

基于注意力机制的事件同指消解方法

程昊熠 李培峰 朱巧明

(苏州大学计算机科学与技术学院 江苏 苏州 215006)

(江苏省计算机信息技术处理重点实验室 江苏 苏州 215006)

摘要 事件同指消解是一项具有挑战性的自然语言处理任务,它在事件抽取、问答系统、阅读理解中有着重要的作用。文中提出了一种基于全局和局部信息,并具有全局推理机制的可分解注意力神经网络模型 DANGL(Decomposable Attention Neural network based on Global and Local information),用于文档级的事件同指消解。神经网络模型 DANGL 与过去大部分以概率模型和图模型为基础的传统方法之间存在很大的区别。DANGL 首先使用 Bi-LSTM 和 CNN 分别获取每个事件句的全局信息和局部信息;然后使用可分解注意力网络获取每个事件句中相对重要的信息;最后使用文档级全局推理模型进一步优化同指链。在 TAC-KBP 语料库上的实验显示,DANGL 使用了少量的特征,且平均性能优于目前最好的基准系统。

关键词 事件同指,全局和局部信息,可分解注意力网络,全局推理

中图法分类号 TP391.1 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.09.029

Event Coreference Resolution Method Based on Attention Mechanism

CHENG Hao-yi LI Pei-feng ZHU Qiao-ming

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

(Provincial Key Laboratory for Computer Information Processing Technology, Suzhou, Jiangsu 215006, China)

Abstract Event coreference resolution is a challenging NLP task. It plays an import role in event extraction, QA system and reading comprehension. This paper introduced a decomposable attention neural network model DANGL with global inference mechanism based on remote and local information to document-level event coreference resolution. The neural network model DANGL is quite different from most traditional methods based on probabilistic models and graph models in the past. DANGL first uses Bi-LSTM and CNN to capture both the remote information and the local information of each event mention. Then, it applies the decomposable attention network to capture relatively important information in event mention. Finally, it employs a document-level global inference mechanism to further optimize the coreference chains. Experimental results on TAC-KBP show that DANGL uses a few features and outperforms the state-of-the-art baseline.

Keywords Event coreference, Remote and local information, Decomposable attention network, Global inference

1 引言

同指消解是自然语言处理中的一个难点。近年来,针对实体同指消解的研究较多,而针对事件同指消解的研究相对较少,因此事件同指消解仍然是一项具有挑战性的工作。事件同指消解是自然语言理解的主要基础之一,如话题检测^[1]、信息抽取^[2]和阅读理解^[3]等都需要用到事件同指消解。

事件同指消解是在判断文本中的两个事件句是否指向了真实世界中的同一个事件的基础上,将它们链接到同一个事件链。以事件句 S_1 和 S_2 为例:

S_1 : The Yida refugee camp in South Sudan was {bombed} on Thursday.

S_2 : Two bombs {fell} within the Yida camp, including

one close to the school.

事件句 S_1 中的触发词是“bombed”,事件句 S_2 中的触发词是“fell”,两个事件句均指向了现实生活中的同一个 Conflict 类型的事件,因此它们被聚合到了同一个事件链中,属于同指关系。

根据事件句是否在同一文档中,可以把事件同指消解任务划分为文档内事件同指消解和跨文档事件同指消解。本文以文档内事件同指消解为研究对象。在以往的研究中,通常使用基于概率的模型或者基于图的模型^[4],这些方法通常不能提取出事件句的全局信息、事件句内单词与单词之间的局部信息以及事件句中相对重要的部分。同时,它们不能提取出触发词之间、论元之间以及事件结构之间的语义信息。近年来,人们将事件同指消解看作基于事件对的事件同指消

到稿日期:2018-08-24 返修日期:2018-10-12 本文受国家自然科学基金(61836007,61472265,61772354)资助。

程昊熠(1995-),男,硕士生,主要研究领域为自然语言处理;李培峰(1971-),男,教授,硕士生导师,主要研究领域为自然语言处理和机器学习,E-mail:pfli@suda.edu.cn(通信作者);朱巧明(1963-),男,教授,博士生导师,主要研究领域为中文信息处理。

解,即把任意两个事件句放在一起组成一对事件对,并判断它们二者是否是同指关系。但这种方法也存在弊端,即各个事件对之间可能会产生相互的冲突,从而影响对事件对是否同指的判断。尽管目前可以施加规则来减轻这种问题的干扰,但这毕竟不是长久之计,如果更换了语料库,规则也要随之改变。

为了解决上述问题,本文设计了基于全局和局部信息的可分解注意力神经网络模型 DANGL 来解决事件同指消解问题。该模型通过 Bi-LSTM 和 CNN 提取全局和局部信息,使用注意力机制提取重要信息,并使用全局优化模型减轻事件对间的冲突问题。在 TAC-KBP 数据集上的实验结果显示, DANGL 模型优于当前性能最优的基准系统。

2 相关工作

目前,相对于对实体同指消解^[5]的研究,对事件同指消解的研究还很少。事件同指消解由实体同指消解演变而来,研究方法同实体同指消解类似。事件同指消解一般分为文档内事件同指消解和跨文档事件同指消解。

目前,关于跨文档事件同指消解的研究还很少。Bejan 等创建了 ECB (Event Coreference Bank) 语料库,这是为跨文档事件同指消解任务提供的首个语料库。同时,他们也提供了一系列无监督型的贝叶斯方法。Cybulska 和 Vossen 指出 ECB 语料库的词汇不具有多样性,基于此,他们建立了一个更庞大的 ECB+ 语料库。目前,该领域大部分的研究都使用扩展型贝叶斯模型,如无参数型贝叶斯模型^[6]和分层贝叶斯模型^[7]。

早期的文档内事件同指消解研究都是建立在实体同指消解的基础上的,使用的是传统机器学习的方法,如基于概率的模型和基于图的模型等。Chen 等^[8]提出了一种基于一系列

事件对属性(如触发词、时态、极性)的事件对同指消解分类器。Chen 等^[9]提出了一种基于图模型的聚类方法,他们使用最小图切割的方法来处理事件同指链,将不属于该链的事件句剔除。Liu 等^[10]从事件句中提取出 100 多种特征放入 SVM 事件对分类器中进行实验。Lu 等^[11]提出了一种基于马尔科夫链的联合推理模型,用于纠正分类器产生的错误结果。Liu 等^[12]采用了基于图的传统方法,将非同指的事件句从图中切割出去,而且该方法引入了许多特征,如框架特征、论元特征、事件句距离等。

如今,深度学习的方法已经成为自然语言处理领域的一种主流方法。神经网络模型在问答系统、相似度比较等任务中都取得了不错的效果。Krause 等^[13]首次使用神经网络模型在事件同指消解任务上进行了尝试,并取得了不错的效果。首先用 CNN 和池化层对事件句和触发词的上下文进行处理,提取出特征,然后引入一些事件对之间的特征来辅助判别事件对之间是否存在同指关系。该方法也成为了目前的主流方法。Choubey 等^[14]先用全连接层对两个事件句进行了变维操作,然后计算两个事件句的 Cosine 距离和欧氏距离,最后通过激活函数得出一个置信度来判定同指关系。

3 基于全局和局部信息的注意力网络

Krause 提出的模型对事件句使用 CNN 来提取出单词的上下文特征信息,只考虑了事件句中单词与单词间的局部信息,并未注意到上下文对同指判断的影响。此外,Choubey 只将事件句放入一个全连接层进行变维操作,并未从事件句中提取出很好的特征信息。为了解决以上问题,本文提出了一种基于全局和局部信息的可分解注意力神经网络模型 DANGL,其架构如图 1 所示。

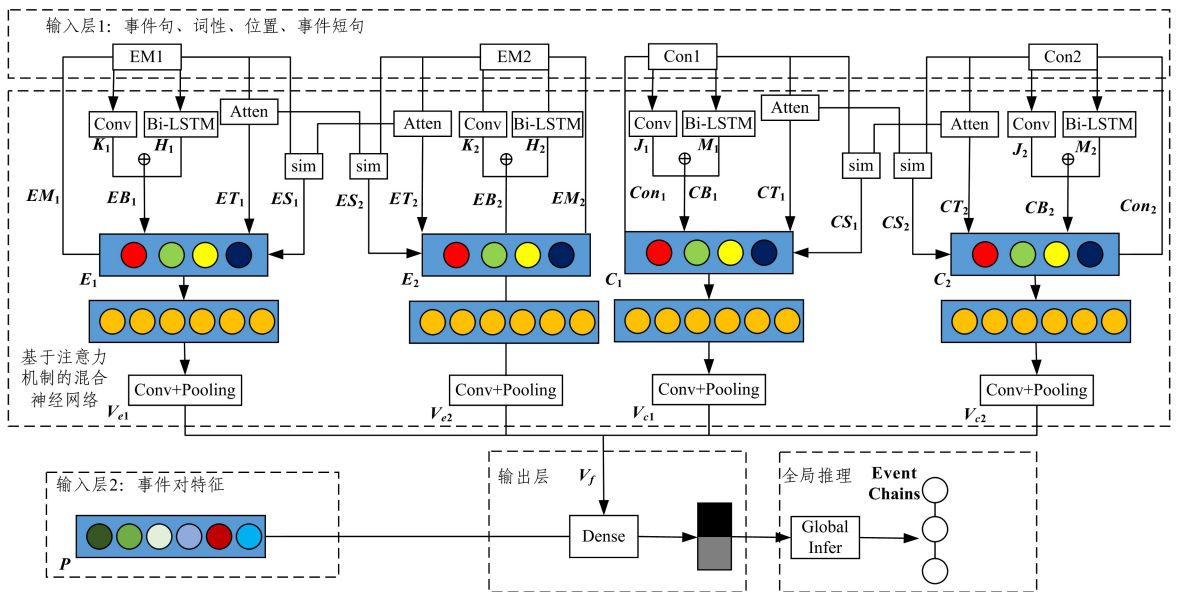


图1 DANGL 结构图

Fig. 1 Structure diagram of DANGL

DANGL 主要通过挖掘事件句的特征信息并用事件对特征来辅助判别两个事件句是否同指。该模型主要分为以下 4 个部分。1) 在输入层中输入事件句、词性信息、位置信息、事件短句和事件对特征,并将事件句、词性信息和位置信息拼接起来。2) 在基于注意力机制的混合神经网络层,使用 CNN

和 Bi-LSTM 来提取事件句的局部和全局信息。CNN 可提取出事件句中相邻单词间的局部信息,从而挖掘出事件句内上下文之间的语义特征,但它的上下文范围有限(取决于设置的移动窗口的大小)。Bi-LSTM 恰好解决了上述问题,它的前馈系统可以将事件句的特征信息从开头一直传递到结尾,能

够将每个单词的上文信息很好地保留并传递下去,同时它的后馈系统将每个单词的下文信息从事件句的结尾传递到开头,因此,Bi-LSTM 提取出了事件句的全局信息;此外,它还使用了可分解注意力机制,该机制可以简化特征信息,筛选出事件句中相对重要的特征信息,从而为后面的计算提供方便,通过计算两个事件句的余弦(Cosine)距离来获取二者的相似度。3)在输出层中,通过激活函数得出置信度,若置信度大于0.5,则判定两事件句为同指关系。4)在全局推理层中,对模型的判定结果进行优化并修正错误,将同指的事件句放入同一个事件链中。

3.1 输入层

单词在事件句中的相对位置对判断事件同指有帮助,因此给每个单词赋予一个它们各自到触发词的距离,并将该距离作为位置信息特征。此外,单词的词性信息在描述事件中有着重要的作用,它可以判断出一个句子中每个词所扮演的语法角色。例如,表示人、事物、地点或抽象概念的名称就是名词;表示动作或状态变化的词为动词;用来修饰名词性成分的词就是形容词,等。同时,它与事件句中单词的位置信息一样,和事件句中的单词都是一对一的关系。因此,本文将事件句、事件句中每个单词的词性、事件句中每个单词的位置、事件短句作为输入。

触发词是一个事件句的核心,根据它能够判断出事件的类型。因此,本文以触发词为中心,在它之前和之后各取3个单词形成一个新的事件短句来描述这个事件。

此外,KBP 语料库还提供了事件类型、事件子类型和时态信息。本文利用这些特征组成事件对匹配特征:1)事件类型是否一致(*type*);2)事件子类型是否一致(*subtype*);3)形态是否一致(*mod*);4)触发词原型是否一致(*trig*)。

事件句、事件短句和词性都是基于词袋模型的。词袋模型忽略了词序、语法和句法,将一个事件句仅仅看作一个词集合,即句中的单词都是相互独立的。本文采用了相同的词嵌入方法,使用了预训练的 glove 的 50 维词向量矩阵作为训练矩阵 M ,并对事件句、事件短句和词性标注特征进行编码。例如,对于事件句中的每个单词(n 个单词),基于训练矩阵 M 得到每个单词的词向量 W_i ,然后将所有词向量合并到一起形成句子向量 Sen :

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

对于需要识别是否存在同指关系的两个事件句 S_1 和 S_2 ,采用以上方法得到事件句向量 Sen_1 和 Sen_2 、事件短句向量 Con_1 和 Con_2 ,以及词性标注特征向量 Pos_1 和 Pos_2 。对于单词的位置信息,使用随机的词嵌入矩阵将它随机映射成 100 维的位置向量 Loc_1 和 Loc_2 。

最后,将事件句向量、词性标注特征向量和位置向量拼接在一起,形成新的事件句向量 EM_1 和 EM_2 ,如式(2)所示:

$$EM_i = Concat(Sen_i, Pos_i, Loc_i) \quad (i=1,2) \quad (2)$$

另外,事件对匹配特征的值为 0 或 1(0 表示不一致,1 表示一致),将事件对特征和两个事件句在文档内的距离 dis 融合成一个向量 P ,如式(3)所示:

$$P = Concat(type, subtype, mod, dis, trig) \quad (3)$$

3.2 基于注意力机制的混合神经网络

由于事件句的单词间存在相互作用,本文使用 CNN 来

挖掘出事件句中相邻单词间的特征,以获取单词间的交互信息,从而获得事件句的局部信息。同时,事件句中的部分单词仅被用来修饰核心词(HeadWord),因此本文引入了词性标注特征。在对事件句和事件短句做卷积操作时,可知每个单词的权重。此外,为了弥补卷积无法获知单词之间的先后顺序的缺点,本文引入了位置信息。

本文将事件句、词性信息和位置信息三者结合起来,通过卷积深层次挖掘出事件句 EM_1 和 EM_2 、事件短句 Con_1 和 Con_2 的局部特征信息 K_1 和 K_2 、 J_1 和 J_2 ,如式(4)所示:

$$K_i = Conv(EM_i) \quad (i=1,2) \quad (4)$$

$$J_i = Conv(Con_i) \quad (i=1,2)$$

在对事件句和事件短句进行卷积操作时,本文同时对它们进行了双向 LSTM 操作,以获取事件句的上下文信息。合并前向 LSTM 和后向 LSTM 的输出,从而提取出事件句的全局信息 H_1 和 H_2 ,以及事件短句的全局信息 M_1 和 M_2 ,如式(5)所示:

$$H_i = [\vec{H}_i, \overleftarrow{H}_i] \quad (i=1,2) \quad (5)$$

$$M_i = [\vec{M}_i, \overleftarrow{M}_i] \quad (i=1,2)$$

最后,将局部信息和全局信息相加,以捕获事件句的全局信息和局部信息 EB_1 和 EB_2 ,以及事件短句的全局信息和局部信息 CB_1 和 CB_2 ,如式(6)所示:

$$EB_i = K_i \oplus H_i \quad (i=1,2) \quad (6)$$

$$CB_i = J_i \oplus M_i \quad (i=1,2)$$

此外,注意力机制^[15]在自然语言处理领域是一种非常有价值的方法,它给句子中的每个单词都赋予了权重,通过权重的大小可以知道各单词在句子中的重要程度,将重要的特征信息提取出来,从而简化计算。

通过将事件句向量 EM_1 和 EM_2 ,以及权重 W_{E1} 和 W_{E2} 做点积操作,并将事件短句向量 Con_1 和 Con_2 ,以及权重 W_{C1} 和 W_{C2} 做点积操作,可以使得事件句和事件短句中每个单词的词向量间的差异性更大,从而获得事件句和事件短句的重要信息 ET_1 , ET_2 , CT_1 和 CT_2 ,如式(7)所示:

$$ET_i = EM_i \odot W_{E_i} \quad (i=1,2) \quad (7)$$

$$CT_i = Con_i \odot W_{C_i} \quad (i=1,2)$$

该模型在上述基础上还计算了两个事件句之间的 Cosine 距离作为相似度 ES_1 , ES_2 , CS_1 和 CS_2 ,如式(8)所示:

$$ES_1 = Sim(EM_1, ET_2) = EM_1^T \cdot ET_2$$

$$ES_2 = Sim(EM_2, ET_1) = EM_2^T \cdot ET_1$$

$$CS_1 = Sim(Con_1, CT_2) = Con_1^T \cdot CT_2$$

$$CS_2 = Sim(Con_2, CT_1) = Con_2^T \cdot CT_1 \quad (8)$$

3.3 输出层

本文将融合向量 EM_1 和 EM_2 、全局和局部信息向量 EB_1 和 EB_2 、重要信息向量 ET_1 和 ET_2 ,以及相似度 ES_1 和 ES_2 拼接在一起,形成向量 E_1 和 E_2 。将事件短句向量 Con_1 和 Con_2 、全局和局部信息向量 CB_1 和 CB_2 、重要信息向量 CT_1 和 CT_2 ,以及相似度 CS_1 和 CS_2 拼接在一起,形成向量 C_1 和 C_2 。相关计算公式如式(9)所示:

$$E_i = Concat(EM_i, EB_i, ET_i, ES_i) \quad (i=1,2) \quad (9)$$

$$C_i = Concat(Con_i, CB_i, CT_i, CS_i) \quad (i=1,2)$$

接着,对向量 E_1 , E_2 , C_1 和 C_2 做卷积和全局最大池化操

作,从中提取出每个卷积核中最有效的特征向量 $\mathbf{V}_{c1}, \mathbf{V}_{c2}, \mathbf{V}_{c1}$ 和 \mathbf{V}_{c2} 来表示事件句,从而简化计算。然后,将特征向量 $\mathbf{V}_{c1}, \mathbf{V}_{c2}, \mathbf{V}_{c1}$ 和 \mathbf{V}_{c2} 与事件对特征向量 \mathbf{P} 拼接在一起,形成向量 \mathbf{V}_f , 如式(10)所示:

$$\mathbf{V}_f = \text{Concat}(\mathbf{V}_{c1}, \mathbf{V}_{c2}, \mathbf{V}_{c1}, \mathbf{V}_{c2}, \mathbf{P}) \quad (10)$$

将 \mathbf{V}_f 向量放入一个多层的全连接分类器中。分类器使用了 relu 激活函数,如式(11)所示:

$$\mathbf{V}_h = \alpha(\mathbf{W}_h * \mathbf{V}_f + \mathbf{b}) \quad (11)$$

通过 sigmoid 层得出事件同指的置信度,如式(12)所示:

$$\text{score} = \text{sigmoid}(\mathbf{W}_0 * \mathbf{V}_h + \mathbf{b}_0) \quad (12)$$

该模型在全连接层中使用了 dropout 来防止过拟合,这也提升了网络模型的鲁棒性,使得模型易于收敛。

最终,用一个闭包将被分类器判定为同指的事件句组成一个事件链。

3.4 全局推理

本文对神经网络模型输出的置信度制定如下规则:若置信度大于 0.5,则判定该事件对是同指关系;否则判定该事件对不是同指关系。但是,该模型没有考虑到事件对之间关系的冲突,例如,有 2 个同指事件对 (T_1, T_2) 和 (T_2, T_3) , T_1 和 T_2 同指, T_2 和 T_3 同指,那么 T_1 和 T_3 也应该同指。但是,该模型可能会将 T_1 和 T_3 判定为非同指。为了防止这种情况的发生,DANGL 模型设计了一个全局推理模块进行后处理,具体如式(13)所示,其分别表示以下推理规则:

- 1) 如果 T_1 和 T_2 同指, T_2 和 T_3 同指,则 T_1 和 T_3 同指;
- 2) 如果 T_1 和 T_2 非同指, T_2 和 T_3 非同指,则 T_1 和 T_3 非同指;
- 3) 如果 T_1 和 T_2 同指, T_2 和 T_3 非同指,则 T_1 和 T_3 非同指;
- 4) 如果 T_1 和 T_2 非同指, T_2 和 T_3 同指,则 T_1 和 T_3 非同指。

$$\begin{cases} Co(T_1, T_2) \oplus Co(T_2, T_3) \rightarrow Co(T_1, T_3) \\ Un(T_1, T_2) \oplus Un(T_2, T_3) \rightarrow Un(T_1, T_3) \\ Co(T_1, T_2) \oplus Un(T_2, T_3) \rightarrow Un(T_1, T_3) \\ Un(T_1, T_2) \oplus Co(T_2, T_3) \rightarrow Un(T_1, T_3) \end{cases} \quad (13)$$

最后,用一个闭包将同指的事件句放入事件链中。

4 实验

4.1 语料和实验设置

TAC-KBP 2015 英文语料库是文档内事件同指消解常用的语料库,该语料库共有 360 篇文档,标注了 8 种事件类型 and 38 种事件子类型。语料库的详细情况如表 1 所列。

表 1 TAC-KBP 2015 语料库的统计数据

Table 1 Statistics of TAC-KBP 2015 corpus

Corpus	# Docs	# Sens	# Mentions	# Chains
TAC-KBP 2015	360	15284	12976	7415

本文使用 TAC-KBP 事件探测任务官网提供的训练集、验证集和测试集划分标准,并且文档内事件两两配对。在评估模型的性能时,与文献[16]一样使用了 4 种常用的评测方法: $MUC^{[17]}$, $B^3^{[18]}$, $BLANC^{[19]}$ 和 $CEAF_e^{[20]}$; 并且也使用上述 4 种方法的均值来评估模型。

在超参数的设置方面,词向量矩阵使用了 $\text{glove}^{[21]}$ 预训

练向量,词向量的维度设为 50,位置向量的维度设为 100。为了防止过拟合,将 Dropout 的值设为 0.2。在编码层,本文将 CNN 卷积核的数量设为 300,将卷积核窗口的大小设为 2,将 Bi-LSTM 神经元的维度设置为 150。该模型的训练轮数为 25,选择验证集中 val_loss 最低的模型来对测试集进行预测。

4.2 实验结果

为了验证 DANGL 模型在事件同指消解上的性能,将它与 3 个基准系统做比较: Krause 的系统, KBP 评测性能最优系统 KBP-TOP 系统,目前性能最好的系统 Liu。表 2 对比了这 4 个系统在 TAC-KBP 语料库上的事件同指消解的性能。

表 2 模型性能的比较

Table 2 Comparison of model performance

System	MUC	B ³	BLANC	CEAF _e	AVG
Krause	60.84	84.16	75.07	76.56	74.16
KBP-TOP	68.08	82.29	76.91	74.12	75.35
Liu	67.81	85.59	77.37	79.65	77.61
DANGL	70.23	85.22	78.81	78.57	78.21

从表 2 中的结果可以看出: 1) 本文的 DANGL 与 Krause 相比,在 4 个性能指标上得到了明显提升,尤其是在 MUC 指标上更是提升了 9.39%。这是因为 Krause 只进行了卷积操作,仅仅提取了事件句中的局部信息。这说明了全局信息的重要性。2) 与 KBP-TOP 系统相比, DANGL 在 4 个评估标准上分别提升了 2.15%, 2.93%, 1.9% 和 2.86%, 平均提升了 2.86%。KBP-TOP 系统引入了许多特征,而本文的 DANGL 只引用了少量特征,而且并没有使用外部资源来扩充特征信息,减少了外部资源的干扰。由此可见, DANGL 模型使用更少的特征取得了更好的判定效果。3) 相比于 Liu 使用的基于图的传统方法,本文模型在 MUC 评测指标上提升了将近 3%, 在 B³, BLANC 和 CEAF_e 评测指标上的值也都接近,在平均值上比 Liu 提高了 0.6%。Liu 的模型中除了使用 KBP 语料库提供的特征外,还使用了许多额外挖掘出来的特征信息,而本文的 DANGL 模型只使用了语料库中提供的少许特征,比 Liu 略优。

4.3 实验结果分析

为了分析 DANGL 模型中各个模块的作用,本文将 5 种简化模型与 DANGL 进行了对比,对比结果如表 3 所列。1) NoCNN: DANGL 模型未使用 CNN; 2) NoBiL: DANGL 模型未使用 Bi-LSTM; 3) NoAtt: DANGL 模型未使用注意力机制; 4) NoP+L: DANGL 模型没有引入词性信息特征和位置信息; 5) NoGI: DANGL 模型没有使用全局推理。

表 3 对比实验

Table 3 Comparative tests

System	MUC	B ³	BLANC	CEAF _e	AVG
DANGL	70.23	85.22	78.81	78.57	78.21
NoCNN	-2.22	-0.66	-1.13	-1.01	-1.26
NoBiL	-2.98	-0.21	-1.80	-0.20	-1.29
NoAtt	-1.98	-1.22	-1.14	-2.4	-1.69
NoP+L	-2.22	-0.89	-1.07	-1.44	-1.41
NoGI	-0.59	-0.48	-0.97	-0.78	-0.71

NoCNN 由于删除了 CNN,未从事件句和事件短句中提取出局部信息,4 种评测性能均有下降。NoBiL 删除了 Bi-LSTM,没有提取出全局信息,这同样导致了性能下降,尤其

是 MUC 下降了近 3%。由此可见,单纯地提取全局信息或者局部信息还不够,需要将事件句的全局信息和局部信息结合起来考虑。

NoAtt 没有使用注意力机制,仅仅对事件句和事件短句做了 CNN 和 Bi-LSTM 操作,4 种评测性能均降低了 1.5% 左右。这反映出注意力机制的重要性,因为它可以提取出事件句中的重要信息。事件句中往往存在对同指识别有价值的重要信息,而注意力机制可对事件句内的每个单词赋予权重,然后将每个单词的词向量和其权重进行点积操作。这样,原本相对重要的信息在句中更加突出,便于模型提取。例如,事件句 S_3 中的重点是“Bush leave France”, S_4 中的重点是“Bush travel Egypt”。在注意力机制下,这两个重点会更明显。如果没有提取出“France”和“Egypt”这 2 个重要信息,那么模型可能会误将这两个事件句判定为同指,因为它们的单词和结构很相似。

S_3 : President Bush will leave France today.

S_4 : Mr. Bush will travel next to Egypt today.

NoP+L 输入时并未在事件句里添加单词的词性信息和位置信息,导致平均性能降低了 1.41%。由此可见,引入词性信息和位置信息对事件同指消解任务起到了促进作用。NoGI 没有使用全局推理模型,它的性能相比上述 4 个模型均有提高,相对于 DANGL 模型,其低了 1% 左右。全局推理模型的使用纠正了一些互相冲突的事件对,从而减少了事件链中的错误。

结束语 本文设计了基于全局和局部信息的注意力机制神经网络模型来解决文档内的事件同指消解任务。该模型引入单词位置信息和词性标注信息来扩充事件句的特征信息,再通过 CNN、Bi-LSTM 和注意力机制来获取事件句中的局部信息、全局信息和重要信息,并将它们融合起来全面考虑。此外,结合事件对的特征来加强模型对事件同指关系的判定。最后,还利用全局推理模型来纠正模型的判定结果。实验结果表明,本文系统取得了最优的性能。未来将重点研究跨文档事件同指消解和端到端事件同指消解。

参 考 文 献

- [1] MOHAMMED M, MARWA A. Efficient Topic Detection System for Online Arabic News[OL]. http://xueshu.baidu.com/usercenter/paper/show?paperid=ea8bdd0d677e034c-c43e7099b2e14b91&site=xueshu_se.
- [2] 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.
- [3] SWAYAMDIPTA S, PARIKH A P, KWIATKOWSKI T. Multi-Mention Learning for Reading Comprehension with Neural Cascades[C]// *Proceedings of the International Conference of Learning Representation*. Vancouver: ICLR Press, 2018.
- [4] CHEN Z, JI H. Graph-based event coreference resolution[C]// *Proceedings of the Association for Computational Linguistics*. Singapore: ACL Press, 2009: 54-57.
- [5] HAGHIGHI A, DAN K. Simple Coreference Resolution with Rich Syntactic and Semantic Features[C]// *Proceedings of the Empirical Methods in Natural Language Processing*. Singapore: EMNLP Press, 2009: 1152-1161.
- [6] BEJAN C A, HARABAGIU S. Unsupervised event coreference resolution with rich linguistic features[C]// *Proceedings of the Association for Computational Linguistics*. Sweden: ACL Press, 2010: 1412-1422.
- [7] YANG B, CARDIE C, FRAZIER P. A Hierarchical Distance-dependent Bayesian Model for Event Coreference Resolution[OL]. <https://arxiv.org/abs/1504.05929>.
- [8] CHEN Z, JI H. Graph-based event coreference resolution[C]// *Proceedings of the Association for Computational Linguistics*. Singapore: ACL Press, 2009: 54-57.
- [9] CHEN Z, JI H, HARALICK R. A pairwise event coreference model, feature impact and evaluation for event coreference resolution[C]// *Proceedings of the Association for Computational Linguistics*. Singapore: ACL Press, 2009: 17-22.
- [10] LIU Z, ARAKI J, HOVY E, et al. Supervised within-document event coreference using information propagation[OL]. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.468.8992>.
- [11] LU J, NG V. Joint Learning for Event Coreference Resolution[C]// *Proceedings of the Association for Computational Linguistics*. Vancouver: ACL Press, 2017: 90-101.
- [12] LIU Z, MITAMURA T, HOVY E. Graph-Based Decoding for Event Sequencing and Coreference Resolution[C]// *Proceedings of the Computational Natural Language Learning*. Belgium: CoNLL Press, 2018.
- [13] KRAUSE S, XU F, USZKOREIT H, et al. Event Linking with Sentential Features from Convolutional Neural Networks[C]// *Proceedings of the Computational Natural Language Learning*. Berlin: CoNLL Press, 2016: 239-249.
- [14] CHOUBEY P K, HUANG R. Event Coreference Resolution by Iteratively Unfolding Inter-dependencies among Events[C]// *Proceedings of the Empirical Methods in Natural Language Processing*. Copenhagen: EMNLP Press, 2017: 2124-2133.
- [15] PARIKH A P, TÄCKSTRÖM O, DAS D, et al. A Decomposable Attention Model for Natural Language Inference[C]// *Proceedings of the Empirical Methods in Natural Language Processing*. Texas: EMNLP Press, 2016: 2249-2255.
- [16] MITAMURA T, LIU Z, HOVY E. Overview of TAC-KBP 2015 Event Nugget Track[C]// *Proceedings of the Tunnelling Association of Canada*. Canada: TAC Press, 2015.
- [17] BAGGA A, BALDWIN B. Algorithms for Scoring Coreference Chains[C]// *LREC Workshop on Linguistic Coreference*. Spain: LREC Press, 1998: 563-566.
- [18] LUO X. On coreference resolution performance metrics[C]// *Proceedings of the Empirical Methods in Natural Language Processing*. Vancouver: EMNLP Press, 2005: 25-32.
- [19] VILAIN M, BURGER J, ABERDEEN J, et al. A Model-Theoretic Coreference Scoring Scheme[C]// *Proceedings of the Association for Computational Linguistics*. Cambridge: ACL Press, 1995: 45-52.
- [20] LUO X. On coreference resolution performance metrics[C]// *Proceedings of the Empirical Methods in Natural Language Processing*. Vancouver: EMNLP Press, 2005: 25-32.