



计算机科学

COMPUTER SCIENCE

基于核心句的端到端事件共指消解

环志刚, 蒋国权, 张玉健, 刘浏, 丁鲲

引用本文

环志刚, 蒋国权, 张玉健, 刘浏, 丁鲲. [基于核心句的端到端事件共指消解](#)[J]. 计算机科学, 2023, 50(11): 185-191.

HUAN Zhigang, JIANG Guoquan, ZHANG Yujian, LIU Liu, DING Kun. [End-to-End Event Coreference Resolution Based on Core Sentence](#) [J]. Computer Science, 2023, 50(11): 185-191.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于注意力机制的概念增强认知诊断模型](#)

Attention Based Concept Enhanced Cognitive Diagnosis

计算机科学, 2023, 50(11): 241-247. <https://doi.org/10.11896/jsjcx.221100169>

[基于AR与DNN联合模型的地理传感器时间序列预测](#)

Geo-sensory Time Series Prediction Based on Joint Model of Auto Regression and Deep NeuralNetwork

计算机科学, 2023, 50(11): 41-48. <https://doi.org/10.11896/jsjcx.230500231>

[基于语义的多架构二进制函数名预测方法](#)

Semantic-based Multi-architecture Binary Function Name Prediction Method

计算机科学, 2023, 50(10): 369-376. <https://doi.org/10.11896/jsjcx.220800175>

[使用Wi-Fi感知连续行为动作的跨域身份认证](#)

Cross-domain User Authentication via Wi-Fi Sensing of Continuous Activities

计算机科学, 2023, 50(10): 299-307. <https://doi.org/10.11896/jsjcx.220900163>

[基于BiLSTM神经网络的多服务器门限服务系统性能分析](#)

Performance Analysis of Multi-server Gated Service System Based on BiLSTM Neural Networks

计算机科学, 2023, 50(10): 266-274. <https://doi.org/10.11896/jsjcx.221000221>

基于核心句的端到端事件共指消解

环志刚^{1,2} 蒋国权¹ 张玉健² 刘浏^{1,3} 丁鲲¹

1 国防科技大学第六十三研究所 南京 210007

2 东南大学网络空间安全学院 南京 211189

3 宿迁学院信息工程学院 江苏 宿迁 223800

(zhiganghuan@seu.edu.cn)

摘要 大多数先前的事件共指消解模型都属于成对相似度模型,通过编码两个事件提及的表示并计算相似度来判断是否共指。但是,当两个事件提及在文档内出现的位置接近时,编码其中一个事件提及的上下文表示会引入另一事件的信息,从而降低模型的性能。针对此问题,提出了一种基于核心句的端到端事件共指消解模型(End-to-end Event Coreference Resolution Based on Core Sentence, ECR-CS),该模型自动抽取事件信息并按照预先设置好的模板为每个事件提及构造核心句,利用核心句的表示代替事件提及的表示。由于核心句中只包含单个事件的信息,因此所提模型可以在编码事件表示时消除其他事件信息的干扰。此外,受到事件信息抽取工具的性能限制,构造的核心句可能会丢失事件的部分重要信息,提出利用事件在文档中的上下文表示来进行弥补。所提模型引入了一种门控机制,将上下文嵌入向量分解为分别与核心句嵌入向量平行和正交的两个分量,平行分量可以认为是与核心句信息维度相同的信息,正交分量则是核心句中不包含的新信息。通过上下文信息和核心句信息的相关度,控制正交分量中被用来补充核心句中缺失的重要信息的新信息的量。在 ACE2005 数据集上进行实验,结果表明,相比最先进的模型,ECR-CS 的 CoNLL 和 AVG 分数分别提升了 1.76 和 1.04。

关键词 事件共指消解;门控机制;神经网络;预训练语言模型;事件核心句

中图法分类号 TPA391

End-to-End Event Coreference Resolution Based on Core Sentence

HUAN Zhigang^{1,2}, JIANG Guoquan¹, ZHANG Yujian², LIU Liu^{1,3} and DING Kun¹

1 The Sixty-third Research Institute, National University of Defense Technology, Nanjing 210007, China

2 School of Cyber Science and Engineering, Southeast University, Nanjing 211189, China

3 School of Information Engineering, Suqian University, Suqian, Jiangsu 223800, China

Abstract Most previous event coreference resolution models belong to pairwise similarity models, which judge whether the two events are coreferences by calculating the similarity between them. However, when two event mentions appear close to each other in the document, encoding one event contextual representation will introduce information from the other event, which degrades the performance of the model. To solve the problem, an end-to-end event coreference resolution method based on core sentence (ECR-CS) is proposed. The model automatically extracts event information and constructs a core sentence for each event mention according to the preset template, and uses the core sentence representation instead of the event representation. Since the core sentence contains only the information of a single event, the model can eliminate the interference of other event information when encoding the event representation. In addition, limited by the performance of event extraction, the core sentence may lose some important information of the event. The contextual representation of the event in the document is used to make up for this problem. To supplement the missing important information in the core sentence with the contextual information, a gated mechanism is introduced to filter the noise in the contextual representation. Experiments on dataset ACE2005 show that the CoNLL and AVG scores of ECR-CS improves by 1.76 and 1.04, respectively, compared with the state-of-the-art baseline model.

Keywords Event coreference resolution, Gated mechanism, Neural network, Pre-trained language models, Event core sentence

到稿日期:2022-10-10 返修日期:2023-03-23

基金项目:中国博士后科学基金面上资助(2021MD703983);国防科技大学校科研计划项目(ZK20-46)

This work was supported by the General Support from China Postdoctoral Science Foundation(2021MD703983) and Scientific Research Program of National University of Defense Technology(ZK20-46).

通信作者:蒋国权(jianggq2001@163.com)

1 引言

事件共指消解(Event Coreference Resolution, ECR)是对文本中指代现实中相同事件的提及进行聚类的任务^[1]。事件共指消解是自然语言处理(Natural Language Processing, NLP)中的一项重要任务,它是很多NLP任务的基础,例如主题检测^[2]、信息抽取^[3]和阅读理解^[4]等。以下面两个事件提及为例:

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中的触发词分别是具有相同语义的“arrested”和“seized”,事件类型均为“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”。因此它们是共指的,指代的是现实世界中的同一事件。

事件共指消解模型大多属于成对相似度模型,首先利用事件触发词和事件论元等特征建模事件提及的表示,然后将事件表示输入神经网络,为每一对事件提及计算一个相似度分数,最后依据该分数判断它们是否共指,特征的质量直接影响模型的性能。基于传统编码器的事件共指消解模型通过编码每一种特征并按照预先设置的顺序组合建模事件,忽略了各个事件特征在文档中的上下文信息和结构信息。相比传统编码器,预训练语言模型(BERT^[5]等)不仅编码字向量,还编码文本向量和位置向量,并且预训练语言模型的注意力机制在编码过程中主动过滤特征中的噪声。基于预训练语言模型的事件共指消解方法^[6-7]将文档作为输入,直接编码整个文档,然后选取事件触发词所在位置的嵌入作为事件提及的表示。这种方法为每个事件表示融合上下文信息,在事件共指消解任务中取得了显著成功。然而,此类方法却忽略了特征中引入噪声的情况:当两个事件提及在文档中的位置比较接近时,如在同一句话中,利用预训练语言模型编码其中一个事件提及的上下文表示时,不可避免地会将另一个事件的信息引入当前事件提及的表示中,从而导致事件的表示受到另一个事件提及的干扰而带有噪声。以下面两个句子为例:

S3: If NATO **pressured** the Ukrainian government to **capture** and **extradite** him, then it was Ukrainian that **arrested** and **imprisoned** him, not the NATO.

S4: I don't see..., when he was **arrested** by Ukrainian police.

S3中有4个触发词指代4个事件提及,而S4中只有1个事件提及。现在,共指消解模型判断S3和S4中触发词为“arrested”的两个事件提及是否共指。如果直接将这两个句子作为模型的输入,S3中的无关信息(其他3个事件提及的信息)将会干扰模型,从而导致模型做出误判,认为两个事件提及不共指。

为解决上述问题,本文设计了一种基于事件核心句的共指消解方法。从文档中识别事件信息,并根据模板为每个事件提及构建单独的事件核心句,利用核心句的表示来代替事件提及的上下文表示。由于事件核心句只包含单个事件的信息,通过编码核心句获得事件的表示可以消除其他事件对事件表示的干扰。但是,受限于事件抽取组件的性能,从文档中抽取的事件信息可能有所缺失甚至包含错误,构造的核心句会丢失事件的部分重要信息。本文提出利用事件提及在文档中的上下文信息,来补充核心句中缺失的重要信息。为缓解上下文表示中噪声的干扰,本文引入了一种门控机制,用于过滤上下文表示中的噪声。本文的贡献主要包括4个方面:

- 1)提出了一种基于核心句的端到端事件共指消解模型;
- 2)构造事件核心句,降低待消解事件的无关信息对模型的干扰;
- 3)引入了一种门控机制,用于过滤上下文信息中的噪声,利用上下文表示中的有用信息来补充核心句中的缺失信息;
- 4)在ACE2005数据集上进行实验,实验结果表明,相比最先进的基线模型,ECR-CS的CoNLL和AVG分数分别提升了1.76和1.04。

2 相关工作

早期文档内事件共指消解使用传统方法,如基于概率的模型^[8]和基于图的模型^[9]。这类方法大多源自实体共指消解,在事件共指消解任务中的效果不佳。Chen等^[10]针对各种注释事件属性,提出了成对事件共指模型。但是传统成对分类器没有考虑文档内的全局信息,成对结果转化为共指事件链时存在很多障碍。近年来,神经网络在各种NLP应用中得到了广泛的应用。Lu等^[6]探索了基于跨度模型的事件共指消解方法,并证明了跨任务依赖和实体共指信息可以被有益地用于事件共指消解。Lai等^[7]设计了一种基于事件触发词的上下文相关的门控神经网络,用于获取无噪声的事件成对特征以解决英文事件共指。信息抽取中上游任务(实体和事件识别、实体共指消解等)的错误会传播到事件共指消解任务中。为解决错误传播,联合学习模型进入了研究者的视野。Kriman等^[11]提出了一种联合训练模型,联合训练实体识别、事件识别、类型识别和事件共指消解4个任务来提升模型应对错误传播的能力。Lu等^[12]将事件共指与触发词检测、实体共指、回指确定、现实检测和论元提取联合训练,并且通过设计惩罚函数将跨任务一致性约束作为软约束纳入学习过程。Peng等^[13]开发了一个具有最小监督的事件检测和事件共指联合模型,将这些任务视为事件提及或事件提及与类型本体之间的语义相似问题,并使用基于语义角色标记的事件表示来利用文本的结构,共同改善事件检测和事件共指消解。联合模型虽然在一定程度上缓解了错误传播的问题,但因为太过复杂而被诟病。

上述基于神经网络的方法引入预训练语言模型,但它们都是以整个文档为输入进行上下文编码,未考虑到当多个事件提及在文档中的位置接近时,各个事件提及的信息会相互干扰,从而导致模型误判。针对此问题,本文提出了基于事件核心句的事件共指消解方法,为每个事件提及构建一个只

包含该提及信息的核心句,结合该提及在文档中的上下文信息进行共指消解。

3 基于核心句的事件共指消解方法

如图 1 所示,ECR-CS 的总体结构为管道结构,共包括两个模块:事件预测模块和共指消解模块。本文模型首先将处理好的数据集输入事件预测模块,进行事件检测和论元检测;其次通过预先设置好的模板,利用预测的事件信息为每个事件提及构建核心句;然后将核心句和原始文档输入共指消解模块,得到两个事件提及的共指得分;最后根据共指得分判断共指关系,构建事件共指链。

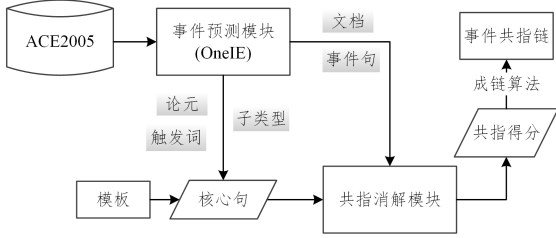


图 1 ECR-CS 的总体结构

Fig.1 Overall structure of ECR-CS

3.1 事件预测模块

本文采用 OneIE^[14] 来识别事件触发词和论元。OneIE

是最先进的信息抽取联合模型,联合学习触发词预测和论元预测任务,在 ACE2005 数据集上获得最佳的效果。ACE2005 数据集定义了 8 种事件类型和 33 种事件子类型,并为每种子类型的事件定义了若干种论元。作为代替,本文修改论元预测的结果,使模型只输出“施事者(arg0)”“受事者(arg1)”“时间(arg_t)”“地点(arg_p)”和“其他(arg_o)”5 种角色的论元,其中“其他”表示无法划分到前四种角色的其他论元。5 种论元的具体说明如表 1 所列。

表 1 本文所使用的论元

Table 1 Arguments used in this paper

角色	说明
arg0	主动发起事件的角色,一般是施事者
arg1	被动参与事件的角色,一般是受事者
arg_t	事件发生的时间
arg_p	事件发生的地点(目的地)
arg_o	无法划分到前四种角色的其他论元

3.2 共指消解模块

先前基于预训练语言模型的共指消解方法的输入是一个文档 D ,其中包含 n 个词元和 k 个事件提及 $\{m_1, m_2, \dots, m_k\}$ 。不同于传统方法直接利用事件提及的上下文信息,本文利用一种简单模板构建每个事件提及的核心句,结合该提及在文档中的上下文信息来完成事件共指消解。共指消解模块的结构如图 2 所示。

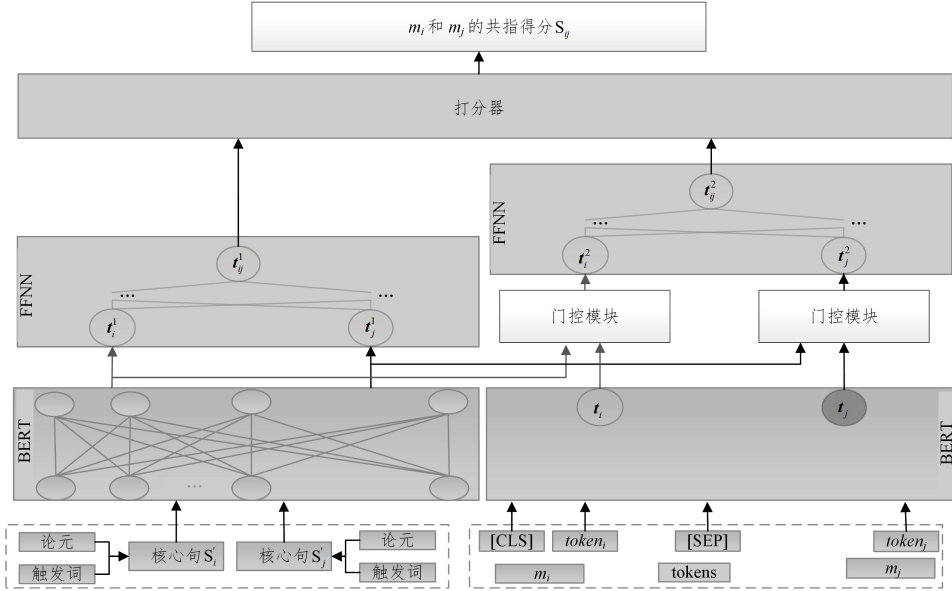


图 2 共指消解模块

Fig.2 Event coreference resolution module

将事件预测模块输出的事件触发词和论元按照【arg_t; arg0; trigger; arg1; arg_p; arg_o】构建每个事件提及的核心句。例如,根据 S3 的信息,构建触发词为“arrested”的事件提及的核心句为“Ukrainian arrested him”。S3 中缺少角色为 arg_t, arg_p 和 arg_o 的论元,因此核心句中缺少事件的时间、地点等信息。将构建好的核心句 S_i' 输入 BERT 对句子进行编码,用 $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ 表示编码器的输出,其中 $\mathbf{X}_i \in \mathbf{R}^d$, d 表示每个词元编码后的向量维度。对于每个提及 m_i ,用 s_i 和 e_i 分别表示触发词的开始和结束索引,它的触发词的表示 t_i^1 被定义为其词元嵌入的平均值:

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

给定两个事件 m_i 和 m_j ,在核心句 S_i' 和 S_j' 中,它们的触发词成对表示 t_{ij}^1 被定义为:

$$t_{ij}^1 = FFNN_i(t_i^1, t_j^1, t_i^1 \circ t_j^1) \quad (2)$$

其中, $FFNN_i$ 是一个 $\mathbf{R}^{3 \times d} \rightarrow \mathbf{R}^p$ 的前馈神经网络, \circ 表示元素级乘法,即对应位置的元素相乘, $t_i \circ t_j$ 编码 t_i 和 t_j 之间的相似性。本文将触发词成对嵌入作为事件核心句成对表示进行事件共指消解。

事件预测包括触发词识别、分类任务和论元识别、分类

任务,在众多 NLP 任务中属于较难的任务,目前提出的事件预测模型的性能仍然有很大的提升空间。本文采用 OneIE^[14] 作为事件预测模型,OneIE 在 ACE2005 数据集上的性能达到最佳,触发词的识别和分类任务的 F1 分数分别达到 75.6 和 72.8,而论元的识别和分类任务的 F1 分数仅为 57.3 和 54.8。受限于事件预测模块的性能,部分论元信息可能未被成功识别或者识别错误,因此构建的核心句中可能缺失事件部分重要信息以及带有部分噪声。为了缓解这个问题,本文交互核心句中的事件表示(触发词嵌入)和文档中的上下文表示(触发词嵌入),利用上下文信息尽可能弥补核心句中缺失的信息。类似式(1),本文采用相同的方法利用文档中触发词的嵌入 t_i 和 t_j 代替原始事件句 S_i 和 S_j 的表示。

受到 Lai 等^[7] 的启发,本文设计了一种门控过滤机制,获取事件上下文表示中最有用的信息,如图 3 所示,利用核心句的特征来有选择地从上下文信息中提取信息。

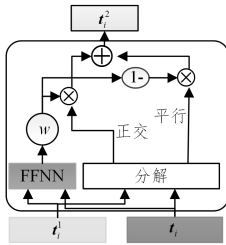


图 3 门控模块的结构

Fig. 3 Structure of gated module

具体地说,首先利用核心句的表示 t_i^1 来对事件上下文表示 t_i 进行正交分解,得到一个平行于 t_i^1 的分量 p_i 和一个与 t_i^1 正交的分量 o_i 。

$$p_i = \frac{t_i \cdot t_i^1}{t_i^1 \cdot t_i^1} t_i^1 \quad (3)$$

$$o_i = t_i - p_i$$

平行分量 p_i 是 t_i 在 t_i^1 方向上的投影,它可以被视为与 t_i^1 维度相同的信息,即与 t_i^1 部分重复的信息。相比之下, o_i 与 t_i^1 正交,因此它可以被视为包含新信息。当上下文表示 t_i 噪声少且具有互补信息时,应利用 o_i 中的新信息,反之亦然。 t_i^1 和 t_i 的相关性越高, t_i 包含的其他事件的信息越少, t_i 包含的信息就越干净,与事件相关的互补信息也越干净。当 t_i^1 和 t_i 的相关性很低时, t_i 包含其他事件信息的可能性很大,当利用这类信息时,虽然会弥补核心句中缺失的信息,但同时会引入更多的噪声,导致模型误判。

本文利用前馈神经网络计算 t_i^1 和 t_i 的相关性,并使用以下方法获取两个分量的权重:

$$\omega_i^o = \sigma(\text{FFNN}_h([t_i^1, t_i])) \quad (4)$$

$$\omega_i^p = 1 - \omega_i^o$$

其中, ω_i^o 和 ω_i^p 分别是正交分量和水平分量上的权重, FFNN_h 是一个 $\mathbf{R}^{3 \times p} \rightarrow \mathbf{R}^n$ 的前馈神经网络, σ 是 sigmoid 激活函数, $[\]$ 表示张量简单拼接操作。 ω_i^o 值越大, t_i 和 t_i^1 的相关性就越高,那么 t_i 中相关的互补信息越干净,则模型应更多地利用 o_i 中的新信息。

然后,得到经过门控模块后的上下文表示 t_i^2 :

$$t_i^2 = \omega_i^o \circ o_i + \omega_i^p \circ p_i \quad (5)$$

给定两个事件 m_i 和 m_j ,上下文成对表示 t_{ij}^2 被定义为:

$$t_{ij}^2 = \text{FFNN}_i(t_i^2, t_j^2, t_i^2 \circ t_j^2) \quad (6)$$

将核心句成对表示 t_{ij}^1 和上下文成对表示 t_{ij}^2 简单拼接,得到最终的事件提及对表示 f_{ij} :

$$f_{ij} = [t_{ij}^1, t_{ij}^2] \quad (7)$$

最后,将事件提及对表示 f_{ij} 输入打分器,得到 m_i 和 m_j 的共指得分 $s(i, j)$:

$$s(i, j) = \text{FFNN}_c(f_{ij}) \quad (8)$$

3.3 训练和推断

模型的目标是输出文档中的所有共指事件链。当一个事件提及的预测先行词是它的真实共指事件时,就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的结果,模型优化所有正确先行词的边际对数似然^[15]:

$$\mathcal{L} = \log \prod_{i=1}^k \sum_{m_j \in y_i \cap \text{GOLD}(i)} P(i, j) \quad (9)$$

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

其中, $\text{GOLD}(i)$ 表示事件提及 m_i 的真实共指事件链,如果 m_i 不存在共指事件提及,则 $\text{GOLD}(i) = \{\epsilon\}$, $P(i, j)$ 表示 m_i 与 m_j 共指的概率。通过优化这个目标函数,模型可以自动学习如何利用真实共指链修剪预测的共指事件链,最终获得最佳的模型性能。

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 ; $m_j \in y_i = \{\epsilon, m_1, m_2, \dots, m_{i-1}\}$ 。虚拟先行词代表 m_i 与前面的所有事件提及都不共指。设定 $s(m_i, \epsilon) = 0$ 。两个提及共指的一个必要的条件是有相同的事件子类型,因此本文仅将具有相同事件子类型的提及对作为候选共指提及对。

通过从每个候选事件提及中找到最好的,即共指得分最高的事件提及,来构建文档内的事件共指链。

$$\hat{y}_i = \arg \max_{j < i} s(i, j) \quad (10)$$

其中, \hat{y}_i 表示 m_i 得分最高的候选共指对。

4 实验和结果

4.1 数据集

本文在 ACE2005 英文数据集上进行所有的实验,它包含 599 个文档。与先前的工作相同^[7,16],选取 30 篇新闻文章作为测试集,随机选择 40 个其他的不同体裁的文档作为验证集进行实验。将剩余的 529 个文档用于训练模型。如表 2 所列,ACE 事件被分为 8 个大类和 33 个小类。

表 2 事件类型和子类型

Table 2 Event types and subtypes

类型	子类型
生活	出生、结婚、离婚、受伤、死亡
移动	运输
事务	所有权转移、转账
商业	成立企业、合并企业、申请破产、结束企业
冲突	攻击、示威游行
交流	会议、电话或书面通信
人事	开始职位、结束职位、提名职位、选举职位
司法	逮捕入狱、释放、审判听证、指控、起诉、定罪、判决、罚款、执行、引渡、无罪释放、上诉、赦免

根据每类事件的特点,ACE2005 定义了事件论元,包括事件参与者和事件发生的时间地点。下面将介绍相关定义。

- 1) 事件提及:描述事件的短语或句子。
- 2) 事件触发词:清楚表示事件发生的主要单词,通常是名词或者是动词。
- 3) 事件论元:实体或短语,包括事件参与者和事件发生的时间和地点。
- 4) 事件类型:由触发词类型决定。

4.2 实验设置

本文使用 SpanBERT (spanbert-base-cased)^[17] 作为编码器。遵循 Lai 等^[7] 的实验参数设置,对于不同的任务,本文设置不同的学习率,SpanBERT 的学习率为 5×10^{-5} ,任务学习率为 5×10^{-4} 。在提及和编码器中,设置 d 为 768,这是 SpanBERT 的编码维度,并且设置 FFNN 的维度 p 为 500,深度为 1。设置 dropout 为 0.5,每次训练的 batch 的大小设置为 8,epoch 为 20。

4.3 评估指标

本文按照 MUC^[18], B³^[19], CEAF_e^[20] 和 BLANC^[21] 标准分别计算精确率 (Precision)、召回率 (Recall) 和综合分数 (F_1),并使用 F_1 评估 GMNN 的性能。

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (11)$$

MUC 是事件共指消解最重要的评估指标,它是基于事件链接的分数。MUC 分数计算了将预测的事件共指链映射到标注的事件共指链(正确的事件共指链)所需插入或删除的最少链接数量。它的缺陷是无法衡量模型预测不存在共指关系的事件的性能。

B³ 是基于事件节点的分数,它弥补了 MUC 对非共指事件评估的不足。B³ 主要对每个事件提及分别计算精确率和召回率,然后以所有事件提及的分数平均值为最终的指标。

CEAF_e 类似于 B³,但它添加了实体相似度来评估事件共指消解的性能。CEAF_e 主要基于这样一种思想:如果两个事件提及是共指的,那么它们对应角色的参数(实体)也是共指的。

BLANC 用于衡量非共指事件和共指事件之间的平均性能。BLANC 实现了兰德指数,主要用于聚类算法。BLANC 正确地处理不存在共指关系的事件,并根据被提及的数量奖励正确的事件提及。BLANC 背后的一个基本假设是,对于给定的事件提及,所有共指链接和非共指链接的总和是不变的。

此外,本文用 CoNLL 和 AVG 指标来衡量结果的 F_1 分数。根据定义,这两个指标是其他标准共指指标 B³, MUC, CEAF_e 和 BLANC 的平均值^[1]。

$$CoNLL = \frac{1}{3} (MUC + B^3 + CEAF_e) \quad (12)$$

$$AVG = \frac{1}{4} (MUC + B^3 + CEAF_e + BLANC) \quad (13)$$

4.4 使用预测提及进行实验的结果

表 3 列出了端到端模型在 ACE2005 数据集上的完整结果。

表 3 ACE2005 数据集上的端到端结果

Table 3 End-to-end results on ACE2005

ACE2005(预测提及)	B ³	MUC	CEAF _e	BLANC	CoNLL	AVG
Peng et al. ^[13]	47.1	59.9	58.7	44.4	55.23	52.53
Lai et al. ^[7]	—	—	—	—	62.07	59.76
ECR-CS	58.5	68.1	64.9	51.7	63.83	60.80

本文中的对比模型均是在 ACE2005 数据集上的成对相似度模型,它们都将整个文档作为模型的输入,没有考虑到在编码时提及之间可能相互干扰,从而导致模型误判。Peng 等^[13] 的方法是基于最小监督的事件检测和事件共指消解的联合方法。与本文方法类似,它解决了阻碍事件研究进展的一些关键问题,包括抽取事件及其关系等。表 3 表明,相比 Peng 等的工作,本文模型在 CoNLL 和 AVG 两个指标上的性能分别提升了 8.60 和 8.27。Lai 等^[7] 通过基于事件触发词的门控神经网络,来获取无噪声的事件成对特征以解决事件共指,在 ACE2005 数据集上取得了最佳的效果。相比此方法,ECR-CS 的性能提升了 1.76 和 1.04。Peng 等联合学习事件检测和事件共指消解两个任务,通过两个任务的相关性(事件检测任务的性能提升可以促进事件共指消解的效果,事件共指消解的结果也能反作用于事件检测)达到同步改善事件检测和事件共指消解的目的。Lai 等为了更加充分地编码事件的表示,引入了 5 个事件属性作为额外的特征,并且通过门控机制过滤事件属性中的噪声。但是他们并没有考虑当两个事件提及在文档中的距离接近或使用事件的额外信息时,可能会引入其他事件同类型的信息带来的噪声,这不仅仅会影响事件共指消解任务,还会对事件检测任务造成负面影响,联合学习方法无法解决这类问题,甚至会扩大它的消极影响。而 ECR-CS 确保每个句子中只有一个事件,避免了其他事件的干扰,同时从事件上下文表示中有选择地过滤信息来补充缺失的事件信息。实验结果表明,ECR-CS 在处理端到端事件共指消解时得到了较为优秀的结果。

4.5 使用标注提及进行实验的结果

为了进一步分析模型的有效性,本文使用数据集中标注的事件提及进行实验,即不使用事件预测模块从数据集中抽取事件,直接使用共指消解模块利用数据集中标注的事件信息进行事件共指消解。实验结果如表 4 所列。

表 4 ACE2005 数据集上使用标注数据的结果

Table 4 Results of using annotation data on ACE2005

ACE2005(标注提及)	B ³	MUC	CEAF _e	BLANC	CoNLL	AVG
Peng et al. ^[13]	74.9	92.8	87.1	83.8	84.93	84.65
Lai et al. ^[7]	—	—	—	—	87.90	88.30
ECR-CS	83.8	94.2	91.9	87.9	89.97	89.45
ECR-CS(-context)	85.5	95.3	93.6	90.2	91.47	91.15

当模型输入的信息很干净时,本文方法构造的核心句干净且具有较完备的信息。表 4 列出了使用标注数据的实验结果。由表 4 可以发现,相比 Peng 等和 Lai 等提出的模型,本文模型在几个关键指标上均取得了最佳成绩,相比最佳模型,ECR-CS 在 CoNLL 和 AVG 两个指标上的性能提升了 2.07 和 1.15。使用数据集中标注的事件信息构建核心句,核心句中包含干净且完整的事件信息,此时使用事件上下文表示不仅不能补充信息而且会额外引入噪声,因此本文进行了另一组

实验 ECR-CS(-context), 未使用上下文信息, 只使用核心句的表示进行共指消解。由表 4 可以看出, 使用标注数据进行实验时, ECR-CS(-context) 比 ECR-CS 在 CoNLL 和 AVG 两个指标上性能提升了 1.5 和 1.7, 这与分析的结果一致。

4.6 消融研究

本文在 ACE2005 数据集上进行了 3 组消融实验, 探究 ECR-CS 各个组件对整个模型的贡献。实验结果如表 5 所列。

表 5 消融研究
Table 5 Ablation study

ACE2005(预测提及)	CoNLL	AVG
ECR-CS	63.83	60.80
- 门控	-1.13	-0.88
- 上下文表示	-5.12	-3.79
- 核心句表示	-5.23	-4.19

第一组实验不使用门控机制, 直接将得到的核心句表示和上下文表示简单拼接后进行实验。由表 5 可以看出, 当不使用门控模块时, 模型的整体性能下降了 1.13 和 0.88, 该结果表明虽然上下文表示中存在大量可用的信息, 但是直接使用上下文表示会带来噪声, 通过门控机制过滤噪声, 挑选出更有用的信息, 可以进一步改善模型的性能。

此外, 第二组和第三组实验表明, 只使用一种表示(核心句表示或上下文表示)进行实验, 模型的性能均会下降, 但只使用核心句表示的模型比只使用上下文表示的模型效果更好。当文档中事件提及的距离较近时, 在编码特征阶段, 提及之间可能会相互干扰, 导致特征中存在噪声。本文使用的核心句只包含一个事件提及的信息, 不会引入其他提及的信息, 相比直接使用上下文特征, 取得了更好的效果。但是, 由于事件预测性能的局限性, 部分事件信息抽取失败, 上下文表示中仍存在重要信息可以解决事件共指, 本文利用上下文表示弥补核心句表示中缺失的重要信息, 模型达到了最佳性能。

结束语 为了应对编码时事件提及之间存在的相互干扰问题, 本文提出了一种基于事件核心句的端到端事件共指消解方法。首先利用抽取出来的事件信息, 直接为每个事件提及构造唯一的不包含其他提及信息的核心句; 然后通过编码每个提及的核心句表示, 并利用一种门控模块结合该提及的上下文表示, 来建模事件提及对表示; 最后为每一对事件提及对表示打分, 并利用共指得分生成共指消解的结果。本文模型在 ACE2005 数据上取得了最佳的结果。未来, 我们将延伸我们的工作以处理更一般的问题, 如跨文档的事件共指消解。

参考文献

- [1] LU J, NG V. Event coreference resolution: a survey of two decades of research[C]// Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. Amsterdam: Elsevier, 2018: 5479-5486.
- [2] MOHAMMED M, MARWA A. Efficient topic detection system for online arabic news [J]. International Journal of Computer Applications, 2018, 180(12): 7-12.
- [3] CHENG L, GAO H K, WANG H B. A news event extraction method in Chinese and Thai languages based on dependency tree elements combined with rules [J]. Software Guide, 2018, 17(7): 49-56, 63.
- [4] WU M Z, MOOSAVI M S, ROTH D, et al. Coreference reasoning in machine reading comprehension[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: 5768-5781.
- [5] 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. Stroudsburg, PA: ACL, 2019: 4171-4186.
- [6] LU J, NG V. Span-based event coreference resolution[C]// Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence. Menlo Park, CA: AAAI, 2021: 13489-13497.
- [7] LAI T M, JI H, BUI T, et al. A context-dependent gated module for incorporating symbolic semantics into event coreference resolution[C]// Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2021: 3491-3499.
- [8] YANG B S, CAIDIE C, FRAZIER P I. A hierarchical distance-dependent bayesian model for event coreference resolution [J]. Transactions of the Association for Computational Linguistics, 2015(3): 517-528.
- [9] LIU Z Z, MITAMURA T, HOVY E H. Graph based decoding for event sequencing and coreference resolution [C]// Proceedings of the 27th International Conference on Computational Linguistic. Stroudsburg: ACL, 2018: 3645-3657.
- [10] CHEN Z, JI H. Graph-based event coreference resolution[C]// Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing. Stroudsburg: ACL, 2009: 54-57.
- [11] KRIMAN S, JI H. Joint detection and coreference resolution of entities and events with document-level context aggregation [C]// Proceedings of the ACL-IJCNLP 2021 Student Research Workshop. Stroudsburg: ACL, 2021: 174-179.
- [12] LU J, 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, PA: ACL, 2021: 4504-4514.
- [13] PENG H R, SONG Y Q, ROTH D. Event Detection and Coreference with Minimal Supervision[C]// In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, Texas. Association for Computational Linguistics, 2016: 392-402.
- [14] LIN Y, JI H, HUANG F. A joint neural model for information extraction with global features[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics

- tics. Stroudsburg: ACL, 2020: 7999-8009.
- [15] LEE K, HE L H, LEWIS M, et al. End-to-end neural coreference resolution[C]// Proc of the 2017 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2017: 188-197.
- [16] CHEN Y B, XU L H, LIU K, et al. 2015. 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, Beijing, China. Association for Computational Linguistics, 2015(1): 167-176.
- [17] JOSHI M, CHEN D Q, LIU Y H, et al. Spanbert: Improving pre-training by representing and predicting spans [J]. Transactions of the Association for Computational Linguistics, 2020(8): 64-77.
- [18] VILAIN M B, BURGER J D, ABERDEEN J S, et al. A model-theoretic coreference scoring scheme[C]// Proceedings of the 6th Conference on Message Understanding. Stroudsburg: ACL, 1995: 45-52.
- [19] BAGGA A, BALDWIN B. Algorithms for scoring coreference chains[C]// The First Conference Language Resources and Evaluation Coreference. ELRA, 1998: 563-566.
- [20] LUO X Q. On coreference resolution performance metrics[C]// Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2005: 25-32.
- [21] RECASENS M, HOVY E H, BLANC; implementing the rand index for coreference evaluation [J]. Natural Language Engineering, 2011, 17(4): 485-510.



HUAN Zhigang, born in 1997, postgraduate, is a student member of China Computer Federation. His main research interests include knowledge graph and natural language processing.



JIANG Guoquan, born in 1978, associate research fellow, master, is a member of China Computer Federation. His main research interests include equipment data engineering and knowledge graph.

(责任编辑:喻藜)