

# 基于多注意力机制的事件同指消解方法

方 杰 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)

(江苏省计算机信息处理技术重点实验室 江苏 苏州 215006)

**摘 要** 事件同指消解是信息抽取的一项重要任务,在信息融合、问答系统、阅读理解中都有着重要的作用。文中提出了一种基于多种注意力机制的卷积神经网络的 CorefNet 方法,用于消解文档级事件同指。该方法通过深层卷积网络抽取事件特征,并使用多种注意力机制获取重要信息。相比过去大部分建立在概率模型和图模型上的传统方法,所提方法仅使用了少量特征;与目前主流的神经网络模型相比,文中方法可以提取深层的事件特征,明显提高了事件同指消解的准确率。在 ACE2005 数据集上的实验验证了 CorefNet 优于目前最优的基准系统。

**关键词** 事件同指,文档级,注意力机制,深层卷积网络

**中图法分类号** TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.08.046

## Employing Multi-attention Mechanism to Resolve Event Coreference

FANG Jie LI Pei-feng ZHU Qiao-ming

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

(Province Key Laboratory of Computer Information Processing Technology of Jiangsu, Suzhou, Jiangsu 215006, China)

**Abstract** Event coreference resolution is an asignificant subtask of information extraction and plays an import role in information fusion, QA system and reading comprehension. This paper introduced a multi-attention-based CNN neural network, called CorefNet, to resolve document-level event coreference. CorefNet uses a deep CNN to extract event features and a multi-attention mechanism to capture important features. Compared with most previous studies with probability-based or graph-based models, the proposed model only uses a few features. Compared with the current main stream nueral network model, this menthod can extract deep event features, and significantly improve the performance of event coreference resolution. The experimental results on the ACE2005 corpus show that this model achieves the state-of-the-art results.

**Keywords** Event coreference, Document-level, Attention mechanism, Deep CNN

## 1 引言

事件同指消解(Event Coreference Resolution)在许多自然语言处理领域(如话题探测<sup>[1]</sup>、信息抽取<sup>[2]</sup>和问答<sup>[3]</sup>等)中都起到了重要的作用。事件同指消解用于判断文档中的多个事件实例(Event Mention)在文本中是否指向真实世界中的同一个事件,并将相互同指的事件进行聚类,形成一个事件链。以下给出两个同指的事件实例。

$S_1$ : A Cuban patrol boat with four heavily armed men landed on American shores.

$S_2$ : These bozos let four armed Cubans land on our shores.

事件  $S_1$  中的触发词“landed”和事件  $S_2$  中的触发词“land”都指向现实生活中的一个移动(Movement)类型的事

件,它们被聚类到同一个事件同指链中。

与已经比较成熟的实体同指消解相比,事件同指消解的研究相对较少。事件同指消解在事件实例的表示形式上非常灵活<sup>[4]</sup>,是一项具有挑战性的任务。根据事件实例是否在同一个文档中,可以把事件同指消解划分为两大任务,即文档内事件同指消解以及跨文档事件同指消解。本文主要研究文档内事件同指消解。

以往的研究通常使用基于概率的模型或者基于图的模型<sup>[5]</sup>,这通常不能很好地抓住触发词间、论元间以及事件结构间的有效语义信息。这是因为在同一篇文档中的同指事件实例在事件触发词或事件论元上都有不同的表达方式,从而导致手工特征在同指事件消解中不可靠。此外,大多数研究都专注于事件对之间的事件同指消解。例如,把事件同指消解任务视为一个判断两个事件实例是否同指的分任务。尽管

到稿日期:2018-07-29 返修日期:2018-10-29 本文受国家自然科学基金(61472265,61772354,61773276)资助。

方 杰(1993-),男,硕士生,主要研究领域为中文信息处理;李培峰(1971-),男,教授,博士生导师,主要研究领域为自然语言处理和机器学习,E-mail:pflj@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。

有各种启发式的规则用于解决这类问题,但其依然会导致不同的事件实例对存在冲突。

为了解决上述问题,本文设计了基于神经网络的 CorefNet 模型用于解决事件同指消解。该模型通过设计深层的神经网络模型作为语言模型来获取事件的语义信息,并使用注意力机制进一步提升模型的性能。实验结果显示,本文的模型在 ACE2005 数据集上获得了不错的效果。本文的贡献有如下几个方面:

- 1) 引入了深层的卷积结构,使得模型可以提取深层的语义特征;
- 2) 使用多种注意力机制,可以筛选出重要的特征信息;
- 3) 本文在 CorefNet 中加入了语义角色特征,触发词附近的论元特征对同指分类有着重要的作用。

## 2 相关工作

相比在实体同指消解<sup>[6]</sup>上的大量研究,事件同指消解则是一个少有人研究的领域。早期在事件同指消解上的方法大部分建立在实体同指消解上<sup>[7]</sup>。最近的方法主要集中于探索事件的句法结构。Chen 等<sup>[8]</sup>提出了一种基于 4 种事件属性(如时态、极性等)的事件对分类器。Liu 等<sup>[9]</sup>提出了图模型聚类的方法,将图切割的规则用于事件同指的聚类。Liu 等<sup>[9]</sup>采用了基于多种特征的 SVM 事件对分类器,利用 100 多种特征来发现事件之间的信息。Lu 等<sup>[10]</sup>提出了一种基于马尔科夫链的联合推理模型来纠正事件对分类器产生的错误。

如今,深度学习的方法在自然语言处理领域被大量使用。神经网络在文本分类<sup>[11]</sup>、相似度<sup>[12]</sup>比较等任务中都有着不错的效果。Krause 等<sup>[13]</sup>首次使用神经网络的方法进行文档内事件同指的消解。它分为两个部分:首先使用卷积神经网络对句子和触发词上下文进行建模,提取特征;然后加入事件对特征,并判断事件是否同指。卷积神经网络(CNN)首先在图像领域取得广泛关注。卷积神经网络的平移不变性、局部连接和参数共享等特点使得在提取特征时有着出色的表现,这在自然语言处理领域中同样适用。通过卷积的操作模型可以发现文本局部中单词与单词之间的联系。多种卷积核也使模型能从多个角度提取文本的特征。这也符合人类的阅读习惯,从多次的阅读中更好地理解文本信息。早期的卷积神经网络由于容易发生梯度消失的情况,通常是浅层的网络,这使模型不能提取深层特征。随后的残差机制解决了这个问题。将不同层的特征相加,使顶层特征传递至最后一层,而不丢失信息。深层的卷积神经网络能够提取深层的特征,大大提升了网络的泛化能力,并改善了模型的效果。

注意力机制(Attention)的思想很早就存在,我们在观察一幅图像时,注意力会自然地集中于图像的某些地方。受到生物学现象的启发,注意力机制在 2014 年首次正式应用在计算机视觉领域。加入注意力机制后,模型变得不易被干扰信息所影响,性能得到了提升。而后,在 2015—2016 年,注意力机制及其在循环神经网络、卷积神经网络中的变种也逐渐出

现在自然语言处理的各项任务中。2017 年,Google 提出的 Transformer 模型使用的多头注意力机制和自注意力机制取得了出色的效果,因此引起了研究人员的重点关注。如今,在自然语言处理任务中,软对齐注意力机制和多头注意力机制被广泛应用。

## 3 CorefNet 模型

CorefNet 是一种带多注意力机制的卷积神经网络模型,主要解决了事件特征难以获取的问题。其通过使用深层的卷积神经网络建立语言模型,自动地抽取事件特征,并使用注意力机制对特征加权,筛选重要的特征。CorefNet 模型如图 1 所示,该模型抽取事件之间的深层特征,并融合两个事件的特征,从而判断两个事件是否同指。

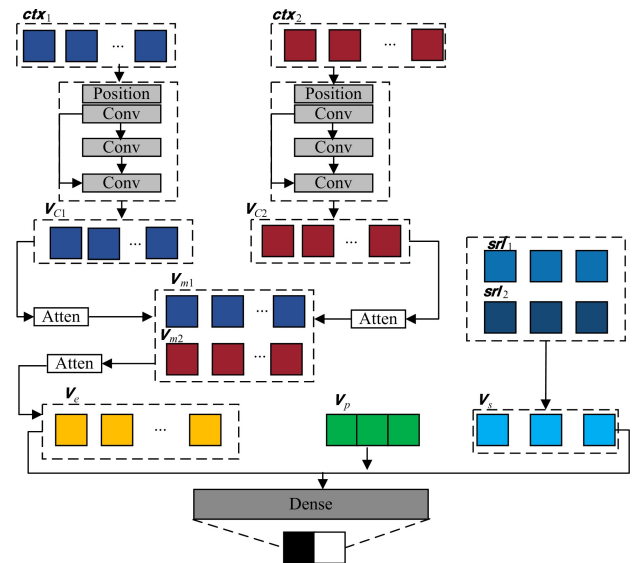


图 1 CorefNet 模型架构

Fig. 1 Architecture of CorefNet model

CorefNet 主要分为以下几个部分:

1) 分别对两个事件进行编码,编码层主要使用卷积来提取事件特征。这是由于卷积可以根据上下文信息做出推断,从而挖掘出上下文之间的语义特征。本模型使用了多层不同采样的卷积,可以进一步挖掘事件之间的深层特征;加入的残差机制也进一步防止了梯度消失。

2) 使用多种注意力机制<sup>[14]</sup>(点积注意力和自注意力)融合两个事件。点积注意力可以发现两个事件具有的共性特征,自注意力机制可以发现句子本身对分类任务有重要划度的特征。

3) 加入了事件对特征以及语义角色特征。事件对特征来自于事件的属性,这对于判断事件是否同指具有重要的意义。语义角色特征抽取了触发词周围的论元特征,进一步提升了模型的性能。

4) 利用分类器消解事件同指。将所有特征进行融合,使用全连接网络进行分类,给出事件同指的置信度。

### 3.1 模型输入

事件同指消解可以看作是一个分类的任务,即给定两个

事件,判定它们是否为同指事件。在一个事件中,触发词是事件的核心,因此,本文以触发词为中心,选取句子中触发词前 10 个单词和后 10 个单词作为描述一个事件的上下文短语特征  $feat_{c1}, feat_{c2}$ ,并记录下每个单词和触发词的距离作为它的位置信息  $feat_{l1}, feat_{l2}$ 。

本文发现语义角色标注在描述事件中也有着重要的作用,它可以根据谓词标记出句子中的施事者和受事者等,因此,本文使用语义角色标注标记出触发词的施事者、受事者、地点、时间。由于部分触发词是以名词的形式组成,因此本文取出在施事者或受事者中含有触发词的语义角色序列特征  $feat_{s1}$  和  $feat_{s2}$ 。语义角色标注描绘了句子中不同对象之间的联系,并大大简化了句子中复杂的表达,使得模型能更好地挖掘句子文本中的事件特征,从而提升模型的性能。

此外,本文还利用了事件对之间相互匹配的信息,如事件类型/子类型是否一致的特征  $feat_{type}$ ,时态是否一致的特征  $feat_{realis}$ ,触发词、触发词原型、词性是否一致的特征  $feat_{trigger}$  等,来进一步提升模型的性能。

### 3.2 词嵌入层

事件上下文特征和语义角色序列特征是一种词袋特征。本文采用词嵌入方法,对于句子中的每个单词,基于词嵌入矩阵  $M$  得到词向量  $w_i$ ,并将词向量合并得到句子向量  $sen$ ,如式(1)所示:

$$sen = Concat(w_1, w_2, w_3, \dots, w_n) \quad (1)$$

本文使用了 50 维预训练的 glove 词向量矩阵作为本文的词嵌入矩阵  $M$ ,对事件对中每个事件的上下文特征  $feat_{c1}, feat_{c2}$  和语义角色特征  $feat_{s1}, feat_{s2}$  分别进行编码,得到上下文向量  $ctx1$  和  $ctx2$ ,及语义角色向量  $srl1$  和  $srl2$ 。此外,句子中单词的顺序也是一个重要的特征,单词在句子中出现的位置不同,所表达的意思也会存在差异。因此,本文加入了单词在句子中的位置信息。位置信息使用了随机的词嵌入矩阵对位置特征进行编码,得到位置向量  $loc1$  和  $loc2$ ,位置向量的维度为 10。

### 3.3 卷积层

卷积是对句子中相邻的单词进行采样。由于相邻单词之间都有较强的关联度,可以抽取到更多的特征,因此卷积在自然语言处理中获得了广泛的运用。本文使用卷积层来提取触发词上下文之间的特征。为了弥补卷积无法很好地获取句子前后的序列信息的缺陷,本文采用了将词向量和位置向量结合的方法。

本文使用三层卷积层,其中第一层卷积核的尺寸为 3,第二层卷积核的尺寸为 4,第三层卷积核的尺寸为 5。设置不同的卷积核尺寸可从多个角度采样,从而获得更好的泛化能力。另外,将卷积核的数量设置为 60。为了防止梯度消失,本文采用了残差网络的思想,将不同层的特征相加以更好地提取特征,这避免了前几层的特征在向后传播的过程中丢失的情况。实验中将词嵌入后获得的句子向量与第三层相加,如式(2)所示,得到向量  $V_{C1}$  和  $V_{C2}$ 。多层相加的机制揉合了不同层的特征,使得浅层中有意义的特征也能被保存下来。

$$V_{C1} = Conv(ctx1) + ctx1 \quad (2)$$

$$V_{C2} = Conv(ctx2) + ctx2$$

其中,  $Conv()$  为卷积运算。

### 3.4 多重注意力机制

注意力机制在机器翻译领域被广泛采用,可以对句子向量中的单词赋予权重,自动地发现对模型分类具有重要作用的特征。本文在 CorefNet 中分别使用了点积注意力和自注意力两种注意力机制。首先,对于  $V_{C1}$  和  $V_{C2}$ ,本文使用点积注意力进行编码,如式(3)所示,分别得到向量  $V_{m1}$  和  $V_{m2}$ 。

$$V_{m1} = Attention(V_{C1}W^{C1}, V_{C2}W^{C2}, V_{m1}W^{m1})$$

$$V_{m2} = Attention(V_{C2}W^{C2}, V_{C1}W^{C1}, V_{m2}W^{m2}) \quad (3)$$

其中,  $W^{C1}, W^{C2}, W^{m1}, W^{m2}$  分别是  $V_{C1}, V_{C2}, V_{m1}, V_{m2}$  的权重。

点积注意力机制可以对两个向量之间有共性的值赋予较高权值。由于同指事件之间具有共性的单词一般会比较,因此点积注意力机制对识别语义相近的同指事件很有帮助。

接着,将  $V_{m1}$  和  $V_{m2}$  合并,如式(4)所示,得到合并的向量  $V_m$ 。

$$V_m = Concat(V_{m1}, V_{m2}) \quad (4)$$

最后,本文使用自注意力机制对  $V_m$  进行编码,得到描述事件的特征向量  $V_e$ ,如式(5)所示。与点积注意力相比,自注意力机制使用的 Attention 函数和点积注意力机制相同,不同之处在于 Attention 函数中的参数,在自注意力机制中,采用了将向量和自身做乘积的方法。自注意力机制可以发现句子内部中重要的单词特征,并在反向传播的过程中自动调整,不断改变权值。通过观察数据集,本文发现在一些同指事件中,部分单词出现的频率很高,对这些单词赋予较高的权值能提升事件同指的性能,自注意力机制就是用于发现这些单词并自动赋予权值的一种机制。

$$V_e = Attention(V_mW^m, V_mW^m, V_eW^e) \quad (5)$$

其中,  $W^m, W^e$  分别是  $V_m, V_e$  的权重。

### 3.5 语义角色特征和事件对特征的引入

将语义角色向量  $srl1$  和  $srl2$  进行合并,得到  $V_s$ ,如式(6)所示:

$$V_s = Concat(srl1, srl2) \quad (6)$$

事件对特征是两个事件的属性相互匹配得到的特征。在 CorefNet 中,它的输入是二元数值或者小数。如图 1 所示,本文先将事件对特征  $feat_{type}, feat_{realis}, feat_{trigger}$  转换成向量  $V_n$ ,然后将它们输入到一个前馈网络中得到  $V_p$ ,用来抽取事件对隐藏的特征。

### 3.6 输出层

本文将事件特征向量  $V_e$ 、语义角色向量  $V_s$  和事件特征向量  $V_p$  合并成最终的向量  $V_f$ ,如式(7)所示,用来表示两个事件最终的语义关系。

$$V_f = Concat(V_e, V_s, V_p) \quad (7)$$

$V_f$  向量会被输入到一个全连接分类器中,分类器使用了 relu 的激活函数,如式(8)所示:

$$V_h = \alpha(W_h * V_f + b) \quad (8)$$

其中,  $\alpha$  是激活函数,  $W_h$  和  $b$  是参数。

如式(9)所示,本文通过一层 sigmoid 层得出同指的置信度。

$$\text{Score} = \text{sigmoid}(\mathbf{W}_{\text{out}} * \mathbf{V}_h + \mathbf{b}_{\text{out}}) \quad (9)$$

其中,  $\mathbf{W}_{\text{out}}$  是模型的参数,  $\mathbf{b}_{\text{out}}$  是偏置项。

模型的损失函数如式(10)所示:

$$J(\theta) = -\log \prod_{i=1}^N p(y^{(i)} / x^{(i)}, \theta^{(i)}) \frac{\lambda}{2} \|\theta\|^2 \quad (10)$$

其中,  $\theta = \{\mathbf{W}_{C1}, \mathbf{W}_{C2}, \mathbf{W}_{S1}, \mathbf{W}_{S2}, \mathbf{W}_e, \mathbf{W}_p, \mathbf{W}_f, \mathbf{W}_h, \mathbf{W}_{\text{out}}, \mathbf{b}_{\text{out}}\}$ 。

最终,本文使用一个闭包将所有被分类器判为同指的事件对连接成一个事件链。

### 3.7 网络正则化

本文在词嵌入层和全连接层中使用了 Dropout, Batch-Norm 技术来防止过拟合,并提升了网络的鲁棒性,也使得模型更易于收敛。

## 4 实验与分析

### 4.1 数据集及实验设置

本文主要在 ACE2005 的英文语料库中,将所提方法与以往事件同指的研究进行比较来验证本文模型的有效性。表 1 介绍了 ACE2005 语料库的统计数据,语料库中包含了 6 种体裁的 599 篇文档,标注了 8 种事件类型和 33 种事件子类型。ACE 语料库是自动内容抽取会议(ACE)的产物。ACE2005 语料库在文档内事件同指消解的研究中被广泛使用,它包含了中文和英文文档。在 ACE 语料库中,事件被定义为“特定情况下发生的事情,通常伴随着状态的改变(包括参与者)”。相比 MUC 会议中提出的语料库,ACE 语料库涵盖了更多的事件类型(子类型)。

表 1 ACE2005 语料库的统计数据

Table 1 Statistics of ACE2005 corpus

corpus	# Docs	# Sens	# Mentions	# Chains
ACE	599	15 494	5 268	4 046

本文实验使用了与 Krause 一样的训练集、验证集和开发集。将 ACE2005 数据集以 8:1:1 的比例划分为训练集、开发集、测试集,文档内的事件都是两两配对的。在模型参数的设置中,词向量的维度为 50,词向量矩阵使用预训练的 glove<sup>[15]</sup> 向量,位置向量使用随机的词向量矩阵,位置向量的维度为 10,Dropout 值设为 0.5;使用 Adam 来优化本文的模型,模型训练轮数为 50。

本文使用 3 种评测方法即 MUC<sup>[16]</sup>, B<sup>3</sup><sup>[17]</sup>, Blanc<sup>[18]</sup> 来评估本文的模型,同时将上述 3 种评测方法计算得出的 F1 值的均值作为评测标准。

MUC 是链级的评价指标,它将同指事件看成一个无向图,相互同指的事件之间用无向的边连接,并统计正确划为同指边的分类情况。

Blanc 是事件级的评价指标,它主要统计了被正确划为同指和被正确划为非同指的分类情况。

B<sup>3</sup> 是实体级的评价指标,它通过统计正确划分的事件中的实体数量来评判分类情况。

### 4.2 实验结果

为了验证本文的 CorefNet 模型在文档内同指消解上的

性能,将它与两个基准系统做比较:1) Liu 等<sup>[10]</sup> 的使用了含 100 多种特征的传统机器学习分类器模型;2) Krause<sup>[13]</sup> 的神经网络模型。表 2 列出了 3 种模型在事件类型已知情况下的事件同指消解性能。

表 2 实验结果

Table 2 Experiment results

System	Blanc			B <sup>3</sup>			MUC			AVG
	P	R	F1	P	R	F1	P	R	F1	
Liu	70.01	70.88	70.43	88.86	89.90	<b>89.38</b>	48.75	53.42	50.98	70.26
Krause	71.80	75.16	73.31	86.12	90.52	88.26	45.16	61.54	52.09	71.22
CorefNet	<b>75.85</b>	<b>77.05</b>	<b>76.44</b>	<b>88.43</b>	<b>88.12</b>	<b>88.27</b>	<b>60.14</b>	<b>61.48</b>	<b>60.80</b>	<b>75.17</b>

(单位:%)

表 2 的结果显示本文的 CorefNet 模型在所有指标上都优于两个基准系统,分别在 F1 值上有 4.91% 和 3.95% 的提升。Liu 的模型使用了 100 多种特征,而本文的模型只使用了少量的特征。这些结果验证了本文的 CorefNet 神经网络模型的有效性。与 Krause 的实验相比,本文的模型在 Blanc 和 MUC 上的 F1 值分别提升了 3.13% 和 8.71%。这些结果都表明了本文的 CorefNet 模型是优于 Krause 的 CNN 模型。本文模型在 B<sup>3</sup> 上并没有太大的提升,这主要是由 B<sup>3</sup> 的机制引起的,如果所有事件都是非同指事件,则孤立事件是最多的, B<sup>3</sup> 将会达到 100%。

### 4.3 实验结果与分析

CorefNet 性能提升的原因主要有以下 3 个方面:1) 深层卷积结构的加入;2) 注意力机制的使用;3) 语义角色特征的加入。表 3 分别列出了加入残差机制、语义角色特征后各项指标在 F1 值上的提升。

表 3 与加入残差信息、注意力机制、语义角色特征的方法在 F1 值上的比较

Table 3 F1 comparison with methods adding res, attention, or SRL

(单位:%)

System	Blanc	B <sup>3</sup>	MUC	AVG
Base	74.24	88.08	56.34	72.88
+ Res	75.09	87.87	57.65	73.53
+ Attention	75.82	89.05	58.49	74.45
+ SRL	<b>76.44</b>	<b>88.27</b>	<b>60.80</b>	<b>75.17</b>

表 3 显示,加入残差机制后,模型在 Blanc 和 MUC 上的 F1 值都分别提升了 0.85% 和 1.31%。这说明深层的卷积神经网络可以提取更多深层的语义特征,使得模型的准确度得到提升,残差机制也在一定程度上起到了过拟合的作用。相比 RNN 的方法,卷积的训练速度有着明显的优势。

表 3 表明了注意力机制的重要性,它可以防止不相关的特征对模型带来干扰。在 Blanc, B<sup>3</sup> 和 MUC 3 个指标中,加入注意力机制的方法分别将 F1 值提升了 1.58%, 0.97%, 2.15%。本文的注意力机制被用于对输入的特征加入权值,它对重要的特征给出较大的权值,而对不重要的特征加入较小的权值,从而使得模型处理数据更加准确。下面给出一组例子。

S<sub>3</sub>: There were no reports of protests against the war from the Gulf States many of whom are host to US military.

$S_4$ : We are warning Israel not to exploit this war against Iraq to carry out more attacks against the Palestinian people in the Gaza strip and destroy the Palestinian authority and the peace process.

由于事件  $S_4$  中含有更多的单词,因此模型难以判断  $S_3$  和  $S_4$  是否同指。本文的注意力机制将会对单词“no/not”“against”“war”等加大权值,其揭示了两个事件内在的联系。

表 3 也表明了语义角色特征的有效性。在 Blanc,  $B^3$  和 MUC 3 个指标中,加入语义角色特征后,模型分别将 F1 值提升了 2.2%, 0.19%, 4.46%。本文的语义角色特征只选取了触发词周围的论元,这减少了其他不相关论元的干扰。相比实体特征,语义角色可以提取出代词的论元,以下是一组例子。

$S_5$ : They primarily talked about the war in Iraq.

语义角色可以抽取出施事者“they”、受事者“war”以及谓词“talked”,这是实体特征无法抽取的。这使得模型可以抽取更精准的特征,从而提升模型的准确度。

**结束语** 本文构建了基于神经网络模型的 CorefNet 方法来消解文档内同指的任务。模型中引入的深度卷积神经网络、注意力机制和语义角色特征等都大大提升了模型的性能。实验结果显示,本文的模型在 ACE 语料库中击败了当前最好的模型。未来可将本文的工作延伸至跨文档事件同指消解和多语言事件同指消解领域。

## 参 考 文 献

- [1] ALLAN J, CARBONELL J G, DODDINGTON G, et al. Topic Detection and Tracking Pilot Study: Final report [C] // Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop. Burlington: Morgan Kaufmann Press, 1998: 194-218.
- [2] LI P F, ZHU Q M, ZHOU G D. Argument Inference from Relevant Event Mentions in Chinese Argument Extraction [C] // Proceedings of ACL 2013. Minneapolis: ACL Press, 2013: 1477-1487.
- [3] NARAYANAN S, HARABAGIU S. Question Answering Based on Semantic Structures [C] // Proceedings of COLING 2004. Switzerland: ACL Press, 2004: 693-702.
- [4] CHOUBEY P K, HUANG R H. Event Coreference Resolution by Iteratively Unfolding Inter-dependencies among Events [C] // Proceedings of EMNLP 2017. Copenhagen: ACL Press, 2017: 2124-2133.
- [5] CHEN Z, JI H. Graph-based Event Coreference Resolution [C] // Proceedings of TextGraphs 4. Singapore: ACL Press, 2009: 54-57.
- [6] HAGHIGHI A, KLEIN D. Simple Coreference Resolution with Rich Syntactic and Semantic Features [C] // Proceedings of EMNLP 2009. Singapore: ACL Press, 2009: 1152-1161.
- [7] NG V, CARDIE C. Identifying Anaphoric and Non-Anaphoric Noun Phrases to Improve Coreference Resolution [C] // Proceedings of COLING 2002. Taipei: ACL Press, 2002: 1-7.
- [8] CHEN Z, JI H, HARALICK R. A Pairwise Event Coreference Model, Feature Impact and Evaluation for Event Coreference Resolution [C] // Proceedings of the Workshop on Events in Emerging Text Types. Borovets: ACL Press, 2009: 17-22.
- [9] LIU Z Z, ARAKI J, HOVY E, et al. Supervised Within-Document Event CoReference Using Information Propagation [C] // Proceedings of LREC 2014. Reykjavik: Springer, 2014: 4539-4544.
- [10] LU J, NG V. Joint Learning for Event Coreference Resolution [C] // Proceedings of ACL 2017. Vancouver: ACL Press, 2017: 90-101.
- [11] KIM Y. Convolutional Neural Networks for Sentence Classification [C] // Proceedings of EMNLP 2014. Doha: ACL Press, 2014: 1746-1751.
- [12] YIH W T, TOUTANOVA K, PLATT J C, et al. Learning Discriminative Projections for Text Similarity Measures [C] // Proceedings of CoNLL 2011. Oregon: ACL Press, 2011: 247-256.
- [13] KRAUSE S, XU F Y, USZKOREIT H, et al. Event Linking with Sentential Features from Convolutional Neural Networks [C] // Proceedings of CoNLL 2016. Berlin: ACL Press, 2016: 239-249.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need [C] // Proceedings of NLPs 2017. California: MIT Press, 2017: 5998-6008.
- [15] PENNINGTON J, SOCHER R, MANNING C, et al. Glove: Global Vectors for Word Representation [C] // Proceedings of EMNLP 2014. Doha: ACL Press, 2014: 1532-1543.
- [16] BAGGA A, BALDWIN B. Algorithms for Scoring Coreference Chains [C] // Proceedings of LREC 1998. Granada: ELRA Press, 1998: 563-566.
- [17] LUO X Q. On Coreference Resolution Performance Metrics [C] // Proceedings of EMNLP 2005. Vancouver: ACL Press, 2005: 25-32.
- [18] VILAIN M, BURGER J, ABERDEEN J, et al. A Model-Theoretic Coreference Scoring Scheme [C] // Proceedings of MUC-6. California: ACM Press, 1998: 45-52.