

门控机制融合多种特征的中文事件共指消解

环志刚, 蒋国权, 张玉健, 刘浏, 刘姗姗

引用本文

环志刚, 蒋国权, 张玉健, 刘浏, 刘姗姗. 门控机制融合多种特征的中文事件共指消解[J]. 计算机科学, 2023, 50(3): 291-297.

HUAN Zhigang, JIANG Guoquan, ZHANG Yujian, LIU Liu, LIU Shanshan. [Employing Gated Mechanism to Incorporate Multi-features into Chinese Event Coreference Resolution](#) [J]. Computer Science, 2023, 50(3): 291-297.

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)

[一种基于GRU的半监督网络流量异常检测方法](#)

Semi-supervised Network Traffic Anomaly Detection Method Based on GRU
计算机科学, 2023, 50(3): 380-390. <https://doi.org/10.11896/jsjcx.220100032>

[文档增强型知识库问答](#)

Document-enhanced Question Answering over Knowledge-Bases
计算机科学, 2023, 50(3): 266-275. <https://doi.org/10.11896/jsjcx.220300022>

[演化循环神经网络研究综述](#)

Survey on Evolutionary Recurrent Neural Networks
计算机科学, 2023, 50(3): 254-265. <https://doi.org/10.11896/jsjcx.220600007>

[一种基于三维卷积的声学事件联合估计方法](#)

Sound Event Joint Estimation Method Based on Three-dimension Convolution
计算机科学, 2023, 50(3): 191-198. <https://doi.org/10.11896/jsjcx.220500259>

[基于迁移学习和多视图特征融合提高RNA碱基相互作用预测](#)

Improving RNA Base Interactions Prediction Based on Transfer Learning and Multi-view Feature Fusion
计算机科学, 2023, 50(3): 164-172. <https://doi.org/10.11896/jsjcx.211200186>

门控机制融合多种特征的中文事件共指消解

环志刚^{1,2} 蒋国权² 张玉健¹ 刘浏^{2,3} 刘姗姗²

1 东南大学网络空间安全学院 南京 211189

2 国防科技大学第六十三研究所 南京 210007

3 宿迁学院信息工程学院 江苏 宿迁 223800

(zhiganghuan@seu.edu.cn)

摘要 事件共指消解是很多自然语言处理任务的基础,旨在识别文本中指代相同真实事件的事件提及。由于中文语法相比英文更复杂,捕获英文文本特征的方法在中文事件共指消解中效果并不明显。为解决文档内中文事件共指,提出了一种门控机制神经网络(Gated Mechanism Neural Network,GMNN)。针对中文具有主语省略、结构松散等特点,引入事件基本属性作为符号特征。在此基础上,提出了一种新的门控去噪机制,对符号特征向量进行微调,过滤符号特征中的噪声,提取在特定上下文语境中的有用信息,进而提高共指事件的识别率。在 ACE2005 中文数据集上进行了实验,结果表明,GMNN 的 AVG 分数提升了 2.66,有效地提高了中文事件共指消解的效果。

关键词: 中文事件共指消解;门控机制;神经网络;预训练语言模型;符号特征

中图法分类号 TP391

Employing Gated Mechanism to Incorporate Multi-features into Chinese Event Coreference Resolution

HUAN Zhigang^{1,2},JIANG Guoquan²,ZHANG Yujian¹,LIU Liu^{2,3} and LIU Shanshan²

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

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

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

Abstract Event coreference resolution is the basis of many natural language processing tasks,aiming to identify event mentions in text that refer to the same real event. Since Chinese grammar is much more complex than English,the method of capturing English text features is not effective in Chinese event coreference resolution. To solve the within-document Chinese event coreference,a gated mechanism neural network(GMNN) is proposed. In view of Chinese characteristics with subject omission and loose structure,event attributes are introduced as symbolic features. On this basis,a novel gated mechanism is proposed,which fine-tunes the symbolic feature vector,filters the noise in the symbolic features,extracts useful information in a specific context,and improves the coreference events recognition rate. Experimental results on the ACE2005 Chinese dataset show that the performance of GMNN improves by 2.66,which effectively improves the effect of Chinese event coreference resolution.

Keywords Chinese event coreference resolution,Gated mechanism,Neural network,Pre-trained language models,Symbolic features

1 引言

事件共指消解是对文本中引用相同真实事件的事件提及进行聚类的任务^[1],以下面两个事件描述为例:

S_1 :菲律宾总统埃斯特拉达 2 日下午宣布,菲律宾政府军当天救出了 12 名被阿布沙耶夫武装分子挟持的教士。

S_2 :埃斯特拉达援引菲律宾武装部队总参谋长安赫洛·

雷耶斯的话说,这些人质是当天在南部霍洛岛获救的。

S_1 中事件触发词为“救出”的事件提及和 S_2 中事件触发词为“获救”的事件提及指代的是现实世界中的同一事件“菲律宾政府军救出了人质”,因此它们是共指的。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,如主题检测^[2]、信息抽取^[3]和阅读理解^[4]等。本文重点研究文档内事件共指

到稿日期:2022-07-14 返修日期:2022-11-04

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)

解析,这对于将其扩展到跨文档事件共指消解至关重要。

与英文相比,中文的事件提及更为复杂,且具有语言特点,中文事件共指消解更具有挑战性,主要包括以下 3 个方面:

(1)英文明确了时态和固定搭配,但是汉语中只要语句通顺即可构成完整的句子,这就导致对于同一中文事件的描述可能存在很大差异,从而降低了某些特征(如触发词)的表达能力。

(2)英文非常强调句子结构,而中文句子结构松散,更加强调通过情景传递信息,有时甚至没有主语。因此,中文事件共指消解需要使用更多的人工标注信息来丰富事件表示。

(3)英文的名词和动词由于时态、语态等原因在单词结构上辨识度很高,而中文的名词和动词通常可以是同一个词,不易区分,这会对事件共指消解造成极大的干扰。

本文使用的符号特征包括事件子类型(Subtype)、时态(Tense)、极性(Polarity)、形态(Modality)和泛型(Genericity)。本文认为使用符号特征对于解决中文事件共指消解具有积极意义,以下面两个句子为例。

S_3 :...这些士兵离开了阿富汗...

(... the soldiers left Afghanistan ...)

S_4 :...这些士兵会在六月份离开阿富汗...

(...the soldiers will leave Afghanistan in June...)

在中文语境下, S_3 和 S_4 中的事件触发词都是“离开”,并且它们的事件子类型均为“Movement: Transport”,这两个事件提及很可能会被识别为共指事件。在英文语境下,“leave”作为一个动词,它在 S_3 中的时态是过去时, S_4 中时态为将来时,因此这两个事件不是共指的。 S_3 中“离开”事件对应时态(tense)的属性值为“Past”,表示这个事情已经发生,而 S_4 中事件对应的属性值为“Future”,通过时态这一符号特征,即使在中文语境下也可以判定这两个事件不是共指的。类似地,其他符号特征也可以在不同的语境下帮助解决事件共指。

Teng 等^[5]使用外部工具提取大量特征训练最大熵模型来解决中文事件共指;Chen 等^[6]针对各种注释事件属性提出了成对事件共指模型,首先利用事件的 4 个基本属性(时态、极性、形态和泛型)解决英文事件共指。上述两个模型使用大量人工标注特征,却没有考虑文档内的全局信息,无法捕捉到隐藏在事件提及中的语义。为了解决这个问题,神经网络模型被引入到事件共指消解中并取得了成功。Cheng 等^[7]除了使用 4 个基本属性外,还使用事件触发词、事件类型、子类型和两个事件提及的距离,共同构成事件成对符号特征。然而,上述方法没有考虑符号特征中的噪声。首先,由于共指消解的输入通常来自信息抽取中的上游组件,因此自动提取的符号特征可能有噪声并包含错误;其次,根据特定的上下文,某些特征可能比其他特征包含更多信息。

本文提出了一种门控机制神经网络,用于解决文档内中文事件共指。针对符号特征中的噪声,本文提出了一种基于事件基本特征的门控机制,过滤符号特征中的噪声,提取在特定上下文语境中的有用信息。首先,利用 BERT 编码输入文档中的每个词元(Token),并生成事件触发词和事件论元的

上下文表示;其次,利用可训练的嵌入矩阵计算符号特征的向量表示;然后,通过门控机制过滤符号特征,结合触发词和论元表示得到最终的事件对表示;最后,将事件对表示传入输出层得到共指得分并输出共指事件链。

本文的主要贡献包括 4 个方面:

(1)提出了一种新的门控机制神经网络 GMNN,用于解决中文事件共指问题;

(2)在中文事件共指消解任务中引入预训练语言模型 BERT,编码事件的上下文特征;

(3)提出了一种基于事件基本特征的门控机制,过滤符号特征中的噪声,获取在特定的上下文语境中更有用的信息;

(4)在 ACE2005 语料库上进行实验,在 AVG 指标上,GMNN 的性能比目前最佳的模型提升了 2.66。

2 相关工作

相比英文事件共指消解,中文语料上的研究较少。针对事件共指关系的研究中多采用事件对分类或聚类方法而忽略了事件之间内在联系的问题,Teng 等^[5]提出了一个中文事件共指消解的全局优化模型,用外部工具提取的触发词、事件论元和事件距离等大量特征训练最大熵模型,并利用对称性和传递性等多种约束条件,将共指消解转化成整数线性规划问题来解决中文事件共指。针对事件的语义信息主要由触发词和事件论元表示这一特点,Huan 等^[8]将事件进行结构化表示并输入一个基于门控注意力机制的模型,进行文档内中文事件共指消解。该模型采用语义角色标注和依存句法分析技术对事件句进行浅层语义分析,并使用多头注意力机制挖掘事件句和事件对之间的重要特征。为了解决事件提及中事件无关信息的影响以及中文句子结构灵活多样的问题,Cheng 等^[7]设计了一种门控注意力神经网络模型,从事件提及中选择事件相关信息,然后过滤噪声信息,该模型不仅使用单个余弦距离来计算两个事件提及之间的线性距离,还引入了双线性距离和单层网络机制,进一步计算线性和非线性距离。Teng 等^[5]和 Huan 等^[8]分别使用了传统方法和神经网络来解决中文事件共指,但这两种方法都没有使用事件基本属性作为符号特征。Cheng 等^[7]使用符号特征提高了中文共指消解模型的性能,但未考虑到不同的符号特征在特定上下文中对共指消解的贡献不同。相比上述方法,本文提出的 GMNN 模型在中文事件共指中引入了 BERT,使用事件基本特征(事件触发词和事件论元)和事件符号特征(子类型、极性、形态、泛型和时态)建模事件,结合一种基于事件基本特征的门控机制,降低符号特征中噪声的影响,解决事件共指。

3 门控机制神经网络

本节首先介绍 GMNN 的框架,然后描述其组成部分,即输入层、单提及表示层、提及对表示层和输出层。

3.1 概述

本文提出了一种新的神经网络 GMNN 来解决文档内的中文事件共指,系统框架如图 1 所示。

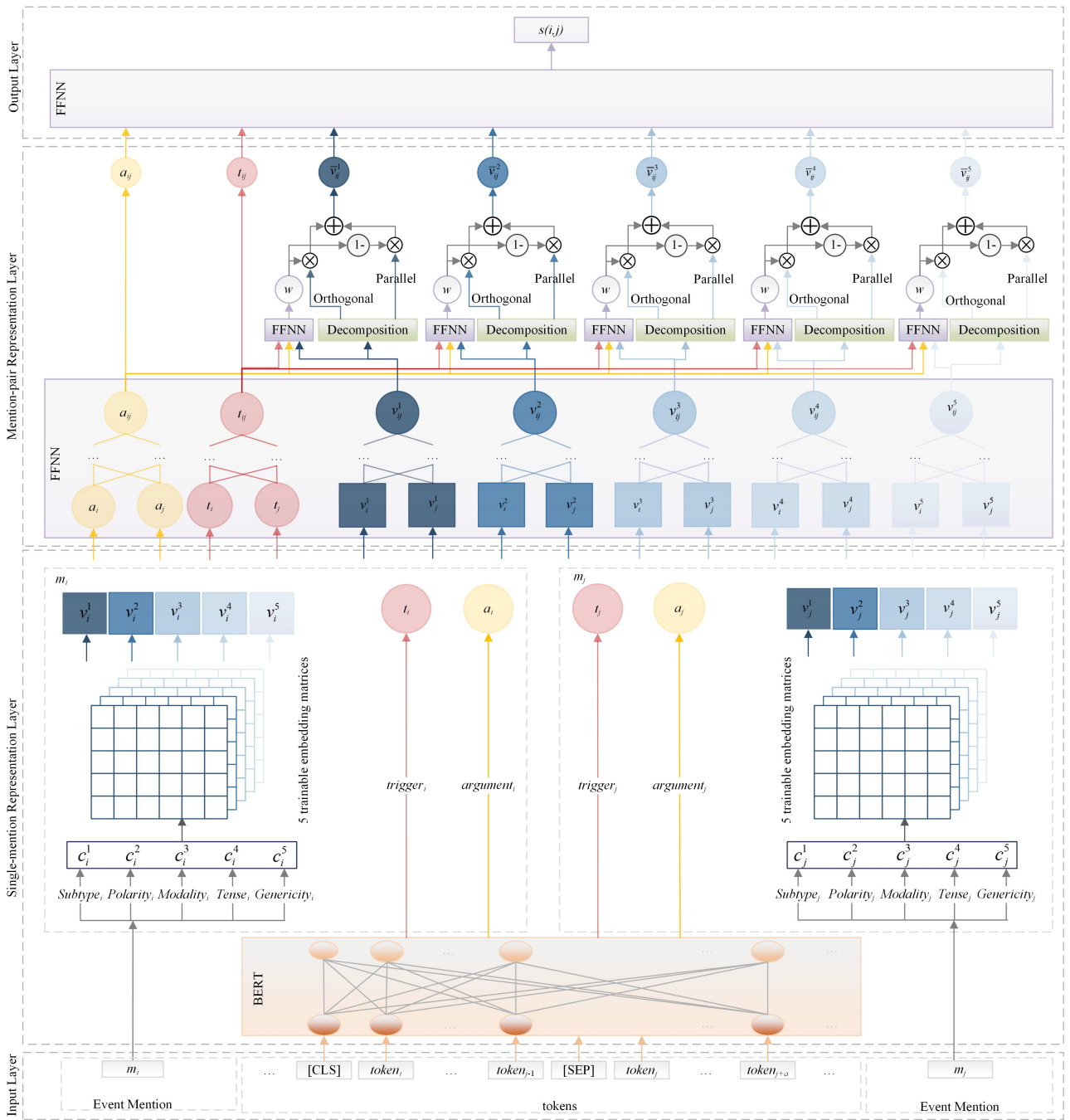


图1 GMNN 系统框架

Fig. 1 Structure of GMNN

GMNN 首先利用事件触发词、事件论元以及事件的 5 个基本属性(事件子类型、极性、形态、泛型和时态)建模每个事件提及的表示;然后使用基于触发词和事件论元的门控机制过滤符号特征中的噪声,获取待消解事件对中更有用的符号特征信息,并通过前馈神经网络形成事件提及对表示;最后将提及对表示输入前馈神经网络输出事件对的共指得分。GMNN 一共分为 4 层,依次为输入层、单提及表示层、提及对表示层和输出层。

3.2 输入层

GMNN 的输入是一个文档 D ,其中包含 n 个词元和 k 个事件提及 $\{m_1, m_2, \dots, m_k\}$ 。 m_i 有 5 种符号特征 $\{c_i^1, c_i^2, \dots, c_i^5\}$,每个 c_i^q 从 $\{1, 2, \dots, N_u\}$ 中取一个离散值,对应一种事件

属性的类型, N_u 表示 c_i^q 的可能取值个数。GMNN 使用的分类特征如表 1 和表 2 所列。

表 1 事件类型和子类型

Table 1 Event type and subtype

| 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 |

表2 事件的基本属性

Table 2 Basic properties of event

| 属性 | 类型 |
|----|-------------|
| 极性 | 否定、肯定 |
| 形态 | 断言、其他 |
| 泛型 | 一般、特殊 |
| 时态 | 未知、过去、现在、将来 |

3.3 单提及表示层

给定文档 D , 首先使用 BERT 为每个词元形成上下文表示。设编码器的输出是 $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, 对于事件提及 m_i , 其触发词表示 t_i 被定义为其词元嵌入的平均值。

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

其中, s_i 和 e_i 表示其触发词的开始和结束索引。

同理, 事件论元表示 \mathbf{a}_i 被定义为所有事件论元的词元嵌入的平均值。

$$\mathbf{a}_i = \frac{1}{n} \sum_{w=1}^n \sum_{j=s_i^w}^{e_i^w} \frac{\mathbf{x}_j}{e_i^w - s_i^w + 1} \quad (2)$$

其中, n 为事件提及 m_i 的论元个数, s_i^w 和 e_i^w 分别表示 m_i 的第 w 个论元的开始索引和结束索引。

此外, 通过 5 个可训练的嵌入矩阵将 m_i 的符号特征 $\{c_i^1, c_i^2, \dots, c_i^5\}$ 转化成 5 个向量 $\{v_i^1, v_i^2, \dots, v_i^5\}$ 。

3.4 提及对表示层

给定两个事件提及 m_i 和 m_j , 其触发词对表示 t_{ij} 定义为:

$$\mathbf{t}_{ij} = \text{FFNN}_t(\mathbf{t}_i, \mathbf{t}_j, \mathbf{t}_i \circ \mathbf{t}_j) \quad (3)$$

其中, FFNN_t 是一个 $\mathbb{R}^{3 \times d} \rightarrow \mathbb{R}^p$ 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, $\mathbf{t}_i \circ \mathbf{t}_j$ 编码 t_i 和 t_j 之间的相似性。

同理, 事件论元对 \mathbf{a}_{ij} 和符号特征对 \mathbf{v}_{ij}^u 的表示分别为:

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

$$\mathbf{v}_{ij}^u = \text{FFNN}_u(v_i^u, v_j^u, v_i^u \circ v_j^u), u \in [1, 5] \quad (5)$$

连接事件基本特征和符号特征最简单的策略是将基本信息(触发词和事件论元)的表示与所有符号特征的表示直接拼接在一起, 即:

$$f_{ij} = [\mathbf{t}_{ij}, \mathbf{a}_{ij}, \mathbf{v}_{ij}^1, \mathbf{v}_{ij}^2, \dots, \mathbf{v}_{ij}^5] \quad (6)$$

但是, 这种方法并不是任何情况都适用的。首先, 这些符号特征在向量化的过程中可能引入噪声甚至错误。其次, 在特定的上下文语境中, 不同的符号特征对解决事件共指的贡献也可能不同。受到 Lai 等^[9]的启发, 本文提出了一种新的门控机制, 如图 2 所示, 通过事件基本特征来有选择地从符号特征中提取信息。

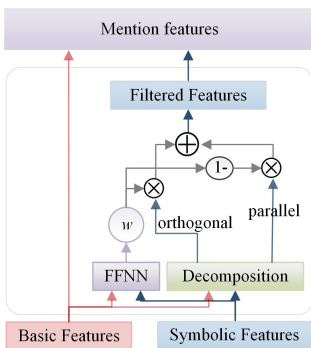


图2 门控机制框架

Fig. 2 Structure of gated mechanism

具体地说, 首先将 \mathbf{t}_{ij} 和 \mathbf{a}_{ij} 相加得到一个新向量 \mathbf{g}_{ij} , 然后利用 \mathbf{g}_{ij} 来对 \mathbf{v}_{ij}^u 进行正交分解, 得到一个平行于 \mathbf{g}_{ij} 的分量 \mathbf{p}_{ij}^u 和一个与 \mathbf{g}_{ij} 正交的分量 \mathbf{o}_{ij}^u :

$$\mathbf{g}_{ij} = \mathbf{t}_{ij} + \mathbf{a}_{ij}$$

$$\mathbf{p}_{ij}^u = \frac{\mathbf{v}_{ij}^u \cdot \mathbf{g}_{ij}}{\mathbf{g}_{ij} \cdot \mathbf{g}_{ij}} \mathbf{g}_{ij} \quad (7)$$

$$\mathbf{o}_{ij}^u = \mathbf{v}_{ij}^u - \mathbf{p}_{ij}^u$$

平行分量 \mathbf{p}_{ij}^u 是 \mathbf{v}_{ij}^u 在 \mathbf{g}_{ij} 方向上的投影, 它可以被视为包含 \mathbf{t}_{ij} 和 \mathbf{a}_{ij} 部分的信息。相比之下, \mathbf{o}_{ij}^u 与 \mathbf{g}_{ij} 正交, 因此它可以被视为包含新信息。当原始符号特征向量 \mathbf{v}_{ij}^u 噪声少且具有互补信息时, 应利用 \mathbf{o}_{ij}^u 中的新信息, 反之亦然。

本文使用以下方法获取两个分量的权重:

$$\mathbf{w}_o = \sigma(\text{FFNN}_h([\mathbf{t}_{ij}, \mathbf{a}_{ij}, \mathbf{v}_{ij}^u])) \quad (8)$$

$$\mathbf{w}_p = 1 - \mathbf{w}_o$$

其中, \mathbf{w}_o 和 \mathbf{w}_p 分别是正交分量和水平分量上的权重矩阵, FFNN_h 是一个 $\mathbb{R}^{3 \times p} \rightarrow \mathbb{R}^p$ 的前馈神经网络, σ 是 sigmoid 激活函数。 \mathbf{w}_o 越高, \mathbf{v}_{ij}^u 和 \mathbf{g}_{ij} 的相关性就越高, 则认为 \mathbf{v}_{ij}^u 中的噪声越少, 相关的互补信息越多, 模型应更多地利用 \mathbf{o}_{ij}^u 中的新信息。

然后, 得到经过门控模块后的符号特征对表示 $\bar{\mathbf{v}}_{ij}^u$:

$$\bar{\mathbf{v}}_{ij}^u = \mathbf{w}_o \circ \mathbf{o}_{ij}^u + \mathbf{w}_p \circ \mathbf{p}_{ij}^u \quad (9)$$

最后, 拼接触发词表示、论元表示以及过滤后的符号特征对表示, 得到最终的事件提及对表示:

$$f_{ij} = [\mathbf{t}_{ij}, \mathbf{a}_{ij}, \bar{\mathbf{v}}_{ij}^1, \bar{\mathbf{v}}_{ij}^2, \dots, \bar{\mathbf{v}}_{ij}^5] \quad (10)$$

3.5 输出层

对于事件提及 m_i 和 m_j , 输出层获得提及对表示层的最终输出 f_{ij} , 然后利用一个前馈神经网络, 输出 m_i 和 m_j 的共指得分:

$$s(i, j) = \text{FFNN}_c(f_{ij}) \quad (11)$$

其中, FFNN_c 是一个 $\mathbb{R}^{7 \times p} \rightarrow \mathbb{R}$ 的前馈神经网络。

对于每个事件提及 m_i , GMNN 将从所有的候选提及 y_i 中为它分配一个得分最高的先行词 m_j 或者虚拟先行词 ϵ :

$$m_j \in y_i = \{\epsilon, m_1, m_2, \dots, m_{i-1}\} \quad (12)$$

$$y_i = \arg \max_{j < i} s(i, j)$$

其中, $s(i, \epsilon) = 0$, y_i 表示 m_i 的得分最高的共指对。虚拟先行词代表 m_i 与前面的所有事件提及都不共指。

GMNN 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, 则认为这个预测的先行词是正确的先行词。为了让模型得到最好的结果, GMNN 优化所有正确的先行词的边际对数似然^[10]:

$$\mathcal{L} = \log \prod_{i=1}^k \sum_{m_j \in y_i \cap \text{GOLD}(i)} P(i, j) \quad (13)$$

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

其中, $\text{GOLD}(i)$ 表示 m_i 的真实共指事件链(数据集中标注的共指事件链)。如果 m_i 不存在真实共指事件, 则 $\text{GOLD}(i) = \{\epsilon\}$, $P(i, j)$ 表示 m_i 与 m_j 共指的概率。

4 实验和结果

本节首先介绍使用的数据集和实验设置, 然后进一步

展示实验结果并对结果进行分析。

4.1 数据集

本文使用 ACE2005 中文语料库。如表 2 所列,ACE 事件被分为 8 个大类,33 个小类。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。下文将介绍相关定义^[11]。

(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。

ACE 中文语料库共计 632 个文档,本文按照 Teng 等^[5]的方法移除不存在共指事件对的文档,最终共提取出 445 个文档。

4.2 实验设置

本文将得到的数据集按照 355:45:45 的比例划分成训练集、验证集和测试集。每个前馈神经网络隐藏层的大小设置为 500,符号特征的维度大小设置为 50,本文使用的 transformer 编码器为 BERT(bert-base-chinese)。在训练阶段,为了防止模型过拟合,dropout_rate 设置为 0.5,epoch 设置为 20,batch_size 设置为 8。遵循 Joshi 等^[12]提出的方法,本文为训练任务和 BERT 设置不同的学习率,任务学习率设置为 5×10^{-4} ,BERT 学习率设置为 5×10^{-5} ,并且在训练过程中两个学习率会线性衰减。

4.3 评估指标

本文按照 4 种标准的共指评估指标 MUC^[13],B³^[14],CEAF_e^[15]和 BLANC^[16]评估 GMNN 的性能,并且使用 4 个指标的平均值(AVG)作为对比指标。

4.4 实验结果

本文选用两个最先进的中文共指消解系统作为对比:1)文献[5]中使用传统模型进行文档内的事件共指消解的系统;2)文献[7]中使用门控注意力神经网络进行文档内事件共指消解的模型(GANN),这是目前最好的中文共指消解模型。此外,本文在 ACE2005 中文数据集上重新训练了 Lai 等^[9]的工作,这是目前在 ACE2005 英文数据集上共指消解效果最好的工作,本文将重新训练后的模型作为第三个对比模型。基线模型(Baseline)是本文只使用基础特征(触发词和事件论元)而不使用符号特征的模型;简单模型(Simple)是本文利用所有符号特征并使用简单连接策略(见式(6))的结果;GMNN 是使用事件基本特征(触发词和事件论元)和符号特征(事件子类型、极性、形态、泛型和时态),并采用门控连接策略(见式(10))完整的模型。表 3 列出了 ACE2005 中文语料库上 6 种模型的性能比较。

GMNN 在各评价指标上都取得了最好的结果。Teng 等^[5]只提供了模型的 MUC 分数,未提供其他指标分数。与传统模型如 Teng 等^[5]提出的模型相比,GMNN 将 MUC 分数提高了 10.31。Teng 等^[5]使用外部工具提取了大量特征,

而 GMNN 只使用语料库提供的少量特征。该结果表明,相比传统模型,神经网络模型在中文事件共指消解上具有更优越的性能。

相比 Cheng 等^[7]的门控注意力神经网络模型,GMNN 在各个指标上均有提升,在 AVG 指标上有 8.2 的提升,主要有两个原因:1)GANN^[7]使用 Word2Vec 编码大量的上下文特征,并利用门控注意力机制过滤噪声,而 GMNN 只使用预训练语言模型 BERT 编码少量必要的特征,并且没有额外引入门控注意力机制,却获得了更好的效果,该结果表明,相比传统编码器,预训练语言模型具有优越的性能,可以使用更少的特征获得更好的结果;2)GANN^[7]对所有符号特征采用简单的处理方式,直接使用独热编码将其向量化,而 GMNN 使用一种基于触发词和事件论元的门控机制,根据上下文语境过滤符号特征中的噪声,获取在特定上下文语境中更有用的信息。

Lai 等^[9]的模型是目前 ACE2005 数据集上效果最好的英文事件共指消解模型,它是一种基于神经网络的模型。本文重新训练了此模型,将它应用在中文数据集上。从表 3 可以看出,与 Lai 等^[9]的模型相比,GMNN 在 AVG 指标上提升了 2.66。

Baseline 是未使用符号特征只使用触发词特征和事件论元信息的模型,与使用基本特征(触发词和事件论元)和符号特征的简单连接模型 Simple 和完整的模型 GMNN 相比,在 AVG 指标上,Simple 和 GMNN 分别比 Baseline 提升了 7.58 和 9.35,这说明利用符号特征对解决中文事件共指是有益的。

为过滤符号特征中的噪声和解决特定语境下各种符号特征对解决事件共指贡献不同的问题,GMNN 采用了一种基于触发词和事件论元的门控连接策略。如表 3 所列,相比使用简单连接的 Simple 模型,GMNN 模型的 AVG 分数提升了 1.77,这表明了引入基于触发词和事件论元信息的门控机制的必要性。

表 3 5 种模型的性能比较

| Systems | MUC | B ³ | CEAF _e | BLANC | AVG |
|-----------------------|-------|----------------|-------------------|-------|-------|
| Teng 等 ^[5] | 73.50 | — | — | — | — |
| GANN ^[7] | 74.59 | 89.18 | 82.56 | 80.32 | 81.66 |
| Lai 等 ^[9] | 81.25 | 90.88 | 90.06 | 86.61 | 87.20 |
| Baseline | 67.70 | 89.66 | 86.72 | 77.82 | 80.51 |
| Simple | 80.82 | 93.36 | 90.74 | 87.44 | 88.09 |
| GMNN | 83.81 | 94.16 | 92.48 | 88.98 | 89.86 |

4.5 结果分析

为了进一步分析 GMNN 中每个组件的有效性,本文对模型的每个组件进行了对比研究。

(1)门控策略

本文探究门控策略对每个符号特征的影响,一共进行了 5 组对比实验,进一步说明门控策略的性能。每组实验分为两个实验,分别使用简单连接策略和门控连接策略,并且每组实验只使用一种符号特征。

如图 3 所示,与使用简单连接策略的模型相比,对每一个

符号特征使用门控连接策略的模型均获得了更好的效果。该结果表明,门控策略对存在噪声的或处在特定上下文语境中的信息具有更强的处理能力,相比简单连接策略它能够识别更多有用的信息。

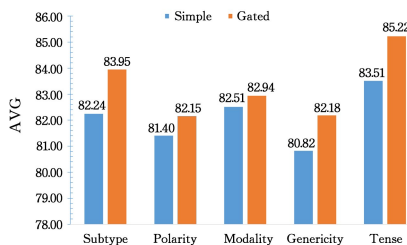


图3 门控策略的影响

Fig. 3 Impact of gated mechanism

(2) 事件论元

预训练语言模型的成功,允许事件共指消解模型利用更少的信息获取更好的模型性能。触发词信息是解决事件共指最基本、最有效的信息,也是必不可少的信息。图4给出了事件论元的效果。

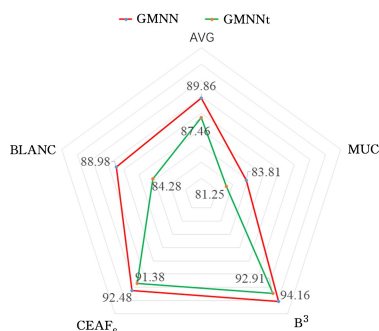


图4 基本特征的影响

Fig. 4 Impact of basic features

如图4所示,GMNN是使用触发词和事件论元作为基本特征的完整模型,GMNNt是只使用事件触发词作为基本特征的模型。两个模型均使用所有符号特征并采用门控策略连接基本特征和符号特征。不难发现,增加事件论元作为基本特征优化了GMNN的模型性能。该结果表明,事件论元信息对于解决事件共指具有重要意义。事件论元中不仅有事件发生的时间和地点可以直接帮助解决事件共指,还有事件的参与者等信息间接帮助事件共指消解。

(3) 符号特征

为说明每个符号特征的必要性,图5给出了每个符号特征对GMNN的影响。本文使用完整GMNN的结果作为参照,分别进行了5组实验,每组实验分别减少1个符号特征,使用其他4个特征。

如图5所示,减少每个符号特征都会直接降低GMNN的性能,这表明本文使用的每个符号特征对解决中文事件共指都有不同程度的贡献。其中,时态(Tense)对解决中文事件共指贡献最大,减少使用时态这一符号特征后,模型性能降低了2.91。如前文的句子示例S₃和S₄所示,ACE2005中文语料库中存在着大量这种因为时态而无法识别共指关系的事件

提及,引入时态这一符号特征恰好可以解决此类问题。

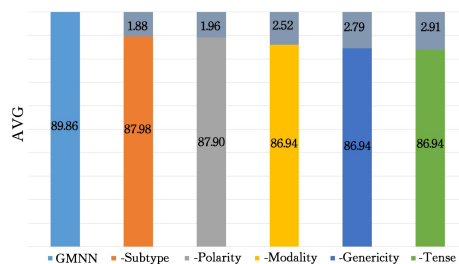


图5 符号特征的影响

Fig. 5 Impact of symbolic features

结束语 本文提出了一种基于事件基本特征的门控机制神经网络GMNN来解决文档内中文事件共指。相比已有的工作,GMNN使用预训练语言模型BERT编码少量必要的事件基本特征,并利用门控机制过滤符号特征中的噪声,在ACE2005中文语料库上取得了最佳的效果。

未来,我们将拓展模型,将其应用到更一般的场景下,具体工作包括3个方面。

(1) 预测事件基本属性。本文使用的事件属性是从ACE2005数据集中抽取的已经标注好的事件基本属性,不具有一般性。未来我们将利用ACE2005数据集训练一个事件基本属性预测模型,首先利用训练好的预测模型预测事件基本属性,然后利用GMNN进行事件共指消解研究,拓展GMNN在其他语料上进行研究。

(2) 端到端事件共指消解。本文直接利用语料库中标注好的事件信息进行事件共指消解研究,虽然可以验证模型的有效性,但是模型很难应用到现实任务中。未来我们将联合训练事件预测和事件共指消解两个任务,实现先事件抽取再共指消解,完全消除模型对标注数据的依赖。

(3) 跨语言、跨文档事件共指消解。本文针对ACE2005中文语料库研究文档内事件共指消解,ACE2005英文数据集与中文数据集结构相似,也可用于事件共指消解。此外,跨文档事件共指消解的研究意义更大。未来我们将拓展模型,实现跨语言、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 Lin-

- guistics and the 11th International Joint Conference on Natural Language Processing, Stroudsburg: ACL, 2021: 5768-5781.
- [5] TENG J Y, LI P F, ZHU Q M. Global inference for co-reference resolution between Chinese events [J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2016, 52(1): 97-103.
- [6] 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.
- [7] CHENG H Y, LI P F, ZHU Q M. Employing gated attention and multi-similarities to resolve document-level Chinese event coreference[C]// International Conference on Asian Language Processing, Piscataway, NJ: IEEE, 2019: 296-301.
- [8] HUAN M, CHEN H Y, LI P F. Resolving Chinese event coreference via structured representation [J]. Journal of Chinese Information Processing, 2019, 33(12): 19-27.
- [9] 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.
- [10] LEE K, HE L H, LEWIS M, et al. End-to-end neural coreference resolution[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2017: 188-197.
- [11] WEI P, CHAO W H, LUO Z C, et al. Selective expression approach based on event trigger for event coreference resolution on Twitter [J]. Computer Science, 2018, 45(12): 130-136.
- [12] JOSHI M, CHEN D Q, LIU Y H, et al. BERT for coreference resolution: baselines and analysis[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, Stroudsburg: ACL, 2019: 5803-5808.
- [13] 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.
- [14] BAGGA A, BALDWIN B. Algorithms for scoring coreference chains[C]// The First Conference Language Resources and Evaluation Coreference. ELRA, 1998: 563-566.
- [15] 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.
- [16] 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:喻黎)