

基于结构化表示的中文事件同指消解方法

宦敏 程昊熠 李培峰

(苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

摘要: 事件同指消解是自然语言处理中一个具有挑战性的任务, 它在事件抽取、问答系统和阅读理解中有着重要的作用。针对事件的语义信息主要由触发词和论元表示这一个特点, 本文将事件进行结构化表示并输入一个基于门控和注意力机制的模型 GAN-SR (Gated Attention Network with Structured Representation), 在文档内进行中文事件同指消解。首先, 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, 抽取事件句信息并表示为一个事件五元组。其次, 该模型将各种事件信息输入 GRU 进行编码, 然后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。该文在 ACE2005 中文语料库上的实验表明, GAN-SR 的性能优于目前性能最好的基准系统。

关键词: 中文事件同指消解; 结构化表示; 注意力机制

中图分类号: TP391

文献标识码: A

Resolving Chinese Event Coreference on Structured Representation

HUAN Min, CHENG Haoyi, and LI Peifeng

(School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu, 215006, China)

Abstract: Event coreference resolution is a challenging task in natural language processing and it plays an important role in many NLP tasks, such as event extraction, question answering and reading comprehension. Since the semantics of an event is mainly represented by its trigger and arguments, this paper introduces the structured representation of events to a neural network model GAN-SR (Gated Attention Network with Structured Representation) based on the gated and attention mechanism to perform Chinese event coreference resolution in document level. Firstly, it applies the semantic role labeling and dependency parsing to analyze the shallow semantics of events, and then uses a quintile to represent their structures. Secondly, we use GRU to encode various kinds of event information and then apply the multi-head attention mechanism to mine the important features between the events and event pairs. The experimental results on the ACE2005 Chinese corpus show that GAN-SR outperforms the state-of-the-art baselines.

Key words: Chinese event coreference resolution; structured representation; attention mechanism

0 引言

事件同指消解是信息抽取领域中的重要课题之一, 是自然语言处理中的一大难点。作为信息抽取的下游任务, 事件同指的准确识别有利于理解文本信息, 扩充文本隐含信息和发现事件之间的关系, 可应用于话题检测^[1], 信息抽取^[2]和阅读理解^[3]等领域。近年来, 在同指

消解这一大方向中, 大多数是实体同指研究, 事件同指的研究相对较少。

事件是描述信息的一个重要载体, 通常是指现实世界的某一人、物等在某一特定的时间、地点相互作用。事件同指消解旨在判断两个实例是否指向现实世界中的同一事件。一般情况下, 同一篇文档中具有同指关系的事件有着相同的参与者、时间和地点, 如例 1 和例 2 所示:

例1: 政府军从9月16日开始对阿布沙耶夫武装发动攻击。

例2: 9月16日以来政府军共与阿布沙耶夫武装交战40次。

事件句例1中的触发词是“攻击”，事件句例2中的触发词是“交战”，虽然两个触发词不完全一致，但是含义相似，都指向了现实世界中同一个“Attack”类型的事件。同时，两个事件句中的实施者都是“政府军”，受事者都是“阿布沙耶夫”，并且两个事件句具有相同的时间“9月16日”。因此，这两个事件句属于同指事件。通过判断事件句是否属于同一文档中，事件同指可分为跨文档事件消解和文档内同指消解。

本文的研究着眼于中文文档内的事件同指消解。近年来，在英文同指领域中，研究人员开始将神经网络模型广泛应用于事件同指的任务中，将任意两个事件句通过网络层组成事件对，判断两个事件句是否同指关系。Krause^[4]采用CNN网络，加入事件对特征，挖掘事件句的内部特征，但是只考虑事件句的局部信息，没有考虑事件句的长远距离信息与两个事件句之间的信息。Fang^[5]采用基于可分解注意力机制的神经网络模型，提取事件句之间的重要信息来挖掘事件句之间的关系。但是，往往存在着大量噪音，并且中文比英文的表述更加复杂，存在着一词多义、多种指代的现象，仅考虑事件句之间的信息还不够。

为了解决上述问题，本文提出了使用结构化语句、基于门控和注意力机制的方法来解决中文事件同指消解的问题：（1）使用自注意力机制和多头注意力机制挖掘事件句自身与事件对之间的重要信息；（2）采用语义角色标注与依存句法分析的方法，从触发词角度出发，挖掘事件句中的实施者、受事者、时间、地点等信息，从而过滤事件描述中的冗余信息，减少噪音。在ACE2005中文数据集上，本文方法的实验结果优于当前性能最优的基准系统。

1 相关工作

相比实体同指消解的研究，在事件同指消解上的研究还很少。事件同指消解任务最早是Ahn^[6]研究事件抽取时提出，通过研究事件对的有效特征，判断事件对之间的相似度，进而完成事件对的同指消解。

早期的事件同指消解的研究方法主要是由实体消解^[7]演变而来。随后将决策树^[8]，最大熵分类器^[9]、支持向量机（SVM）^[10]、深度神经网络^[11]等事件对二分类器应用于事件同指消解。之后，通过对语言学特点的研究，手动提取特征，对事件同指消解进行探索。Bejan^[12]使用非参贝叶斯网络对事件对的具体化特征建立事件同指消解模型。Liu^[13]使用手工提取的100多种特征，结合WordNet、FrameNet等外部知识信息进行实验。使用事件对模型往往会忽略事件对之间的联系，为了解决这种问题，滕佳月^[14]提出了中文事件同指消解全局优化模型。Peng^[15]等人提出采用语义角色标注来结构化表述一个事件并使用一种弱监督的算法解决事件同指问题。

如今，深度学习成为了自然语言处理领域的一种主流方法，且都达到了出色的效果。在英文事件同指消解任务上，Krause^[4]等使用卷积神经网络对事件对的两个句子，相对位置信息以及触发词的上下文建模，提取事件对的特征，引入事件特征最后得出事件同指结果。Fang^[5]采用基于可分解注意力机制的神经网络模型，结合事件级特征和事件对级特征，挖掘事件对之间的信息，得到事件是否具有同指关系。

2 事件同指消解模型

Krause^[4]使用CNN模型提取事件句信息，但CNN模型只考虑了词与词之间的局部信息，并没有考虑整个事件句内的信息。单独考虑事件句信息，无法挖掘事件对之间的信息。Fang^[5]提出采用注意力机制将事件句中相对重要的信息提取出来，但信息提取的过程中存在噪音，影响了对是否具有同指关系事件的判断。Peng^[15]提出的MSEP框架使用语义角色标注描

述事件，但事件句中的触发词不是都为谓语且并不都充当事件句的核心词。只通过使用语义角色标注抽取句子信息，会使得事件的描述存在较大的失误，进而影响判断事件同指的正确性。为了解决上述不足，本文提出了一个使用结构化语句和门控注意力机制的神经网络模型 GAN-SR (Gated Attention Network with Structured Representation)，如图 1 和图 2 所示。首先，基于事件句结构化规则将事件句结构化为事件五元组的表示形式；然后，将数据流通过 GAN-SR 模型，得到同指置信度，判断事件对是否具有同指关系。

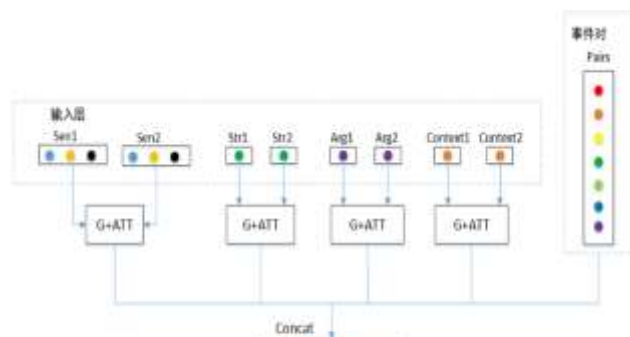


图 1 GAN-SR 框架

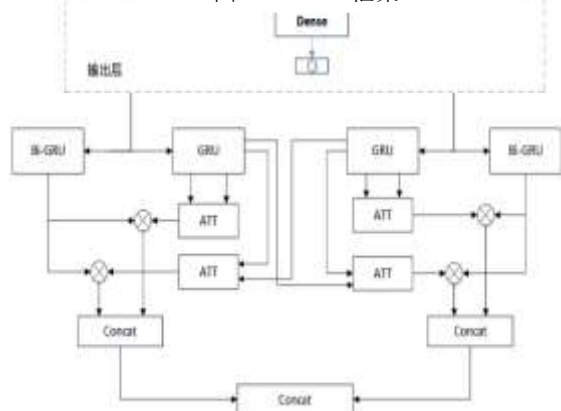


图 2 G+ATT 框架

GAN-SR 主要是通过门控处理长远距离依赖，使用注意力机制捕获事件句有利于同指判断的重要信息，并且结合事件对特征判断两个事件句是否具有同指关系。GAN-SR 主要包括以下三个部分：(1) 在输入层中，输入事件句，词性信息，位置信息，结构化语句，上下文信息，论元信息和事件对特征。其中，将事件句、词性信息和位置信息拼接起来；结构化语句是事先处理好的事件五元组；论元和触发词上下文信息是对结构化语句的补充。然后，使用

GRU 和 Bi-GRU 进行编码，捕获当前词之前词的特征以及前后词的特征。(2) 在注意力机制层中，使用了多头自注意力机制^[16]挖掘单个事件句自身的重要信息和事件对之间的重要信息。然后，通过门控来控制信息流，过滤噪音，减少噪音对模型判断的影响。(3) 在输出层中，通过激活函数得出置信度，如果置信度大于 0.5，判为同指，反之非同指。

2.1 事件句结构化表示

事件结构化是指从无结构的文本中抽取有价值的信息，然后通过一个填槽过程，用结构化的方式表示出来。在一个事件中，触发词是事件的核心，这个词代表事件的发生，通常是决定事件类型的重要特征，而论元一般则是指事件的参与者，是对事件信息的补充。

事件的论元包含了许多信息，往往具有同指关系的两个事件拥有相同的基本论元信息。本文利用这一特性，将事件信息用触发词和论元表示，最终得到结构化的事件句表示方式。

本文使用了哈工大社会计算与信息检索研究中心研制的语言技术平台 (LTP)^[17]。利用 LTP 对事件句进行浅层分析，主要进行语义角色标注和依存句法分析。

语义角色标注 (SRL) 是以句子中的谓词为中心，分析句子中各个成分与谓词之间的关

奥巴马	昨晚	在白宫	发表	了	演说
A0	TEM	LOC	发表		A1

图 3 SRL 实例

系，确定其他论元以及它们的论元角色，用谓词-论元的结构来描述句子。

图 3 句子中的谓词为“发表”，实施者 (A0) 为“奥巴马”，受事者 (A1) 为“演说”，时间 (TEM) 为“昨晚”，地点 (LOC) 为“在白宫”。

依存句法分析则是通过分析句子内的各个成分之间的依存关系构建句法结构。句子中的核心通常为谓语中的一个动词，其他成分直接或间接的与它产生依存 (修饰) 联系。句子中只有一个成分是独立的，即核心词，其他成分则都从属于其他成分，且任何成分都只依存一

一个成分。成分之间发生依存关系，通常使用一个依存树的结构来描述语句。

从图4中句子的依存句法分析可以看出，句子的核心词为“参加”，主语为“西门子”，宾语为“建设”，而不是“中国”，且“三峡”、“工程”与“建设”为定中关系，对“建设”信息进行了扩充。

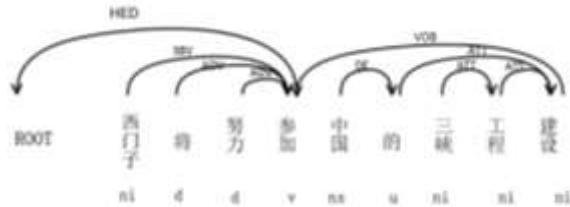


图4 依存句法实例

2.2 结构化规则

本文结合语义角色标注与依存句法分析进行事件表示，当语义角色标注无法按规则抽出事件信息时，则使用依存句法分析抽取信息，最后将事件句结构化表示。形式上，事件句可以表示为 E ，定义为一个五元组：

$$E = (Arg1, V, Arg2, T, L)$$

其中事件 E 中的元素称为事件元素，分别表示论元1 ($Arg1$)、动作 (V)、论元2 ($Arg2$)、时间 (T)、地点 (L)。

在结构化过程中，制定如下规则：

语义角色标注规则 (R1)：利用 LTP 的语义角色标注，基于判断触发词是否有 A0 和 A1，如若有 A0 和 A1 则进行五元组的信息抽取，直接获取主谓宾与时间地点信息，按槽填空。

在语义角色标注规则下，触发词往往为动词，并且有明确的主语与宾语信息，如下例3所示（空缺信息使用空格代替）：

例3：米洛舍维奇被迫逃离贝尔格勒。

$$E_3 = (\text{米洛舍维奇}, \text{逃离}, \text{贝尔格勒}, -, -)$$

依存句法分析规则 (R2)：如果语义角色标注为空，则从触发词位置出发使用依存句法分析，查询“SBV”、“VOB”等信息；如有，则直接对主语或宾语信息抽取论元。

在依存句法分析规则下，触发词往往为动词，但触发词并不是事件句的核心词，主语与

宾语信息并不一定都齐全，需要使用依存句法分析挖掘事件信息，如例4所示：

例4：在南斯拉夫米洛舍维奇下台之后欧洲联盟决定减缓对塞尔维亚的制裁。

$$E_4 = (\text{米洛舍维奇}, \text{下台}, -, -, \text{南斯拉夫})$$

动词映射规则 (R3)：如果 $Arg1$ 、 V 和 $Arg2$ 都为空，则抽取距离触发词最近的动词，采用 R1 和 R2 的规则基于动词进行事件信息的抽取。

在动词映射规则下，触发词往往是动名词，语义角色标注与依存句法分析的方法将触发词判为其他用于修饰的信息，如例5所示：

例5：刚刚宣誓就任的行政院长张俊雄也应邀参加成立典礼。

$$E_5 = (\text{张俊雄}, \text{参加}, \text{典礼}, -, -)$$

上下文规则 (R4)：基于以上规则， $Arg1$ 、 V 和 $Arg2$ 仍然为空，即事件句无法抽取论元信息，则对事件触发词的上下文进行抽取。

在上下文规则下，事件句结构不明确，触发词往往为动名词，且语义角色标注与依存句法分析无法准确分析触发词语义，如例6所示：

例6：至楼梯口下车道别回家。

$$E_6 = (\text{道别}, \text{回家}, -, -, -)$$

事件信息抽取成功之后，利用事件的五元组表示方式，进行填槽，将信息结构化表示。

事件句例1、2、7和8可以表示为 E_1 、 E_2 、 E_7 和 E_8 ：

例7：杨富家教授起程赴世界著名大学英国诺丁汉大学就任校长。

例8：杨富家教授1991年当选中科院院士。

$$E_1 = (\text{政府军}, \text{攻击}, \text{阿布沙耶夫}, \text{9月16日}, -)$$

$$E_2 = (\text{政府军}, \text{交战}, \text{阿布沙耶夫}, \text{9月16日}, -)$$

$$E_7 = (\text{杨富家教授}, \text{就任}, \text{校长}, -, \text{英国})$$

$$E_8 = (\text{杨富家教授}, \text{当选}, \text{中科院院士}, \text{1991年}, -)$$

将无结构的事件句，转换为简单的五元组，可以清晰简洁地捕获事件信息。由 E_1 、 E_2 ，可以容易看出，它们的触发词（“攻击”，“交战”）虽然不一样，但是语义相似，同时拥有相同的论元 $Arg1$ 、 $Arg2$ 和时间 T 。因此，这两个事件都指向现实世界的同一个事件（“Attack”事件），互为同指事件。

而事件7与事件8虽然有相同的论元 $Arg1$ ，与相似的动作 V ，但是 $Arg2$ （“校长”，“中科

院院士”)却是完全不同的论元信息。因此,这两个事件为非同指事件。

2.3 输入层

词语在句子中距离触发词的位置对事件同指决策同样有帮助,因此,本文还考虑了句中各个词语距离触发词的信息。另外词性信息可以有效解决部分词语一词多义的问题,判断出词语在句法中所充当的成分。此外,加入触发词上下文信息以及论元信息,能够有效地弥补抽取结构化事件五元组过程中产生的误差。

综上,本文的输入有:1)事件句 S , 2)事件句中每个词语离触发词的距离 Loc , 3)事件句每个词语的词性 Pos , 4)结构化的事件五元组 $E5$)标注的论元 Arg , 6)触发词上下文 Con 。

触发词是事件句的核心,可用于判断出事件类型。此外,本文还结合 ACE2005 中文语料库中的一些基础属性来辅助判别同指关系。事件的基础属性是对事件的全局描述,一般具有同指关系的事件实例拥有相同的基础属性。因此,利用语料库中的标注信息来过滤负例(非同指事件对),这有利于模型的准确判断,这些基础属性包括事件类型($type$)、事件子类型($subtype$)、形态(mod)、极性(pol)、泛型(gen)、时态($tense$)和触发词($trig$)。

由于事件句中的词语都是相对独立,以上输入信息可用词袋模型组织。本文采用相同的方法将输入信息嵌入模型,使用维基百科 300 维作为训练矩阵 M ,对事件句 S 、事件五元组 E 、论元 Arg 、触发词上下文 Con 进行编码。一个句子中有 n 个词语,基于训练矩阵 M 得到每个词语的词向量 $W_i(0 < i < n)$ 。然后将事件句中的所有词向量拼接成句子向量 Sen ,如式(1)所示:

$$Sen = (W_1, W_2, W_3, \dots, W_n) \quad (1)$$

将需要判断是否有同指关系的两个事件句采用上述方法对事件句信息 S 编码,得到向量 Sen_1, Sen_2 (本文中的下标 1 和 2 分别表示事件句 1 和事件句 2)。使用同样的方法对词性信息 Pos 和位置信息 Los 生成向量 Pos_1, Pos_2

和 Loc_1, Loc_2 。然后将事件句向量和词性信息、位置信息拼接成新的事件向量 S_1, S_2 :

$$S_i = \text{Concat}(Sen_i, Pos_i, Loc_i) \quad (i = 1, 2) \quad (2)$$

分别使用 GRU 和 Bi-GRU 对事件句 S_1, S_2 , 事件五元组 E_1, E_2 , 论元 Arg_1, Arg_2 和触发词上下文 Con_1, Con_2 进行编码,得到事件向量 Se_1, Se_2, Seb_1, Seb_2 , 事件五元组向量 $Str_1, Str_2, Strb_1, Strb_2$, 论元向量 A_1, A_2, Ab_1, Ab_2 和触发词上下文 C_1, C_2, Cb_1, Cb_2 。

另外,将事件对的基本属性拼接成一个事件对向量,如式(3)所示:

$$P = \text{Concat}(type, subtype, mod, pol, gen, tense, trig) \quad (3)$$

2.4 注意力机制层

最近几年,注意力机制在深度学习领域被广泛使用。根据查询信息,计算信息中的每个元素和这个查询信息的匹配程度,然后通过一个 softmax 函数生成注意力分布,最后根据注意力分布计算信息的加权平均。

本文使用多头注意力机制,旨在探寻单个事件句自身的重要特征,与两个事件句之间的重要特征。

首先,将输入层的输出分别输入到多头注意力机制中,生成自注意力和句间注意力:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

$$head_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (5)$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(head_1, \dots, head_n)W^O \quad (6)$$

自注意力编码过程中,先计算 Q 和每个 K 的点乘,然后除以一个缩放因子,最后用 softmax 函数计算权重。句间注意力编码则是先对 Q, K, V 做一次映射,然后重复使用点积注意力机制计算出结果,将得到的结果进行合并,最后将结果合并进行线性变化。多头注意力机制与传统注意力机制相比,可以允许模型在不同的维度和表示子空间里学到相关信息。

如式(7)所示,探寻句子自身的重要信息,再与 $Seb_1, Seb_2, Strb_1, Strb_2, Ab_1, Ab_2, Cb_1, Cb_2$ 依次进行元素相乘,放大词语之间的向量差距,进一步增强事件句的重要信息,进而得到事件句向量 ST_1, ST_2 , 事件五元组向量 StT_1, StT_2 , 论元向量 AT_1, AT_2 , 触发词上下文向量

CT_1, CT_2 :

$$ST_i = Seb_i \cdot \text{Attention}(Se_i, Se_i, Se_i) \quad (i=1,2)$$

$$StT_i = Strb_i \cdot \text{Attention}(Str_i, Str_i, Str_i) \quad (i=1,2)$$

$$AT_i = Ab_i \cdot \text{Attention}(A_i, A_i, A_i) \quad (i=1,2)$$

$$CT_i = Cb_i \cdot \text{Attention}(C_i, C_i, C_i) \quad (i=1,2) \quad (7)$$

如式(8)所示,探索事件句之间的重要信息,同样与 $Seb_1, Seb_2, Strb_1, Strb_2, Ab_1, Ab_2, Cb_1, Cb_2$ 依次进行元素相乘,得到事件句向量 SMT_1, SMT_2 , 事件五元组向量 $StMT_1, StMT_2$, 论元向量 AMT_1, AMT_2 , 触发词上下文向量 CMT_1, CMT_2 :

$$\begin{aligned} SMT_i &= Seb_i \cdot \text{Attention}(Se_i, Se_j, Se_i) \\ StMT_i &= Strb_i \cdot \text{Attention}(Str_i, Str_j, Str_i) \\ AMT_i &= Ab_i \cdot \text{Attention}(A_i, A_j, A_i) \\ CMT_i &= Cb_i \cdot \text{Attention}(C_i, C_j, C_i) \\ &(i, j \in \{1,2\}, i \neq j) \end{aligned} \quad (8)$$

最后将所得的结果拼接,分别经过 \tanh 和 sigmoid 函数的全连接层,使用两种函数进行互补,过滤信息流,得到事件句向量 SU_i , 事件向量五元组 StU_i , 论元向量 AU_i , 触发词上下文向量 CU_i :

$$\begin{aligned} St_i &= \tanh(W^t(ST_i, SMT_i)) \quad (i=1,2) \\ Ss_i &= \text{sigmoid}(W^s(ST_i, SMT_i)) \quad (i=1,2) \\ SU_i &= St_i \cdot Ss_i + (1 - St_i) \cdot Ss_i \quad (i=1,2) \end{aligned} \quad (9)$$

其中,按式(9)步骤得到事件向量五元组 StU_i , 论元向量 AU_i , 触发词上下文向量 CU_i 。

最后将上述的信息流通过全局最大池化,得到事件向量 SX_i , 事件五元组向量 StX_i , 论元向量 AX_i , 触发词上下文向量 CX_i :

$$\begin{aligned} SX_i &= \text{GlobalMax}(SU_i) \\ StX_i &= \text{GlobalMax}(StU_i) \\ AX_i &= \text{GlobalMax}(AU_i) \\ CX_i &= \text{GlobalMax}(CU_i) \quad (i=1,2) \end{aligned} \quad (10)$$

2.5 输出层

本文将通过最大池化层的向量和事件基础属性进行拼接:

$$V = \text{Concat}(SX_1, SX_2, StX_1, StX_2, AX_1, AX_2, CX_1, CX_2, P)$$

最后将向量 V 放入以使用 relu 激活函数的全连接分类器中,通过 sigmoid 层得到事件同指的置信度:

$$\text{score} = \text{sigmoid}(\alpha(W \cdot V + b)) \quad (12)$$

本文模型使用了 dropout 和 normalization 来防止过拟合。最终将判为同指的事件对组成一个事件链,利用同指事件的自反性、对称性和传递性,进一步优化结果。

3 实验

3.1 实验设置

ACE2005 中文语料库是文档级别中文事件同指消解任务的唯一语料库。由于语料库中有一半的文档内没有同指事件,导致大量的负例(非同指事件对)使得正负比例失调。因此本文参照滕佳月^[18]的方法将语料库中不含同指事件对的文档剔除,以文档为单位,剔除不含有事件的句子,再将事件组成事件对,剔除事件类型不同的事件对,使得同指事件正负比例约为 1:4。

本文参照滕佳月^[14]的方法采用 5 倍交叉验证的方法,在每倍中按照 8:1:1 的比例划分出了训练集、验证集和测试集。本文使用 4 个标准的事件同指消解评测指标用于模型的测评,分别为: $MUC^{[19]}$ 、 $B^{3[20]}$ 、 $BLANC^{[21]}$ 和 $CEAF_c^{[22]}$ 。

MUC: MUC 为事件同指消解最主要的评测指标,主要基于同指链中同指的边数量来评判模型性能,但没有考虑非同指,也没有区别对待不同同指链中的错误。

B³: B³ 针对 MUC 的缺陷,使用了同指链中的实例作为计算目标,但对非同指事件处理不佳。

BLANC: 同时考虑了同指与非同指的关系。

CEAF_c: 与 B³ 算法类似,但在实例映射中使用了 Kuhn-Munkres 算法,解决了 B³ 的缺陷。

综合使用 MUC、B³、BLANC 和 CEAF_c 四种评测标准进行评测,将会从多种角度衡量模型的性能,因此更为客观。

在实验中,本文使用维基百科的预训练向量,词向量维度设为 300,词性和位置向量设为 50,Dropout 的值为 0.2。另外,编码层中 GRU 的神经元数为 100, Bi-GRU 的神经元数为 50,模型训练批次为 20 轮。

3.2 实验结果

为了验证本文的 GAN-SR 模型在中文事件消解任务上的性能, 本文将 Krause^[4]和 Fang^[5]的神经网络模型分别在 ACE2005 中文语料库上复现, 作为本文的基准系统, 与 GAN-SR 模型做对比。表 1 指出了两个基准系统和 GAN-SR 的性能比较, 从表 1 可以看出:

表 1 模型性能比较

System	MUC	B ³	BLANC	CEAF _e	AVG
Krause	66.52	86.08	75.31	77.34	76.31
Fang	69.87	87.77	77.34	80.53	79.00
GAN-SR	76.13	90.86	82.92	84.81	83.68

与 Krause^[4]的系统相比, 本文的系统在四个指标上都有很明显的提升, 平均提升了 7.37%。这是因为 Krause^[4]的系统对事件句只使用卷积神经网络进行操作, 这使得系统只提取了事件句的局部信息, 并且没有考虑事件句的长远距离依赖, 同时没有筛选出事件的重要特征。本文使用注意力机制挖掘了事件句和事件句之间的重要信息, 而且加入事件五元组, 向模型提供了事件对是否同指的重要信息。这说明 GAN-SR 模型中的注意力机制和事件五元组信息的有用性。

与 Fang 的系统相比, 在 MUC 指标上提高了 6.26%, 在其他的三个指标上, 也分别有了很大的提高, 4 个指标平均提高了 4.68%。这是因为本文除了使用注意力机制挖掘事件句的重要信息外, 还考虑了事件对之间的重要信息, 并且对提取的重要特征进行了过滤, 进而减少噪音对系统判断的影响。

总体来说, GAN-SR 模型相比其他几个模型, 在 MUC 指标上的提升最为明显。这是由于本文的模型在加入事件五元组和事件对之间的注意力之后, 使得模型对正例(同指事件)的判断更为准确。

3.3 实验分析

和基准系统相比, 本文系统性能的提升主要源于事件五元组的加入和自身注意力与事件对之间注意力信息的结合。

如果将两个基准系统加入结构化的事件五

元组信息后, 表 2 显示 MUC 都有 5%左右的提升, 总体指标平均提高 2.5%左右。GAN-SR 模型是在 GATT 模型基础上加入结构化的事件五元组信息。与 GATT 模型相比, GAN-SR 引入事件五元组使得 MUC 指标提升了 1.17%, 总体指标平均提高 0.78%。由此可见, 结构化的事件句信息的加入, 更加有利于同指事件的判断。这是因为, 去除冗余信息的事件五元组表述更简洁明了, 相对重要信息的突出, 更有利于模型的学习。

表 2 对比实验

System	MUC	B ³	BLANC	CEAF _e	AVG
Krause	66.52	86.08	75.31	77.34	76.31
+Str	71.63	81.21	81.15	81.46	78.86
Fang	69.87	87.77	77.34	80.53	79.00
+Str	74.89	88.97	80.03	82	81.47
GATT	74.96	90.33	81.88	84.2	82.84
GAN-SR	76.13	90.86	82.92	84.81	83.68

在 R1 规则下, 事件结构化准确抽取了事件句的实施者、受事者、时间地点等信息。加入 R2 规则, 弥补由于触发词不是事件句中的核心词所导致的信息不完全性。由于在语料库中触发词的词性并不都为动词, 存在许多词性为动名词的触发词, 因此采用 R3、R4 规则寻找动名词所在的主句的主要信息, 增加了事件句表述的重要信息。

因为同指的事件往往有着相同的参与者、时间、地点等信息, 事件五元组信息的加入更加有利于判断事件对之间的同指关系或非同指关系。如例 7、8 所示, 虽然事件的参与者是“杨富家教授”, 如果没有论元 Arg2、时间 T 与地点 L 信息的补充, 本文的模型会判断这一组事件为同指关系, 因为这两个事件句的触发词的语义相近并且属于同一个事件类型。但事实上, 事件触发词的宾语信息却完全不同。事件主要信息的加入, 使得模型判断同指关系更为准确。

因此, 结构化的事件五元组信息对识别同指事件是有效的。另外, 论元信息、触发词上下文信息的加入是为了弥补抽取事件五元组过程中的失误。

为了进一步证明事件句自身注意力与事件对之间的注意力信息加入有利于本文模型性能

的提升, 本文分别去除事件句自身注意力与事件对之间的注意力信息与 GATT 模型进行比较。

GATT-self: 在 GATT 模型基础上去除自身注意力信息;

GATT-con: 在 GATT 模型基础上去除事件对之间注意力信息;

GATT-all: 在 GATT 模型基础上去除自身与事件对之间注意力信息。

性能对比为如表 3 所示。与 GATT 模型相比, GATT-con 和 GATT-self 模型在 MUC 性能上分别下降了 2.25%、2.48%, 总体指标平均下降 1.52%、1.7%。

表 3 性能对比实验

System	MUC	B3	BLANC	CEAF _F	AVG
GATT	74.96	90.33	81.88	84.2	82.84
GATT-con	72.71	89.88	79.83	82.85	81.32
GATT-self	72.48	89.49	79.84	82.76	81.14
GATT-all	71.62	89.75	79.57	82.95	80.97

由此可见, 事件句自身和事件对之间的注意力信息对模型判别两个实例是否同指起着很大的作用, 因为注意力机制可以提取事件句的重要信息。事件对之间往往存在着对同指判断有价值的信息, 事件句自身的重要信息更有利同指的识别。

与 GATT 模型相比, GATT-all 去除注意力机制, MUC 性能下降 3.34%, 总体指标平均下降 1.87%。没有注意力机制的模型有利于处理较短的输入句, 但当输入句较长时, 句子中的所有语义通过一个中间语义向量来表示, 词语自身信息的消失导致丢失很多细节信息, 而引入注意力机制可以有效地缓解这一问题。

由此可见, 在考虑事件句自身以及事件对之间的重要信息, 对同指事件的判断是有效的。进一步抽取事件句的主要信息, 使用结构化的事件五元组表示方法在中文事件同指消解的任务上是有帮助的。

4 总结与展望

本文提出使用结构化语句来进行中文事件同指消解, 并且采用基于门控注意力机制的模型

来解决文档内中文事件同指任务。模型中引入事件五元组将事件句结构化, 提取事件句中的重要信息, 同事件句及其他基础特征信息, 通过 GRU 和 Bi-GRU 进行编码, 然后将事件句和事件对之间的重要信息进行挖掘与过滤, 最后结合事件对特征来加强模型对判断事件是否有同指关系的性能。实验表明, 本文的系统取得了明显的提升, 获得了最优的效果。接下来的工作将探讨基于少量人工标注, 面对生语料的事件同指消解任务。

参考文献

- [1] Mohammed M. Fouad, Marwa A. Atyah. Efficient Topic Detection System for Online Arabic News[J]. International Journal of Computer Applications, 2018, 180(12):7-12.
- [2] 程良, 郜洪奎, 王红斌. 基于依存树与规则相结合的汉泰新闻事件要素抽取方法[J]. 软件导刊, 2018, 17(07):53-60+67.
- [3] Swabha Swayamdipta, Ankur P. Parikh, Tom Kwiatkowski. Multi-Mention Learning for Reading Comprehension with Neural Cascades[C]//Proceedings of ICLR. 2018.
- [4] Sebastian Krause, Feiyu Xu, Hans Uszkoreit and Dirk Weisßenborn. Event Linking with Sentential Features from Convolutional Neural Networks[C]//Proceedings of Signll Conference on Computational Natural Language Learning. 2016:239-249.
- [5] Jie Fang, Peifeng Li, Guodong Zhou. Employing Multiple Decomposable Attention Networks to Resolve Event Coreference[C]//Proceedings of CCF International Conference on Natural Language Processing and Chinese Computing. 2018:246-256.
- [6] David Ahn. The Stages of Event Extraction[C]// Proceedings of the Workshop of the ACL on Annotating and Reasoning about Time and Events, 2006: 1-8.
- [7] Haghghi Aria, Klein Dan. Simple Coreference Resolution with Rich Syntactic and Semantic Features[C]// Proceedings of EMNLP, 2009:1152-1161.
- [8] Agata Cybulska and Piek Vossen. Translating Granularity of Event Slots into Features for Event Corefer-

- ence Resolution[C]//Proceedings of the Third Workshop on EVENTS: Definition, Detection, Coreference, and Representation, 2015: 1–10.
- [9] Zheng Chen, Heng Ji, and Robert Haralick. A Pairwise Event Coreference Model, Feature Impact and Evaluation for Event Coreference Resolution[C]// Proceedings of the Workshop on Events in Emerging Text Types, 2009: 17–22.
- [10] Chen Chen and Vincent Ng. SinoCoreferencer: An End-to-End Chinese Event Coreference Resolver[C]// Proceedings of the Ninth Language Resources and Evaluation Conference, 2004: 4532–4538.
- [11] Thien Huu Nguyen, Adam Meyers, and Ralph Grishman. New York University 2016 System for KBP Event Nugget: A Deep Learning Approach[C]//Proceedings of the Text Analysis Conference, 2016: 45–52.
- [12] Cosmin Adrian Bejan and Sanda Harabagiu. Unsupervised Event Coreference Resolution. [J]. Computational Linguistics, 2014, 40(2):1412-1422.
- [13] Zhengzhong Liu, Jun Araki, Eduard Hovy, and Teruko Mitamura. Supervised Within Document Event Coreference Using Information Propagation[C]// Proceedings of the Ninth Language Resources and Evaluation Conference, 2014: 4539–4544.
- [14] 滕佳月, 李培峰, 朱巧明. 基于全局优化的中文事件同指消解方法[J]. 北京大学学报(自然科学版), 2016, 52(1):97-103
- [15] Haoruo Peng, Yangqi Song, and Dan Roth. Event Detection and Co-reference with Minimal Supervision[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2016: 392–402.
- [16] Ashish Vaswani, Noam Shazeer, and Niki Parmar. Attention Is All You Need[C]//Proceedings of NIPS 2017, 2017:61-75.
- [17] Wanxiang Che, Zhenghua Li, Ting Liu. LTP: A Chinese Language Technology Platform[C]//Proceedings of the Coling 2010: 13-16.
- [18] 滕佳月. 中文事件同指消解方法研究[D]// 硕士论文, 2016:18-29.
- [19] Marc Vilain, John Burger, John Aberdeen, Dennis Connolly, and Lynette Hirschman. A Model-Theoretic Coreference Scoring Scheme[C]//Proceedings of MUC-6, 1995.
- [20] Amit Bagga and Breck Baldwin. 1998. Algorithms for Scoring Coreference Chains[C]//Proceedings of LREC 1998, 1998: 563–566.
- [21] Marta Recasen and Eduard Hovy. Blanc: Implementing the Rand Index for Coreference Evaluation. [J]. Natural Language Engineering, 2011:485–510.
- [22] Xiaoqiang Luo. On Coreference Resolution Performance Metrics[C]//Proceedings of EMNLP 2005, 2005:25-32.