

·人工智能·

DOI:10.15961/j.jsuese.202201096



本刊网刊

一种端到端的事件共指消解方法

刘 浏^{1,2}, 蒋国权^{1*}, 环志刚¹, 刘姗姗¹, 刘 茗¹, 丁 鲲¹

(1.国防科技大学第六十三研究所, 江苏南京 210007; 2.宿迁学院信息工程学院, 江苏宿迁 223800)

摘要: 事件共指消解任务主要是判断不同事件提及是否指向同一事件。事件共指消解不仅能有效缓解事件抽取任务中存在的信息冗余问题, 而且为事件内容补全提供了有效途径。尽管许多学者利用深度学习方法对事件共指消解进行了大量研究。但是大部分事件共指消解模型中仍然存在显式信息表示不足、论元噪声引入以及共指事件分布稀疏等问题。针对上述问题, 提出了一种利用显式论元信息和重构事件链的端到端事件共指消解方法。首先, 使用名为OneIE事件抽取模型提取事件的触发词和论元以获取事件的结构化信息; 随后, 使用Transformer编码器对事件提及上下文进行表示, 并将置信分数引入论元信息编码以缓解其可能带来的误差传递; 同时, 采用门控机制对论元在触发词的水平和垂直方向上的信息进行分解, 并根据论元和触发词的相关系数融合两个方向的信息, 过滤论元中的噪声; 然后, 使用前馈网络计算事件提及对共指得分; 最后, 通过重构事件链验证事件提及的合法性以纠正由共指事件稀疏性带来的模型训练结果偏差。为了验证方法的有效性, 本文基于数据集ACE2005进行实验。结果表明, 本文模型在端到端事件共指消解任务上具有一定的先进性, 其中CoNLL和AVG指标平均高出基线模型5.67%和6.24%。

关键词: 事件共指消解; 自然语言处理; 预训练语言模型

中图分类号: TP391.1

文献标志码: A

文章编号: 2096-3246(2024)01-0082-07

事件共指消解是自然语言处理领域的重要任务之一, 被广泛应用于机器阅读理解、信息抽取、机器问答等场景^[1-2]。相较于实体共指消解, 事件共指消解具有更高的挑战性, 并且研究相对较少。它的任务是将指向现实世界中同一事件提及的不同事件提及聚合在共同的事件链中^[3-4]。事件主要由触发词和事件论元构成。触发词是触发事件的词, 能够用于直接描述事件, 而论元包括事件的其他重要信息, 如施事者、受事者、时间和地点等^[5]。以事件S1和S2为例:

S1: On Friday morning, in an unusually rapid response, the Mumbai police arrested one of the five attackers at Volte Gallery, and the police said they had identified the others.

S2: "One of the five has been seized and he admitted to being in Volte Gallery along with some others", Singh said at a news conference Friday afternoon.

S1和S2的触发词分别是“arrested”和“seized”, 事件类型为“Justice: Arrest-Jail”。在S1和S2中, 受事者分别为有着相同语义的“one of the five attackers”和“One of the five”。此外, S1和S2中事件发生的地点相同, 均为“Volte Gallery”。因此, S1和S2是共指的事件提及, 并将其归为同一事件链: {arrested, seized}。

已有的大部分事件共指消解工作通常直接使用标注好的事件信息^[6], 这类方法严重依赖标注数据并且泛化性差。现阶段的研究主要在未标注语料中进

收稿日期:2022-10-12

基金项目:国家自然科学基金项目(71901215);江苏省“333工程”培养资金资助项目(BRA2020418);中国博士后科学基金资助项目(2021MD703983);国防科技大学科研计划项目(ZK20-46);宿迁市科技计划项目(K202128)

作者简介:刘 浏(1988—),男,讲师,博士。研究方向:自然语言处理;人工智能;知识图谱。E-mail: 260344762@qq.com

*通信作者:蒋国权,副研究员, E-mail: jianggq2001@163.com

网络出版时间:2023-06-30 10:35:14 网络出版地址:https://link.cnki.net/urlid/51.1773.TB.20230628.1738.002

行事件共指消解,这类方法更具挑战性和实际意义^[3,7-8]。Peng等^[7]设计一个事件抽取和事件共指消解的端到端模型,从未标注文本中抽取事件信息并进行共指消解,但没有过滤的事件信息会导致错误传递。Lu等^[8]提出一种联合学习模型,联合学习事件检测任务、事件回指检测和事件共指消解任务,联合模型的各个子任务相互促进,可以有效缓解错误传递问题,使模型达到最佳的性能。

事件论元作为事件的关键信息,在事件共指消解任务中被广泛使用^[3,9-10]。使用事件论元解决共指消解通常基于如下假设:如果两个事件提及是共指的,那么它们对应角色的论元也是共指的。Lu等^[3]使用BERT隐式建模事件论元,联合学习事件检测和事件共指消解,提出了显式建模论元的积极效果。Lu等^[9]不区分事件和论元,使用span代替它们,基于SpanBERT模型联合学习事件共指消解和实体共指消解,并通过添加软约束和硬约束来限制结果。Chen等^[10]只利用触发词和部分论元解决事件共指,遗漏了部分重要论元信息。目前主要有3种使用论元信息的方法:1)用预训练语言模型隐式建模事件论元^[3];2)显式建模事件论元,但对论元角色不予区分^[9];3)显式建模论元并区分其角色,但论元不充分,丢失部分重要信息^[10]。本文采用的是显式的建模论元信息,并且将论元划分为5种角色,使其更具一般性。这种划分方式既能够区别对待不同角色的论元,又能够防止一些未知角色论元的遗漏。

由于BERT和SpanBERT等预训练语言模型^[11-12]在自然语言处理任务上获得显著成功,大多数事件共指消解方法使用该模型隐式地获取事件提及的上下文信息。虽然预训练语言模型强大的词嵌入能力允许触发词隐式地蕴含上下文信息,但论元作为事件的重要组成部分,显式引入它们作为一种特征对解决事件共指问题有重要意义^[13-14]。同时,相较于触发词,事件论元更为复杂。一方面,不同类型的事件具有不同角色的事件论元(事件子类型“Marry”的事件论元有3种角色,而事件子类型“Attack”有5种),并且同种角色论元可能不止一个。另一方面,事件提及中的部分论元可能缺失,例如:S1中角色为“交通工具”的论元为“A 30-foot Cuban patrol boat”,但S2中没有对应角色的论元。为了合理使用论元信息解决事件共指消解,本文首先将所有事件的角色划分为:施事者(agent)、受事者(patient)、时间(time)、地点(place)和其他(other),前4种是事件的基本信息,“其他”包含无法划分到前4种中的论元信息。例如,事件类型“Attack”的论元信息包含在{agent: Attacker; patient: Target; time: Time; place: Place; other: In-

strument}集合中,其中论元角色“Instrument”被划为“other”。这种改进的划分方式,既可以满足分别处理对应角色论元的需求,也保证所有的论元都囊括其中,不会致使某些论元缺失。但是论元也容易引入噪声,并且不同事件类型对不同论元的敏感度具有差异性。

此外,事件共指消解的输入通常依赖于上游事件抽取的结果,这种管道方式容易造成错误的传递。同时,共指事件对于事件提及的数量来说具有明显的稀疏性,即单链事件占据相对较大的比重,这就导致模型容易出现偏差,降低其泛化能力。

针对上述问题,本文提出了一种端到端的事件共指消解方法。该方法主要贡献有:第一,使用置信分数和门控机制自适应地控制论元的信息流,从而降低噪声干扰的同时提高论元对不同事件类型的敏感度。第二,通过重建事件链缓解由事件共指链分布稀疏导致的模型学习偏差。第三,在ACE2005数据集上验证了本文模型的先进性。

1 事件共指消解模型

1.1 模型结构

本文模型主要包括4个组件:事件抽取、提及编码器、共指得分器和共指链重构,工作流程如图1所示。第1步,通过事件抽取模块获取事件提及的触发词和论元信息;第2步,使用提及编码器获取触发词和论元的嵌入表示;第3步,对上述表示进行二次计算,并进行拼接以获取事件提及的表示;第4步,通过得分器计算事件对的共指分数;第5步,使用修剪算法对事件共指链进行重构。

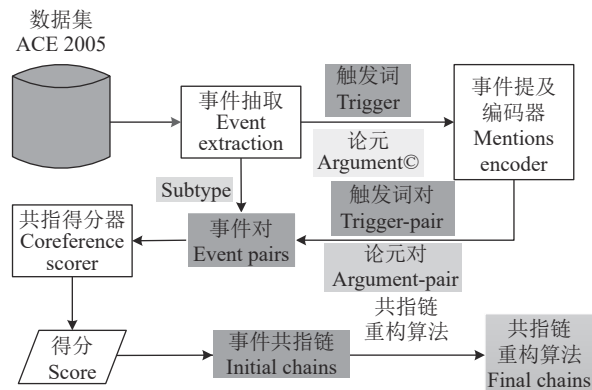


图1 事件共指消解模型结构

Fig. 1 Architecture of event coreference resolution model

1.2 事件抽取

为得到较为准确的事件信息,采用OneIE^[15]方法来提取事件提及的触发词和论元信息。OneIE作为性能比较突出的事件抽取方法之一,在ACE2005数据

集上获得了较好的效果。根据OneIE模型预测的论元信息,将其划分为前文中提到的5种角色,并且为每个论元保留一个置信分数。

1.3 提及和编码器和共指得分器

提及和编码的输入是一个包含 n 个tokens和 k 个事件提及 $\{m_1, m_2, \dots, m_k\}$ 的文档 D 。提及和编码器如图2所示。

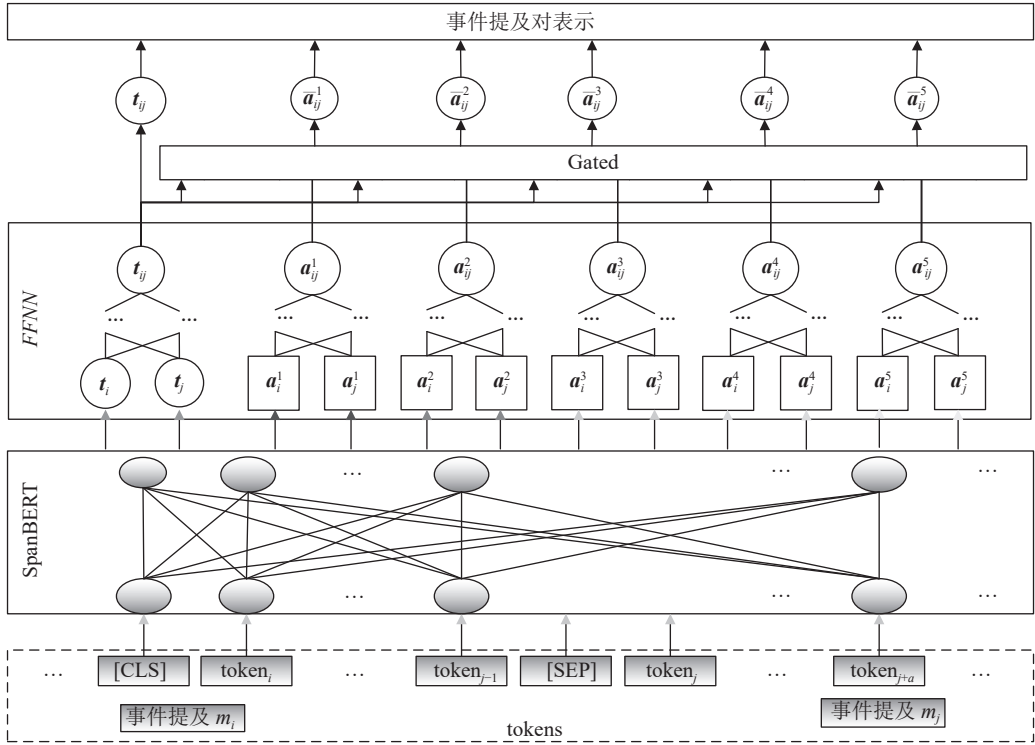


图 2 提及编码器的结构

Fig. 2 Structure of mention encoder

模型首先使用Transformer编码器为每个输入token形成上下文表示,用 $\mathbf{X}=(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ 表示编码器的输出,其中 $\mathbf{x}_i \in \mathbb{R}^d$, d 表示每个token编码后的向量维度。对于每个 m_i ,用 s_i 和 e_i 分别表示触发词(或论元)的开始和结束索引,它的触发词 t_i 被定义为其token嵌入的平均值:

$$\mathbf{t}_i = \sum_{j=s_i}^{e_i} \frac{\mathbf{x}_j}{e_i - s_i + 1} \quad (1)$$

式中, t_i 表示提及 m_i 的触发词。但是,如上所述,提及 m_i 对应的论元可能不止一个,并且来自信息抽取中的错误可能会对事件共指消解产生负面影响。为此,在事件抽取时为每个论元分配一个置信分数 $c \in (0, 1]$,它表示该论元是提及 m_i 对应角色 r 的论元的概率。当论元置信分数 c 越接近1时,使用它引入错误的可能性越小,反之亦然。为了缓解错误传递带来的消极影响,在论元的表示中引入置信分数,对应角色 r 的论元 \mathbf{a}_i^r 定义如下:

$$\mathbf{a}_i^r[l] = c \cdot \sum_{j=s_i^l}^{e_i^l} \frac{\mathbf{x}_j}{e_i^l - s_i^l + 1}, l \in [1, u] \quad (2)$$

$$\mathbf{a}_i^r = \text{Pooling}(\mathbf{a}_i^r[1], \mathbf{a}_i^r[2], \dots, \mathbf{a}_i^r[u]) \quad (3)$$

式(2)~(3)中, $r \in \{\text{agent}, \text{patient}, \text{time}, \text{place}, \text{other}\}$, $\mathbf{a}_i^r[l]$ 为提及 m_i 对应角色 r 的第 l 个论元, s_i^l 和 e_i^l 分别表示第 l 个论元的开始和结束索引, c 表示第 l 个论元的置信分数, u 表示 m_i 对应角色 r 的论元个数。在获得 m_i 对应角色 r 的所有论元表示后,采用池化策略(*Pooling*)得到最终的论元 \mathbf{a}_i^r 。当 m_i 对应角色 r 的论元为缺失或不存在时,使用一个 d 维0向量表示。

给定两个提及 m_i 和 m_j ,触发词对和对对应角色 r 论元对的表示分别被定义为:

$$\mathbf{t}_{ij} = \text{FFNN}_t(\mathbf{t}_i, \mathbf{t}_j, \mathbf{t}_i \circ \mathbf{t}_j) \quad (4)$$

$$\mathbf{a}_{ij}^r = \text{FFNN}_t(\mathbf{a}_i^r, \mathbf{a}_j^r, \mathbf{a}_i^r \circ \mathbf{a}_j^r) \quad (5)$$

式(4)~(5)中, FFNN_t 是 $\mathbb{R}^{3 \times d} \rightarrow \mathbb{R}^d$ 的前馈神经网络, $\mathbf{a}_i^r \circ \mathbf{a}_j^r$ 编码 m_i 和 m_j 的论元级相似性。

为进一步缓解错误传递和获取事件提及上下文中最有用的信息,受到Lai等^[13]的启发,设计了一种门控过滤机制,利用触发词过滤论元中的噪声,如图3所示。

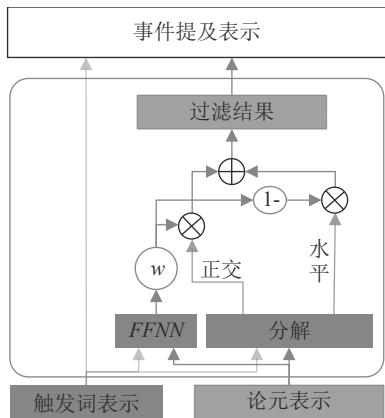


图3 门控模块的结构

Fig. 3 Structure of the gated module

首先,基于触发词表示 t_{ij} 将论元分解为平行分量 p_{ij}^r 和正交分量 o_{ij}^r 。平行分量 p_{ij}^r 是 a_{ij}^r 在 t_{ij} 方向上的投影,它可以被视为包含已经是 t_{ij} 部分的信息。相比之下, o_{ij}^r 与 t_{ij} 正交,被视为包含新信息,即只包含原始向量 a_{ij}^r 与 t_{ij} 线性无关信息。当原始论元表示中噪声少且具有互补信息时,应利用 o_{ij}^r 中的新信息,反之亦然。

$$p_{ij}^r = \frac{(a_{ij}^r \cdot t_{ij})t_{ij}}{|t_{ij}|^2} \quad (6)$$

$$o_{ij}^r = a_{ij}^r - p_{ij}^r \quad (7)$$

然后,使用以下方法获取两个分量的权重系数:

$$w_o = \sigma(\text{FFNN}_p([t_{ij}, a_{ij}^r])) \quad (8)$$

$$w_p = 1 - w_o \quad (9)$$

式(8)~(9)中, w_o 和 w_p 分别是正交分量和水平分量上的权重系数, FFNN_p 是一个 $\mathbb{R}^{2 \times p} \rightarrow \mathbb{R}$ 的前馈神经网络, σ 是sigmoid激活函数。 w_o 越高, a_{ij}^r 与 t_{ij} 的相关性就越高,则认为 a_{ij}^r 中的噪声越少,相关的互补信息越多,则模型应更多地利用 o_{ij}^r 中的新信息。

因此得到过滤后的论元对表示为:

$$\bar{a}_{ij}^r = w_o \circ o_{ij}^r + w_p \circ p_{ij}^r \quad (10)$$

将基于触发词的表示和所有基于论元的表示进行拼接构造最终的提及对表示:

$$f_{ij} = \text{concat}(t_{ij}, \bar{a}_{ij}^{\text{agent}}, \bar{a}_{ij}^{\text{patient}}, \bar{a}_{ij}^{\text{time}}, \bar{a}_{ij}^{\text{place}}, \bar{a}_{ij}^{\text{other}}) \quad (11)$$

最后,事件提及 m_i 和 m_j 的共指得分 $s(i, j)$:

$$s(i, j) = \text{FFNN}_a(f_{ij}) \quad (12)$$

式中, FFNN_a 是 $\mathbb{R}^{6 \times p} \rightarrow \mathbb{R}$ 的前馈网络。

1.4 事件链重构

对于每个事件提及 m_i , 模型将从所有的候选提

及 y_i 中为它分配一个先行词 m_j 或者虚拟先行词 ε : $m_j \in y_i, y_i = \{\varepsilon, m_1, m_2, \dots, m_{i-1}\}$ 。虚拟先行词代表两种情况: 1) m_i 不是事件提及; 2) m_i 是事件提及, 但它与前面的所有事件提及都不共指。本文设定 $s(i, \varepsilon) = 0$ 。两个提及共指的一个必要的条件是有相同的事件子类型, 因此仅将具有相同事件子类型的提及对作为候选共指提及对。

最直接的构建事件共指链的方法是从每个候选事件提及中找到最好的(共指得分最高的)事件提及^[8]:

$$\hat{y}_i = \underset{j < i}{\text{argmax}} s(i, j) \quad (13)$$

式中, \hat{y}_i 表示 m_i 得分最高的候选共指对。

算法1 事件链重构

输入: 初始化事件链 $C_1 = \{c_1, c_2, \dots, c_l\}$

共指得分scorer

输出: 事件链 C_2

1. for $i = 1, 2, \dots, l$ do
2. // 验证单链(提及数=1)
3. if c_i is a singleton do
4. for $j = 1, 2, \dots, i-1, i+1, \dots, l$ do
5. $s(i, j) \leftarrow \text{scorer}(c_i, c_j)$;
6. $(c_i, c_j) \leftarrow \underset{j < i}{\text{argmax}} s(i, j)$;
7. if $s(c_i, c_j) > \omega_1$ do
8. merge c_i into c_j ;
9. $C_2 \leftarrow \text{update}(C_1)$;
10. // 验证长链(提及数>2)
11. if c_i is a long-chain do
12. for each mention m in c_i do
13. $s \leftarrow \text{scorer}(m, c_i - \{m\})$;
14. if $s < \omega_2$ do
15. del m from c_i ;
16. set m is a singleton;
17. $C_2 \leftarrow \text{update}(C_1)$;
18. Return C_2

但这种贪婪算法只考虑局部一致性, 无法保证全局最佳。为此, 文献[12]设计了一种类型指导的解码机制保证事件链的全局一致性。不同于上述方法, 本文基于共指链分布的稀疏性设计了一种新的成链算法。从大量的实验中观察到单链(singleton)占比远比共指链占比高, 因此需要再次考虑单链的合法性。此外, 将提及数大于2的事件链视为长链(long-chain)。事件链的复杂度随事件链的长度增加而提升, 因此额外验证长链中每个提及的合法性。具体地, 给定文档D中的事件提及 $\{m_1, m_2, \dots, m_k\}$, 模型首先通过上述贪婪算法得到最初的事件共指链, 然后再使用事件

链重构算法1得到最终的事件共指链。

在算法1中,对于D中每个事件链 c_i ($c_i \in C$, $C = \{c_1, c_2, \dots, c_i\}$),用链中所有提及表示的平均池化表示该链。算法1中第2~8行验证单链。对于文档D中的每个单链 c_i ,使用模型训练得到的scorer(第2.3节)分别计算 c_i 与其他事件链 c_j ($c_j \in (C - \{c_i\})$)的共指得分,如果 c_i 和 c_j 的得分最高且大于阈值 ω_1 ,合并 c_i 和 c_j 并更新事件链。算法1中第9~15行,对于每个长度大于2的事件链 c ,依次计算链中每个事件提及 m 和 $c - \{m\}$ 的得分,如果得分小于阈值 ω_2 ,就认为 m 和 c 中其他提及共指的概率较低,因此移出 c 并将其作为一个单链更新事件共指链。然后再次使用算法1,对D中的所有单链和长链进行一次重构。

1.5 训练

模型的目标是输出文档中的所有事件共指链。当一个事件提及的预测先行词是它的真实共指事件时,认为这个预测的先行词是正确的先行词。为了让模型得到最佳结果,本文优化所有正确的先行词的边际对数似然^[16]:

$$\mathcal{L} = \lg \prod_{i=1}^k \sum_{m_j \in y_i \cap GOLD(i)} P(i, j) \quad (14)$$

$$P(i, j) = \frac{\exp(s(i, j))}{\sum_{y' \in y_i} \exp(s(i, y'))} \quad (15)$$

式(14)~(15)中, $GOLD(i)$ 表示 m_i 的真实共指事件链,如果 m_i 不存在真实共指事件,则 $GOLD(i) = \{\varepsilon\}$, $P(i, j)$ 表示 m_i 与 m_j 共指的概率。

2 实验和结果

2.1 数据集及指标

在ACE2005数据集^[17]上进行所有的实验,该数据集包含599个文档。为实验结果的公正性,选取与文献[8]和[14]相同的40个新闻文章作为测试集进行实验,并且随机选择30个其他的不同题材的文档作为验证集,剩余529个文档用作模型的训练集。

实验结果分析指标沿用文献[8]中的CoNLL和AVG指标来衡量。CoNLL分数是B3^[18]、MUC^[19]和CEAF_e^[20]3个指标的平均值,AVG分数是B3、MUC、CEAF_e和BLANC^[21]的平均值。

2.2 实验设置

使用SpanBERT作为Transformer编码器^[12, 22]。在实验中对于不同的任务设置不同的学习率,SpanBERT的学习率为 5×10^{-5} ,任务学习率为 5×10^{-4} 。在提及编码器中,设置SpanBERT的编码维度 $d=768$,并且设置FFNN的维度 $p=500$,深度为1。在算法1中,设

置阈值 ω_1 和 ω_2 均为0。设置dropout=0.5,每次训练的batch的大小设置为8,epoch=50。

2.3 对比模型

1) Baseline模型仅采用触发词作为事件共指消解的特征,即式(11)的 f_{ij} 只包含式(4)中 t_{ij} 。

2) SSED+SupervisedExtended与SSED+MSEP的不同是前者使用文献[23]中的事件表示方法,后者在前者的基础上提出了MSEP模型,该模型将事件元素抽象成5类元素从而实现事件的结构化向量表示。

3) CDGM^[8]使用事件提及对的触发词以符号特征来计算事件共指得分。为了使模型能够更好地学习稳定信号,CDGM+Noise模型在其训练过程中对符号特征随机增加噪声。

2.4 基于事件预测的实验

表1为端到端模型在ACE2005数据集上整体的结果。本文使用OneIE来抽取事件提及、类型和论元。表1表明,相比于先前的工作,本文模型取得了最佳的效果,在CoNLL和AVG两个指标与其他经典模型相比平均提升了5.67%和6.24%。相较于方法SSED,本文的方法和CDGM方法都有明显的优势,这是因为在对论元信息表示上都采用了门控机制过滤噪声,因此效果提升较为明显。虽然本文与CDGM方法都使用OneIE作为事件抽取模块,同时也采用门控机制对事件信息进行过滤,但是后采用的是隐式手段将事件提及中的论元信息包含在触发词和符号特征中。同时,SSED和CDGM这两种方法没有处理由事件共指链稀疏性导致的模型训练偏差问题。因此本文模型获得了较优的结果。

表1 端到端模型结果对比

ACE(预测事件信息)	CoNLL	AVG
SSED+Supervised _{Extended}	55.23	52.53
SSED+MSEP	53.80	51.38
baseline	58.93	55.78
CDGM	58.99	56.32
CDGM+Noise	62.07	59.76
本文模型	63.48	61.40

2.5 基于事件标注的实验

为了更好地分析模型的有效性,本节使用数据集中标注的事件提及进行实验。为此,实验将论元的置信分数 c 全部设置为1,这表示论元信息是完全可信的,实验结果见表2。由表2可知,相比于其他模型,本文模型在真实的事件提及上的实验结果平均提升了5.17%和4.9%,进一步说明了本文方法的有效性。但是相较于表1的结果,容易发现目前端到端模型还需要进一步优化,即使对事件信息抽取结果进行过

滤,依然无法弥补其误差导致的错误传递问题。

表2 基于事件标注的实验结果

Tab. 2 Experimental results based on event annotation

ACE(人工标注信息)	CoNLL	AVG
baseline	81.62	81.49
CDGM	87.90	88.30
CDGM+Noise	85.40	85.38
本文模型	90.15	89.96

2.6 消融实验

为进一步探索模型各组件的效果,在ACE2005数据集上进行了消融实验。表3为对事件论元进行消融实验的结果。由表3可知:首先,当删除论元置信分数(-confidence)时,即所有论元置信分数设置为1,CoNLL和AVG指标分别下降了0.74%和1.04%,这说明引入论元置信分数对缓解信息抽取阶段的错误具有积极作用。其次,当删除对应角色为‘other’的事件论元(-other)时,即只使用角色为agent、patient、time和place的4种论元,CoNLL和AVG指标分别下降了1.25%和1.58%,表明角色为‘other’的论元中包含了被遗漏信息,这论证了本文对论元角色划分的合理性。当删除所有论元信息(-all argument),即只使用触发词时,由表3可以看出,模型的效果下降2.90%和4.03%。实验结果再次验证了显示引入论元信息的有效性。最后,本文不区分论元角色,直接使用所有论元表示的平均池化替代区分后的论元(+all argument(mean)),由表3可以看出,不区分论元角色的模型效果下降了1.03%和1.44%。论元作为参与事件的关键角色,区分不同角色的论元,更容易让模型理解论元中的信息,提升模型的效果。

表3 消融实验结果

Tab. 3 Results of ablation experiment

ACE(predicted mentions)	CoNLL变化幅度/%	AVG变化幅度/%
-confidence(arguments)	-0.74	-1.04
-other	-1.25	-1.58
-all argument	-2.90	-4.03
+all argument(mean)	-1.03	-1.44
-reconstruct	-0.85	-1.48

最后对重构共指链算法(-reconstruct)进行消融研究的结果表明,当直接使用贪婪算法构建事件链并将其作为最终的事件共指链时,CoNLL和AVG指标分别下降了0.85%和1.48%。这就验证了本文重构事件共指链算法对提高共指消解模型性能具有正面的效果。

3 结论与展望

本文提出了一种利用显式论元信息和重构事件

链的端到端事件共指消解方法。该方法利用门控和置信分数等方法缓解事件自动化抽取导致的错误信息传播问题。同时通过重构事件链算法减少由共指事件稀疏带来的模型学习偏差。相较于现有模型,本文模型在端到端事件共指消解任务上具有一定的先进性,其中在CoNLL以及AVG指标上平均高出基线模型5.67%和6.24%。但该模型依然存在改进的空间。在未来工作中,将进一步研究如何利用联合学习尝试解决端到端模型中不同任务模块间存在的信息错误传递问题。同时还需要针对跨文本事件共指消解这一难题进行研究。

参考文献:

- [1] Lu Jing,Ng V.Event coreference resolution:A survey of two decades of research[C]//Proceedings of the Twenty-seventh International Joint Conference on Artificial Intelligence. Stockholm:IJCAI,2018:5479-5486.
- [2] Bejan C,Harabagiu S.Unsupervised event coreference resolution[J].*Computational Linguistics*,2014,40(2):311-347.
- [3] Lu Yaojie,Lin Hongyu,Tang Jialong,et al.End-to-end neural event coreference resolution[J].*Artificial Intelligence*, 2022,303:103632.
- [4] Lu Jing,Ng V.Constrained multi-task learning for event coreference resolution[C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics:Human Language Technologies. Stroudsburg:Association for Computational Linguistics, 2021:4504-4514.
- [5] Linguistic Data Consortium.ACE (Automatic Content Extraction) English Annotation Guidelines for Events Version 5.4.3[EB/OL].(2005-07-01)[2022-10-12].<https://www ldc upenn edu/>.
- [6] Krause S,Xu Feiyu,Uszkoreit H,et al.Event linking with sentential features from convolutional neural networks[C]// Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning. Berlin.Stroudsburg:Association for Computational Linguistics,2016:239-249.
- [7] Peng Haoruo,Song Yangqiu,Roth D.Event detection and Co-reference with minimal supervision[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin.Stroudsburg:Association for Computational Linguistics,2016:392-402.
- [8] Lu Jing,Ng V.Joint learning for event coreference resolution[C]//ACL (1).2017:90-101.
- [9] Lu Jing,Ng V.Span-based event coreference resolution[C]. *Proceedings of the AAAI Conference on Artificial Intelligence*,2021,35(15):13489-13497.
- [10] Chen Xinyu,Xu Sheng,Li Peifeng,et al.Sentence rewriting with few-shot learning for document-level event coreference resolution[C]//Mantoro T, Lee M, Ayu MA, et al. *International Conference on Neural Information Processing*. Cham:Springer,2021:152-164.
- [11] Devlin J,Chang M, Lee K, et al.BERT:Pre-training of deep bidirectional transformers for language understanding[EB/OL].[2022-10-12].<https://arxiv.org/abs/1810.04805>.

- [12] Joshi M, Chen Danqi, Liu Yinhan, et al. SpanBERT: Improving pre-training by representing and predicting spans[J]. *Transactions of the Association for Computational Linguistics*, 2020, 8: 64–77.
- [13] Lai Tuan, Ji Heng, Bui T, et al. A context-dependent gated module for incorporating symbolic semantics into event coreference resolution[EB/OL]. [2022-10-12]. <https://arxiv.org/abs/2104.01697>.
- [14] Bejan C A, Harabagiu S. Unsupervised event coreference resolution with rich linguistic features[C]// *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Uppsala: ACM, 2010: 1412–1422.
- [15] Lin Ying, Ji Heng, Huang Fei, et al. A joint neural model for information extraction with global features[C]// *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg: Association for Computational Linguistics, 2020: 7999–8009.
- [16] Lee K, He Luheng, Lewis M, et al. End-to-end neural coreference resolution[EB/OL]. [2022-10-12]. <https://arxiv.org/abs/1707.07045>.
- [17] Christopher Walker, Stephanie Strassel, Julie Medero, et al. Ace 2005 multilingual training corpus. Linguistic Data Consortium[EB/OL]. (2006-02-15)[2022-10-12]. <https://catalog.ldc.upenn.edu/LDC2006T06>.
- [18] Amit Bagga and Breck Baldwin. 1998. Algorithms for scoring coreference chains[C]. *Proceedings of the LREC Workshop on Linguistic Coreference*, Granada, 1998: X563–566.
- [19] Vilain M, Burger J, Aberdeen J, et al. A model-theoretic coreference scoring scheme[C]// *Proceedings of the 6th conference on Message understanding-MUC6'95*. Columbia. Morristown: Association for Computational Linguistics, 1995.
- [20] Luo Xiaoqiang. On coreference resolution performance metrics[C]// *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. New York: Association for Computing Machinery, 2005: 25–32.
- [21] Recasens M, Hovy E. Blanc: Implementing the rand index for coreference evaluation[J]. *Natural Language Engineering*, 2011, 17(4): 485–510.
- [22] Wolf T, Debut L, Sanh V, et al. Transformers: state-of-the-art natural language processing[C]// *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Stroudsburg: Association for Computational Linguistics, 2020: 38–45.
- [23] Sammons M, Peng Haoruo, Song Yangqiu, et al. Illinois CCG TAC 2015 event nugget, entity discovery and linking, and slot filler validation systems[C]// *Theory and Applications of Categories*, 2015.

An End-to-end Event Coreference Resolution Method

LIU Liu^{1,2}, JIANG Guoquan^{1*}, HUAN Zhigang¹, LIU Shanshan¹, LIU Ming¹, DING Kun¹

(1. The Sixty-Third Research Inst., National Univ. of Defense Technol., Nanjing 210007, China;

2. School of Info. Eng., Suqian Univ., Suqian 223800, China)

Abstract: The event coreference resolution (ECR) is mainly to determine whether different event mentions refer to the same event. ECR not only effectively alleviates the problem of information redundancy in event extraction tasks, but also provides an effective way for event completion. Although many scholars have conducted extensive research on ECR using deep learning methods and achieved significant achievements, there are still issues in most ECR models, such as insufficient explicit information representation, noise introduced by arguments, and sparse distribution of coreference events. Aiming at the above problems, an end-to-end ECR method using explicit argument information and event chain reconstruction was proposed. First, an event extraction model called OneIE was used to extract event triggers and arguments. Then, a Transformer encoder is used to express the context of the event mentions, and the confidence score was introduced into the argument information coding to mitigate the error transmission. Meanwhile, the information of the argument in the horizontal and vertical directions of the trigger was decomposed by the gating mechanism, and the noise of the argument was filtered by fusing the information of the directions according to the correlation coefficient of the argument and the trigger. Afterwards, the coreference score of the event pairs was calculated by the feed forward network. Finally, to verify the validity of the event mentions, the event chains were reconstructed to correct the deviation of the model caused by the sparse event coreference. In order to verify the effectiveness of our method, the proposed model is trained and tested on the public dataset ACE2005. The experimental results showed that our model in end-to-end ECR task is 5.67% and 6.24% higher than the other models in the scores of CoNLL and AVG on average.

Key words: event coreference resolution; NLP; pre-trained language model

(编辑 陈雪)

引用格式: Liu Liu, Jiang Guoquan, Huan Zhigang, et al. An end-to-end event coreference resolution method[J]. *Advanced Engineering Sciences*, 2024, 56(1): 82–88. [刘浏, 蒋国权, 环志刚, 等. 一种端到端的事件共指消解方法[J]. *工程科学与技术*, 2024, 56(1): 82–88.]