识图谱（temporal knowledge graph, TKG）<sup>[30-31]</sup>数据集，包含了中东地区2015年2月19日至2022年3月17日的新闻文章。此外，MidEast-TE按照文档内容对大量新闻文档进行了聚类，将其划分为多个不同的话题。每个话题下包含多篇文档，这些文档均围绕该话题展开叙述。后续需要对新闻文档进行两两配对，以判断篇章间的事件关系，为了控制样本规模，本文选取了文档数量适中的话题进行研究。最终共选取20个话题，每个话题包含40~50篇新闻文档。

接下来进行样本组织。在得到新闻文档之后，首先需要对其进行划分。具体来说，本文将每个话题下的文档按照2:1:1的比例划分为训练集、验证集和测试集。在生成数据样本之前进行文档划分的目的是防止数据泄露，确保训练集、验证集和测试集中的文档内容没有重叠，从而保证

模型评估的公平性。在划分文档时保证不同数据集之间没有重叠，可以更真实地评估模型在未见数据上的泛化能力。之后，分别对训练集、验证集和测试集中的文档进行两两配对，得到相应的文档对。具体来说，首先根据文档的发布时间对所有文档进行排序，然后按顺序进行组合。与此同时，为了缩减样本规模，在组合过程中只保留一半的文档对，即只保留正向样本（文档1，文档2），删除反向样本（文档2，文档1）。最终将组合得到的文档对作为数据集中的样本。

最后是标签标注阶段。目前大语言模型在文本理解方面展现出了较强的能力，其可以充分理解新闻文档的内容。然而在实际应用中，大语言模型往往需要消耗大量的计算资源，推理速度也相对较慢，因此在实际部署中存在一定的限制。本文尝试将大语言模型的能力蒸馏到小模型（如RoBERTa）中，以兼顾性能与效率。具体来说，本文使用先进的大语言模型DeepSeek-V3<sup>[32]</sup>对每个样本的因果、时序和共指关系进行标注。对于因果关系，需要判断篇章主题事件之间是否存在因果关联；对于时序关系，需判断篇章主题事件之间的时序关系类型，包括“before”（先发生）、“after”（后发生）和“vague”（不确定）；对于共指关系，需要判断两个篇章主题事件是否指代同一个事件。最终，

构建一个适合于篇章间事件关系抽取任务的数据集DTER。数据集的统计信息见表1。

## 3.2 评估指标

与篇章内事件关系抽取任务类似<sup>[4]</sup>，本文将篇章间事件关系抽取任务也视为一个分类任务，采用最常用的精确率（ $P$ ）、召回率（ $R$ ）和F1值（F1-score）作为评估指标。另外，在时序关系抽取中，所有的标签类型都被视为正类（包括VAGUE），因此精确率、召回率和F1值结果相同。

## 3.3 实验设置

### 3.3.1 实现细节

该方法中使用的预训练语言模型是RoBERTa-base，并且在训练过程中对其进行微调。文本编码的向量表示和图上节点表示的维度均为768，滑动窗口机制中的窗口大小设置为256。模型中前馈神经网络的激活函数使用ReLU函数，dropout比率设置为0.2。图神经网络的层数设置为1，数据中负样本的采样比例设置为0.5。篇章间事件因果关系抽取和共指关系抽取中 $\lambda$ 设置为0.5，时序关系抽取中 $\lambda$ 设置为0.8。预训练语言模型的参数学习率和图神经网络的参数学习率都设置为 $1 \times 10^{-5}$ ，其他参数的学习率设置为 $1 \times 10^{-4}$ 。所有参数都通过AdamW梯度优化策略进行优化。数据batch大小设置为3，梯度累积步数设置为4，模型训练10个epoch。该方法使用NetworkX工具包来获取两个事件之间的最短路径。模型均使用PyTorch框架和HuggingFace的Transformer库实现。所有实验均在Tesla V100上进行。

表1 DTER数据集的统计信息

数据集	文档数	事件对数	因果系数	时序系数	共指系数
全部	893	6 991	2 858	4 525	1 910
训练集	440	4 636	1 926	2 954	1 344
验证集	216	1 063	395	589	362
测试集	237	1 292	537	982	204

### 3.3.2 基线方法

为了评估KACM的性能，本文使用以下代表性方法作为基线模型。

- RoBERTa<sup>[26]</sup>是常用的预训练语言模型，已经在大规模语料上进行过预训练，用于下游任务时只需要对其进行微调。使用该模型可以对文本的语义进行编码，从而判断篇章主题事件之间的关系。

- DVA<sup>[33]</sup>是一种深度向量平均的方法，不仅可对文档语义表示的平均值进行处理，而且可捕捉序列中的局部上下文信息，能够更好地处理长序列文档中的信息。

- HMAN<sup>[33]</sup>是基于层次化多注意力网络的分类方法，对文档进行了多层次建模，通过在不同层次上应用不同的注意力机制，动态地分配不同部分的权重，从而提高分类的准确性。

- TextFCG<sup>[34]</sup>是基于图神经网络的分类方法，对文档的标题和正文内容进行建模，利用图结构来捕捉文本之间的关系，旨在通过融合上下文信息来提高分类的性能。

- Llama2-7b-chat<sup>[35]</sup>是Meta推出的LLaMA2系列中的一款对话语言模型，有70亿（7B）参数，适用于聊天和对话任务。相比基础的Llama模型，它经过额外

的监督微调和强化学习，提升了在对话任务中的表现，特别是在指令跟随、事实准确性、上下文理解等方面的能力更加强大。本文使用Few-Shot提示和思维链技术来提升其事件关系抽取能力。

- Llama3-8b-instruct<sup>[36]</sup>是Meta发布的Llama3系列模型之一，有80亿（8B）参数。本文使用Few-Shot提示和思维链技术来提升其事件关系抽取能力。

### 3.4 实验结果

表2展示了KACM模型在DTER数据集上的实验结果。总体来看，对于篇章间事件关系抽取任务，KACM方法优于所有的基线方法。在因果、时序和共指关系上，KACM模型相比于基线模型，F1值分别提升了2.3%、3.4%和2.6%，这表明了本文提出的基于知识增强的线索挖掘模型对事件之间的关联线索进行挖掘以及对外部知识的充分建模，可以有效地识别篇章主题事件之间的关系。在因果关系上，KACM模型取得了最高的召回率，与HMAN模型相比提升了3.5%，这说明该模型可以预测出更多其他模型难以识别到的因果关系。

表2 DTER数据集上的实验结果

模型	因果关系			时序关系			共指关系		
	P	R	F1	P	R	F1	P	R	F1
RoBERTa	54.8%	65.9%	59.9%	59.5%	59.5%	59.5%	25.8%	60.3%	36.1%
DVA	57.2%	64.6%	60.7%	68.9%	68.9%	68.9%	26.8%	61.3%	37.3%
HMAN	52.7%	73.0%	61.2%	60.4%	60.4%	60.4%	25.8%	66.7%	37.2%
TextFCG	55.1%	71.3%	62.2%	70.5%	70.5%	70.5%	27.6%	58.8%	37.6%
Llama2-7b-chat	42.4%	54.4%	47.7%	29.3%	29.3%	29.3%	19.0%	54.4%	28.2%
Llama3-8b-instruct	30.5%	8.0%	12.7%	22.8%	22.8%	22.8%	73.3%	5.4%	10.1%
KACM	55.8%	76.5%	<b>64.5%*</b>	73.9%	73.9%	<b>73.9%*</b>	29.9%	61.3%	<b>40.2%*</b>

注：\*表示模型结果具有统计学显著性提升( $p < 0.05$ )。

出现这一现象的主要原因是KACM模型找到了更多事件之间的关联线索，一方面是在文本内容中挖掘出了更多的语义关联线索，另一方面是利用外部知识来补充事件之间的关联线索，最终取得了较好的结果。除此之外，KACM模型的精确率和HMAN模型相比提升了0.7%，这说明该模型在预测出更多因果关系的同时，仍然保持着较高的预测精准性，可以精确地识别出因果关系。在时序关系上，不同模型表现出较大的差异，可能的原因是时序关系的判断高度依赖于事件发生的具体时间信息，不同模型对这类信息的建模能力存在差异。在共指关系上，KACM模型显著提升了精确率，可以更准确地判断两个篇章主题事件是不是同一个事件。

另外，与直接使用大语言模型Llama2-7b-chat和Llama3-8b-instruct相比，尽管KACM模型的参数量较少(0.125B)，但是经过学习之后，其关系预测能力远远优于这两个大语言模型。Llama3-8b-instruct的结果低于Llama2-7b-chat，原因是其对较长指令的理解能力较差，并没有根据指令要求给出问题的答案，输出的内容对于结果判断没有任何帮助。

### 3.5 消融实验

为了说明KACM模型中各个组件的作

用，本节对其进行了消融实验，实验结果见表3。其中，KACM(-clue)表示去掉关联线索挖掘器，只保留外部知识编码器的结果；KACM(-know.)表示去掉外部知识编码器，只保留关联线索挖掘器的结果；而KACM(-clue & -know.)表示同时去掉关联线索挖掘器和外部知识编码器，只使用预训练语言模型得到的文本表示进行分类。

#### 3.5.1 关联线索挖掘器的影响

在因果关系上，KACM(-know.)和KACM(-clue & -know.)相比，F1值提升了1.6%，这说明对文档内容中的关联线索进行深入挖掘有助于识别事件之间的因果关系。另外，模型的精确率提升了0.5%，这说明通过文档具体内容可以更准确地理解篇章主题事件，并找到它们之间的关联关系。模型的召回率提升了3.4%，这说明模型挖掘到了更多隐含的关联线索，识别到了更多的事件因果关系。在时序关系上，KACM(-know.)和KACM(-clue & -know.)相比，F1值提升了9.1%，这说明时序关系的判断主要依赖于文本内容，对其中的关联线索进行挖掘可以有效地提升模型判断时序关系的能力。在共指关系上，KACM(-know.)和KACM(-clue & -know.)相比，F1值提升了3.3%，这表明关联线

表3 KACM模型消融实验结果

模型	因果关系				时序关系				共指关系			
	P	R	F1	$\Delta$	P	R	F1	$\Delta$	P	R	F1	$\Delta$
KACM	55.8%	76.5%	64.5%	-	73.9%	73.9%	73.9%	-	29.9%	61.3%	40.2%	-
-clue(A)	51.3%	81.4%	62.9%	-1.6%	65.1%	65.1%	65.1%	-8.8%	27.5%	63.2%	38.3%	-1.9%
-know.(B)	55.3%	69.3%	61.5%	-3.0%	68.6%	68.6%	68.6%	-5.3%	28.1%	65.7%	39.4%	-0.8%
-A & -B	54.8%	65.9%	65.9%	-4.6%	59.5%	59.5%	59.5%	-14.4%	25.8%	60.3%	36.1%	-4.1%

索挖掘器同样有助于识别篇章主题事件之间的共指关系。另外，模型的精确率提升了4.1%，这进一步证明了挖掘文档内容之间的关联线索对于事件关系的识别是十分重要的。

### 3.5.2 外部知识编码器的影响

在因果关系上，KACM(-clue)和KACM(-clue & -know.)相比，F1值提升了3.0%，这说明对外部知识建模可以提供更多的辅助关联线索，可以有效地提高模型事件因果关系的识别能力。另外，模型的召回率出现了大幅度的提升（提升了15.5%）。出现这一现象的原因可能是有大量的事件对没有直接的关联线索，利用外部知识可以补充它们之间的关联线索，从而预测出更多的因果关系。但是，与此同时，模型的精确率降低了3.5%，这说明单纯利用外部知识进行预测时，部分结果是不准确的，需要结合具体的上下文来推断事件之间的关系。在时序关系上，KACM(-clue)和KACM(-clue & -know.)相比，F1值提升了5.6%，这说明在时序关系的判断过程中，常识知识在一定程度上也能辅助关系的识别。另外，在因果关系上，外部知识编码器起的作用大于关联线索挖掘器，而在时序关系上有相反的结论。这是因为常识知识可以直接帮助模型确定因果关系，而时序关系更多地是由事件实际

发生的时间来决定的。在共指关系上，KACM(-clue)和KACM(-clue & -know.)相比，F1值提升了2.2%，这表明外部知识编码器同样有助于识别篇章主题事件之间的共指关系。

### 3.5.3 关联线索的互补性

在此基础上，将KACM和KACM(-clue & -know.)进行对比，模型在因果、时序和共指关系上的F1值分别提升了4.6%、14.4%和4.1%，这说明对文本内容的关联线索进行挖掘和对外部知识进行建模是互补的，它们关注了不同类型的关联线索，可以相互促进。将这两部分关联线索信息进行融合，可以同时提高模型的精确率和召回率，实现事件关系的准确识别。

## 3.6 细节分析

### 3.6.1 图编码器的影响

为了分析图编码器对实验结果的影响，本节使用3种不同的图编码器进行实验，分别是GCN<sup>[37]</sup>、DCGCN<sup>[38]</sup>和GAT<sup>[28]</sup>。实验结果见表4。从实验结果可以看到，采用GAT作为图编码器取得了最好的结果。这表明，利用GAT模型可以有效地对外部知识进行建模。GAT模型应用注意力机制，为每个节点的邻居分配不同的权重。这种机制使模型不仅能够识别出节点之间

表4 在DTER上使用不同图编码器的实验结果

模型	因果关系			时序关系			共指关系		
	P	R	F1	P	R	F1	P	R	F1
GCN	54.5%	73.9%	62.7%	57.6%	57.6%	57.6%	25.3%	72.6%	37.5%
DCGCN	51.1%	81.6%	62.8%	61.8%	61.8%	61.8%	28.1%	66.7%	39.5%
GAT	55.8%	76.5%	<b>64.5%</b>	73.9%	73.9%	<b>73.9%</b>	29.9%	61.3%	<b>40.2%</b>

的连接关系,还能够衡量不同邻居对目标节点的贡献程度。这意味着节点与其邻居的关系是动态的,模型可以根据节点之间的相对重要性自动调整权重。

### 3.6.2 组件平衡因子的影响

组件平衡因子 $\lambda$ 是模型中一个非常重要的超参数,它决定了关联线索挖掘器和外部知识编码器对最终结果的影响程度。本节对组件平衡因子进行调节,从而观察其对实验结果的影响,其中 $\lambda \in [0, 0.2, 0.5, 0.8, 1.0]$ 。其中, $\lambda=0$ 表示关联线索挖掘器不起作用,只有外部知识编码器起作用; $\lambda=1.0$ 表示不使用外部知识编码器,只使用关联线索挖掘器。实验结果如图3所示。从图3可以看到,在因果关系上关联线索挖掘器和外部知识编码器同时起作用时,可以取得更好的结果,而且 $\lambda=0.5$ 时取得的结果最好。在时序关系上, $\lambda=0.8$ 时取得的结果最好,这表明关联线索挖掘器对结果的影响程度大于外部知识编码器,外部知识对时序关系识别的作用相对较小。在共指关系上, $\lambda=0.5$ 时同样取得了最好的结果,这表明关联线索挖掘器和外部知识编码器在识别共指关系时都起到了非常重要

的作用。

## 4 结束语

本文提出了一个基于知识增强的线索挖掘模型用于篇章间事件关系抽取。该模型主要包含关联线索挖掘器和外部知识编码器两个组件。其中关联线索挖掘器主要用于挖掘文本中各个部分之间的关联信息,识别文本中隐含的关联线索;外部知识编码器主要用于引入外部知识,并使用外部知识来补充事件之间的关联线索。在DTER数据集上的实验结果表明,KACM模型通过对多种关联线索进行建模,能够有效地捕捉到事件之间的关联线索,从而更准确地对篇章间的事件关系进行识别。

## 参考文献:

- [1] Bai L, Guan S P, Guo J F, et al. Integrating deep event-level and script-level information for script event prediction[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Strouds-

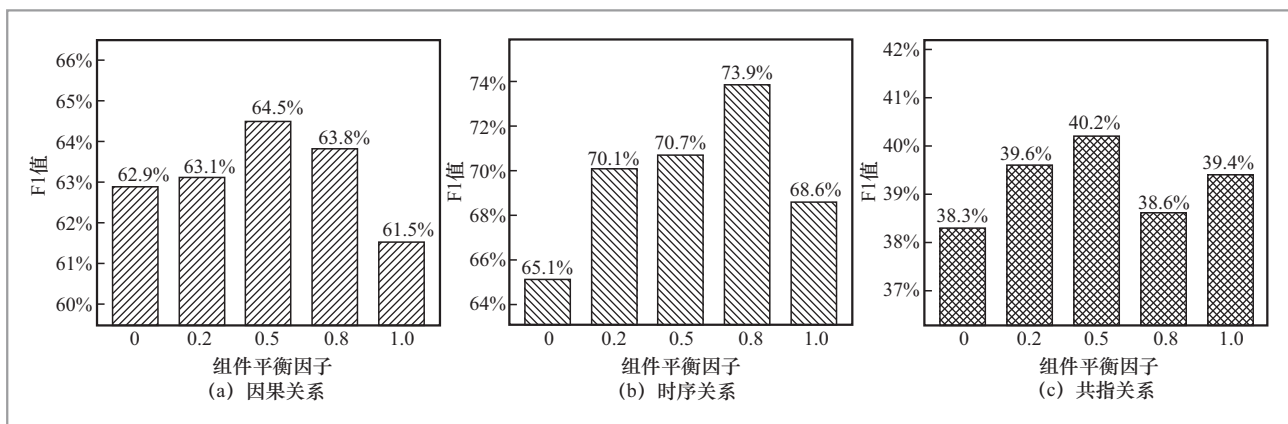


图3 不同组件平衡因子下的实验结果

- burg: ACL, 2021: 9869–9878.
- [2] Kočliský T, Schwarz J, Blunsom P, et al. The NarrativeQA reading comprehension challenge[J]. Transactions of the Association for Computational Linguistics, 2018, 6: 317–328.
- [3] Kwiatkowski T, Palomaki J, Redfield O, et al. Natural questions: a benchmark for question answering research[J]. Transactions of the Association for Computational Linguistics, 2019, 7: 453–466.
- [4] Tran Phu M, Nguyen T H. Graph convolutional networks for event causality identification with rich document-level structures[C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: ACL, 2021: 3480–3490.
- [5] 汪诗蕊, 解博涵, 丁玲, 等. 知识与句法融合的因果关系抽取网络[J]. 大数据, 2024, 10(3): 82–92.
- Wang S R, Xie B H, Ding L, et al. Knowledge and syntax fusion causality extraction network[J]. Big Data Research, 2024, 10(3): 82–92.
- [6] Fan C, Liu D X, Qin L B, et al. Towards event-level causal relation identification[C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2022: 1828–1833.
- [7] Yuan C S, Huang H Y, Cao Y X, et al. Discriminative reasoning with sparse event representation for document-level event-event relation extraction[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2023: 16222–16234.
- [8] Chen M, Cao Y, Deng K, et al. ERGO: Event Relational Graph Transformer for Document-level Event Causality Identification[C]//Proceedings of the 29th International Conference on Computational Linguistics. [S.l.:s.n], 2022: 2118–2128.
- [9] Liu C, Xiang W, Wang B. Identifying while learning for document event causality identification[C]//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2024: 3815–3827.
- [10] Speer R, Chin J, Havasi C. ConceptNet 5.5: an open multilingual graph of general knowledge[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2017, 31(1): 4444 – 4451.
- [11] Patwardhan S, Riloff E. A unified model of phrasal and sentential evidence for information extraction[C]//Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing Volume 1. Morristown: ACL, 2009: 151.
- [12] Huang R, Riloff E. Peeling back the layers: detecting event role fillers in secondary contexts[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Portland: Association for Computational Linguistics, 2011: 1137–1147.
- [13] Huang R H, Riloff E. Modeling textual cohesion for event extraction[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2012, 26(1): 1664–1670.
- [14] Chen P, Yang H, Liu K, et al. Reconstructing event regions for event extraction *via* graph attention networks[C]//Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association

- for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing. Stroudsburg, 2020: 811–820.
- [15] Du X Y, Rush A, Cardie C. GRIT: generative-roller-transformers for document-level event entity extraction[C]//Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. Stroudsburg, 2021: 634–644.
- [16] 徐道柱, 赵凯琳, 康栋, 等. 面向低资源场景的实体知识获取研究综述[J]. 大数据, 2024, 10(1): 46–61.
- Xu D Z, Zhao K L, Kang D, et al. Survey on entity extraction for low-resource scenarios[J]. Big Data Research, 2024, 10(1): 46–61.
- [17] Zhou H Z, Mao K Z. Document-level event argument extraction by leveraging redundant information and closed boundary loss[C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: ACL, 2022: 3041–3052.
- [18] Wang W, Zhao D Y, Zou L, et al. Extracting 5W1H event semantic elements from Chinese online news[C]//Web-Age Information Management. Heidelberg: Springer, 2010: 644–655.
- [19] Wang W. Chinese news event 5W1H semantic elements extraction for event ontology population[C]//Proceedings of the 21st International Conference on World Wide Web. New York: ACM, 2012: 197–202.
- [20] Hamborg F, Breiting C, Gipp B. Giveme5W1H: a universal system for extracting main events from news articles[C]//Proceedings of the 7th International Workshop on News Recommendation and Analytics in Conjunction. Copenhagen: CEUR-WS.org, 2019: 35–43.
- [21] Kar D, Sarkar S, Goyal P. ArgFuse: a weakly-supervised framework for document-level event argument aggregation[C]//Proceedings of the 4th Workshop on Challenges and Applications of Automated Extraction of Socio-Political Events from Text. Stroudsburg: ACL, 2021: 20–30.
- [22] Liu J, Chen Y B, Zhao J. Knowledge enhanced event causality identification with mention masking generalizations[C]//Proceedings of the 29th International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann, 2020: 3608–3614.
- [23] Zuo X Y, Cao P F, Chen Y B, et al. LearnDA: learnable knowledge-guided data augmentation for event causality identification[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. Stroudsburg: ACL, 2021: 3558–3571.
- [24] Gao L, Choubey P K, Huang R H. Modeling document-level causal structures for event causal relation identification[C]//Proceedings of the 2019 Conference of the North. Stroudsburg: ACL, 2019: 1808–1817.
- [25] Liu Z Y, Hu B T, Xu Z R, et al. PPAT: progressive graph pairwise attention network for event causality identification[C]//Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization, 2023: 5150–5158.
- [26] Liu Y H, Ott M, Goyal N, et al. Ro-

- BERTa: a robustly optimized BERT pretraining approach[EB]. arXiv preprint, 2019, arXiv:1907.11692.
- [27] Devlin J, Chang M W, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019: 4171-4186.
- [28] VelićKović P, Cucurull G, Casanova A, et al. Graph attention networks[C]//Proceedings of the International Conference on Learning Representations. [S.l.:s.n.], 2018.
- [29] Ma Y S, Ye C C, Wu Z J, et al. SCTc-TE: a comprehensive formulation and benchmark for temporal event forecasting[EB]. arXiv preprint, 2023, arXiv:2312.01052.
- [30] Li Z X, Jin X L, Li W, et al. Temporal knowledge graph reasoning based on evolutionary representation learning[C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2021: 408-417.
- [31] Xu Y, Ou J J, Xu H, et al. Temporal knowledge graph reasoning with historical contrastive learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(4): 4765-4773.
- [32] DeepSeek-AI, Liu A, Feng B, et al. DeepSeek-V3 technical report[R]. 2025.
- [33] Huang Y R, Chen J J, Zheng S M, et al. Hierarchical multi-attention networks for document classification[J]. International Journal of Machine Learning and Cybernetics, 2021, 12(6): 1639-1647.
- [34] Wang Y Z, Wang C X, Zhan J Y, et al. Text FCG: fusing contextual information *via* graph learning for text classification[J]. Expert Systems with Applications, 2023, 219: 119658.
- [35] Touvron H, Martin L, Stone K, et al. Llama 2: open foundation and fine-tuned chat models[EB]. arXiv preprint, 2023, arXiv:2307.09288.
- [36] Grattafiori A, Dubey A, Jauhri A, et al. The Llama 3 herd of models[EB]. arXiv preprint, 2024, arXiv:2407.21783.
- [37] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[C]//Proceedings of the 5th International Conference on Learning Representations. [S.l.:s.n.], 2017.
- [38] Guo Z J, Zhang Y, Teng Z Y, et al. Densely connected graph convolutional networks for graph-to-sequence learning[J]. Transactions of the Association for Computational Linguistics, 2019, 7: 297-312.

## 作者简介



胡志磊 (1996-), 男, 中国科学院计算技术研究所博士生, 主要研究方向为自然语言处理、知识图谱、事件关系抽取。



李紫宣 (1995-), 男, 博士, 中国科学院计算技术研究所副研究员, 主要研究方向为知识图谱、知识增强大模型。



魏继耀 (1998-), 男, 中国科学院计算技术研究所博士生, 主要研究方向为知识计算、自然语言处理、知识图谱。



张瑾 (1978-), 男, 博士, 中国科学院计算技术研究所高级工程师, 主要研究方向为信息检索、自然语言处理、话题分析。



靳小龙 (1976-), 男, 博士, 中国科学院计算技术研究所研究员, 主要研究方向为知识图谱、知识计算、大数据知识工程。

收稿日期: 2025-10-29

通信作者: 靳小龙, jinxiaolong@ict.ac.cn

基金项目: 中国科学院战略先导项目(No.XDB0680102);国家自然科学基金项目(No.62306299, No.62406308)

**Foundation Items:** The Strategic Priority Research Program of the CAS (No.XDB0680102), The National Natural Science Foundation of China (No.62306299, No.62406308)