融合篇章表征的事件指代消解研究

吴瑞荣[1]　孔芳[1,†]

1. 苏州大学计算机科学与技术学院，苏州 215006；
† 通信作者，E-mail: kongfang@suda.edu.cn

摘要　事件指代消解作为自然语言处理领域多项任务的技术支撑，对篇章理解具有重要意义。该任务与实体指代消解相比难度大，其主要原因有事件描述在非结构化文本中分布稀疏；且不具备同指关系的单链占有很大比例；同时事件自身承载的语义信息比实体更加丰富。针对事件指代消解任务的以上特点，提出一种融合篇章表征的事件指代消解平台。该平台通过 CRF 有效的区分非事件句，单链以及同指链，同时利用分级 attention 机制捕捉句子级别和篇章级别的重要信息。在 KBP2015、2016 数据集上进行的事件指代消解实验验证了本模型的有效性，其中在 CoNLL 评测标准下达到了 43.07%的 F1 值。

关键词　事件指代消解; 篇章表征; 分级 attention

中图分类号　X123

Event Coreference Resolution with Document Representation

WU Ruiying[1]，KONG Fang[1,†]

1. School of Computer Sciences and Technology, Soochow University, Suzhou 215006; †Corresponding author, E-mail:kongfang@suda.edu.cn

Abstract　Event coreference resolution as a technical support for multiple tasks in the field of Natural Language Processing, it is of great significance to text comprehension. This task is more difficult than entity coreference resolution, the main reason is that the event mention in the unstructured texts is sparse, and most of them do not have the coreference relationship, at the same time, the semantic information carried by the event itself is richer than entity. For the above characteristics of event coreference resolution, an event coreference resolution platform with text representation is proposed. This platform effectively distinguishes non-event mention, single-chain and coreference event mention through CRF, and uses hierarchical attention mechanism to capture important information at sentence level and text level. Experiments on KBP2015 and 2016 datasets verify the validity of the model, and the CoNLL evaluation standard reaches 43.07% of the F1 value.

Key words　event coreference resolution; document representation; hierarchical attention

阅读文本、信息抽取的过程往往是对非结构化文本语义信息的理解和抽取，事件作为语义信息的主要载体，在自然语言处理的很多领域中都起着关键的作用。从概念上来说，事件对应与某种状态的变化，这种变化通常包括时间、地点、参与者，例如某地区发生了恐怖袭击事件，某个国家的总统宣布辞职事件，某家公司收购事件等。

Vendler[1]曾提出，在非结构化文本中，事件被认为是文本跨度（Text spans），通常是作为指示状态变化的动词或名词。在与事件相关任务的研究中这些指示状态变化的词被称为触发词(Trigger)，某个具体事件的表述被称为事件描述(Event mention)或者事件句，找到正确的触发词即可认为找到了相应的事件描述。例如在 KBP2015 标注语料中，事件描述“A pedestrain was killed in an accident Sunday morning on the Upper West Side of Manhattan.”描述了一个 life-die 类型的事件，其中“killed”作为触发词，触发了该事件的产生。

非结构化文本中的事件描述之间常存在着多种复杂的关系，例如时序关系、包含关系、同指关系等，了解这些事件描述之间的关系能够更好的帮助理解篇章内容，捕捉关键的语义信息，本文主要研究的是其中的同指关系。

基金项目：国家自然可想而基金面上项目（61876118）；国家自然科学基金重点项目（61836007）

事件指代消解(Event coreference resolution)是指在非结构化文本中找到正确的触发词，并判断这些触发词所触发的事件描述是否指向现实世界中的同一个事件，若是，则认为它们之间构成同指关系，并将具有同指关系的事件描述构成一个同指链，否则为单链。如图 1 中的示例所示，这篇文本包含了 10 个句子，其中，加粗的部分为触发词，下划线标注的部分为实体，具有同指关系的触发词之间用箭头连接，例如“suicide”和“death”具有同指关系，并构成一个同指链，而“attack”,“arrested”等触发词不和任何其他触发词构成同指关系，即为单链。

<doc id="1d2911e09a6746b942c3e7b3cbdcb0ce">
1.Former Pakistani dancing girl commits **suicide** 12 years after horrific acid **attack** which **left** her looking 'not human'.
2.She had undergone 39 separate surgeries to repair **damage** .
3.Leapt to her **death** from sixth floor Rome building earlier this month.
4.Her ex-husband was charged with attempted murder in 2002 but has since been acquitted .
5.Bilal Khar was **arrested** in 2002 and charged with attempted murder following the attack, only to be released on bail after five months.
6.Khar, an **ex-parliamentarian** and son of a wealthy Pakistani governor,....
7.They let him off, huh?
8.Expect to see more of this, then.
9.Poor woman a terrible price to pay for accepting a misongynist cultural norm.
10.I do not choose to get past my ethnocentric disgust on this one.
</doc>

图 1　事件指代消解示例图

Fig.1　Example diagram of Event Coreference Resolution

目前，事件指代消解任务的研究方法主要是通过构建大量的人工特征，或完全借鉴实体指代消解任务的研究思路。然而大量的特征不仅需要手动设计，费时费力，而且很难全面的捕捉篇章级别的语义信息，也会因为一些不正确的规则约束导致错误传播影响模型的最终性能。尽管事件指代消解与实体指代消解在是形式上很相似，但是相比后者，事件指代消解有以下几个难点，完全套用实体指代消解的方法并不能有效的处理该任务：（1）篇章中触发词的数量很少，分布稀疏，如示例中的篇章一共只有 7 个触发词，这种比例要远远少于实体在篇章中的分布情况。（2）篇章中的事件多为单链，被反复提及的事件很大程度上与篇章的主题相关。示例中单链的数量为 3，而同指链的数量为 2，并且构成指代链的事件句都描述了这篇文章的主题--“前巴基斯坦舞女自杀和受伤”事件。(3)事件的语义信息更加丰富，例如“Former Pakistani dancing girl”该实体能够很清晰的指向某个具体的人，而“suicide”触发的事件却包含了时间、地点、参与者等等信息。

针对以上问题，本文提出一种新颖的事件指代消解平台，该平台在使用词级别的特征同时融合了篇章级别的特征信息，并利用双向循环神经网络的变体和分层注意力机制学习不同层次的语义信息，捕捉与篇章主题相关度高的内容。同时，借助 CRF 序列化标注，将篇章内容划分为非事件句，只构成单链的非同指事件句，以及构成指代链的同指事件句（以下简称非事件句，单链，同指链），尽可能保留更多的同指事件句参与后期的指代消解任务，从而有效的提升模型的整体性能。

1　相关工作

事件指代消解任务起步晚，研究少，主要受限于语料资源的匮乏。随着 ACE2005 语料对事件标注的扩充，TAC KPB 相关评测任务的展开，以及跨篇章事件同指语料 ECB，ECB+的公布，出现了事件指代消解任务的语料资源，为该务的研究奠定了基础。

目前，针对事件指代消解任务的研究思路主要有基于管道模型和基于联合学习模型。其中基于管道模型的处理方法是先进行事件识别（Event Detection），再研究事件指代消解，如 Liu[2]先通过 Bi-GRU 抽取事件，再通过浅层先行树计算同指事件。 Lu[3]将指代消解分成事件类型识别，时态识别，事件同指三个模

块分别进行研究，Nguyuen[4]也同样采用管道模型的思路，利用深度神经网络对该任务进行研究。管道模型符合常见的指代消解研究思路，但具有一个显著的缺点即错误传播，事件识别任务性能的好坏会很大程度上影响后续的指代消解任务。也有研究人员在假设事件都被正确抽取的情况下，将指代消解部分作为一个单独的子任务，如 Chen[5]等人抽取词性、事件论元等特征作为事件描述对的特征输入到最大熵模型中构建指代消解平台，Lu[6]利用卷积神经网络判断事件描述对之间是否具有同指关系。常见的基于联合学习的事件指代消解模型，有 Lu 和 Ng[7]于 2017 提出的结合事件识别，事件先行词识别的联合学习，也有 Lee[8]在跨篇章语料中结合实体指代任务的联合学习方法。联合学习模型多是基于各项任务的特征共享，这种方法往往需要构建大量的特征工程。

2017 年 Lee[9]在实体指代消解任务中提出了一种基于端到端的神经网络模型，该模型在实体指代消解任务上取得了当时最好的性能，并且后续仍有不少研究人员[10-12]的工作是在该平台基础上进行改进，因此，我们将该方法迁移到事件指代消解任务研究中，并结合上述对事件同指情况的分析，搭建一种融合篇章表征的事件指代消解平台。

需要说明的是，由于各类语料库对事件同指的定义稍有区别，这里，我们遵循 TAC KBP 中 Event Nugget Detection and Coreference 对该任务的定义，只针对篇章内的事件指代消解任务进行研究。

2 事件指代消解模型框架

融合篇章表征的事件指代消解模型框图如下图 2 所示，该模型沿用 Lee[9]针对实体指代的基本思路，首先对输入文本的每个 Token 进行特征表征 $\boldsymbol{w}_d(1 \le d \le N)$，随后从文本中抽取出满足条件的候选元素 $\boldsymbol{p}_k$，并通过打分函数 Span Score 保留得分最高的候选元素作为事件描述，参与最终的事件描述对打分（Event Mention Pair Score）。事件描述对打分模块如图 4 所示，将保留的事件描述两两组成事件描述对，最终判断这些事件描述对的同指关系。此外，针对事件指代消解任务特点和难点对实体指代模型进行了修改，最终构建的事件指代消解的模型，主要包括以下几个部分：（1）篇章级别的特征表征；（2）指代消解模块；（3）利用 CRF 区分非事件句，单链以及同指链进而优化模型；

以下分别详细介绍模型的三个模块。

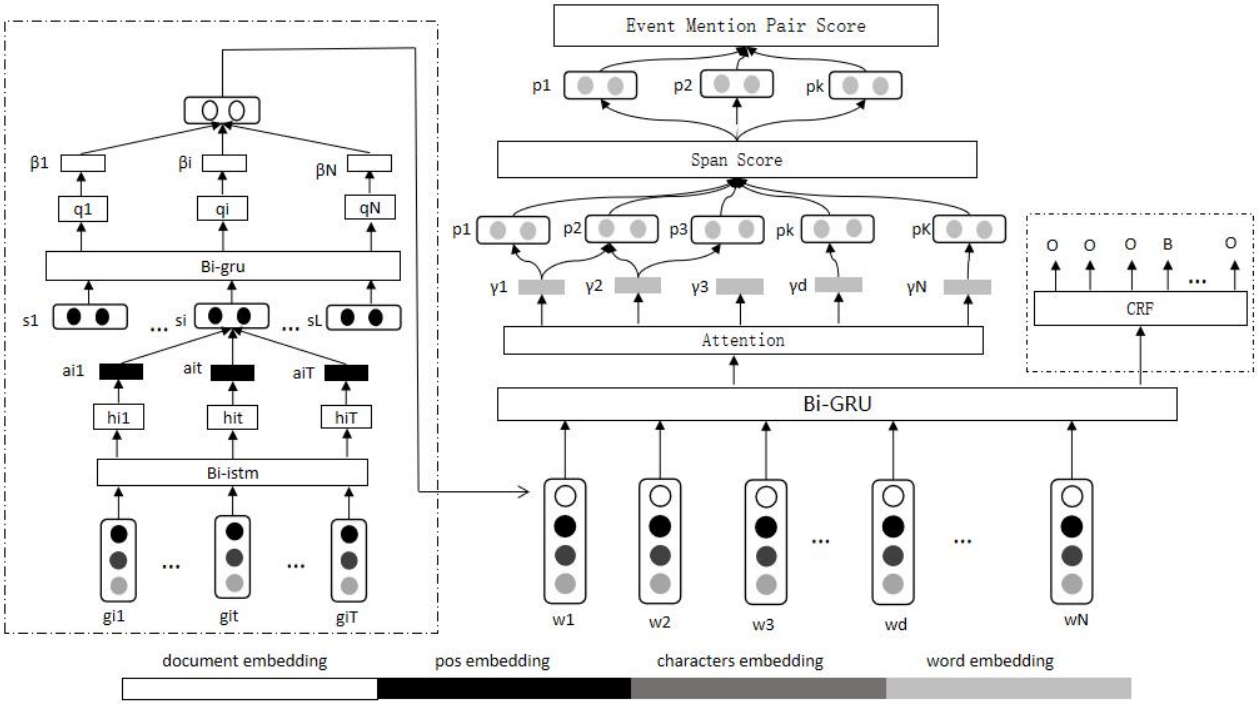


图 2　融合篇章表征的事件指代消解模型框架图

Fig.2　Model Framework Diagram of Event Coreference Resolution with Document Representation

2.1 篇章级别的特征表征

为了获得 Document embedding，模型使用了分层注意力机制如图 2 所示，它是由 Bi-LSTM 编码输出的词级别表征和 Bi-GRU 编码输出的句级别的表征组成，并且分别通过 Attention[13]机制捕捉触发词语义信息和事件句语义信息。

2.1.1 词级别表征

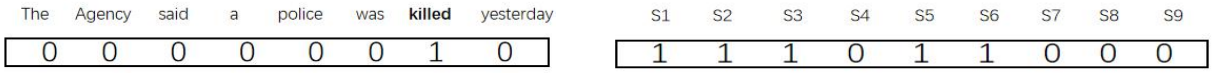
假设文本包含 L 个句子，给定一个句子 $\boldsymbol{s}_i(1 \le i \le L)$，其中每个 Token 的向量表征 $\boldsymbol{g}_{it}(1 \le t \le T)$ 是由向量 Word embedding[14]，字符向量的 Characters embedding[15]，经过斯坦福词性标注工具标注后训练得到的词性向量 Pos Embedding 拼接而成。经过 Bi-LSTM 编码得到隐层状态的输出 $\boldsymbol{h}_{it}$：

$$\boldsymbol{h}_{it} = [\vec{Lstm}(\boldsymbol{g}_{it}), \overleftarrow{Lstm}(\boldsymbol{g}_{it})], \tag{1}$$

接着模型利用 Attention 机制为每个隐层输出分配新的权值 α_{it}，同时为了使句子中的触发词能够获得更多的注意力，本文构建了一个标准的词级别注意力权值 α_{it}^*，如图 3(a)所示，将是触发词的 Token 权重设为 1，否则为 0，并计算 α_{it} 与 α_{it}^* 之间的损失 Loss_h。

$$\alpha_{it} = \mathrm{softmax}(\boldsymbol{C}_w^T(\tanh(\boldsymbol{W}_w\boldsymbol{h}_{it} + \boldsymbol{b}_w))), \tag{2}$$

$$\mathrm{Loss}_h = \sum_{i=1}^{L}\sum_{t=1}^{T}(\alpha_{it}^* - \alpha_{it}), \tag{3}$$



(a) (b)

图 3 标准注意力权值

Fig.3 Standard Attention Weight

2.1.2 句级别表征

经过上述的处理，模型加权求和获得每个句子的向量表征 $\boldsymbol{s}_i$，为了减少模型在训练时需要学习的参数，本文使用循环神经网络的另一种变体 Bi-GRU 来编码获得句向量的隐层输出 $\boldsymbol{q}_i$。

$$\boldsymbol{s}_i = \sum_{t=1}^{T}\alpha_{it}\boldsymbol{h}_{it}, \tag{4}$$

$$\boldsymbol{q}_i = [\vec{Gru}(\boldsymbol{s}_i), \overleftarrow{Gru}(\boldsymbol{s}_i)], \tag{5}$$

同样使用 Attention 机制为每个句子分配新的权重 β_i，并加权求和得到篇章表征 $\boldsymbol{d}$。为了凸显文本中的事件句，我们同样计算标准句级别注意力权值 β_i^* 与模型预测值 β_i 之间的损失 Loss_s，β_i^* 如图 3(b)所示，它对应图 1 中的 1-9 句的事件的分布情况，包含事件的句子设为 1，否则为 0。

$$\beta_i = \mathrm{softmax}(\boldsymbol{C}_s^T(\tanh(\boldsymbol{W}_s\boldsymbol{q}_i + \boldsymbol{b}_s))), \tag{6}$$

$$\boldsymbol{d} = \sum_{i=1}^{L}\beta_i\boldsymbol{q}_i, \tag{7}$$

$$\mathrm{Loss}_s = \sum_{i=1}^{L}(\beta_i^* - \beta_i), \tag{8}$$

2.2 事件指代消解

2.2.1 候选元素抽取

模型将融合了 Document embedding 的向量表征 $\boldsymbol{w}_d$ 输入到 Bi-GRU 网络中编码学习篇章的上下文信息得到隐层输出 $\boldsymbol{x}_d$。

$$\boldsymbol{x}_d = [\vec{Gru}(\boldsymbol{w}_d) + \overleftarrow{Gru}(\boldsymbol{w}_d)], \tag{9}$$

经统计，我们发现 KBP 语料中触发词长度范围集中在单个 Token 以及两个 Token 组成的词组，很少出现三个及三个以上的 Token 组成的词，并且触发词的词性的分布也具有特殊性性，因此，我们选取了篇章中长度小于 3 且具有某些词性[1]单词作为候选元素。假设篇章中候选元素的个数为 K，我们用 $start(k)$，$end(k)$ 分别表示每个候选元素 $\boldsymbol{p}_k$ $(1 \le k \le K)$ 的起始位置和结束位置。

[1] V*，NN，NNS，NNP，PRP，TO，JJ，DT，IN，RB，RP，AD（其中 V*表示所有动词）

$$\gamma_d = W_a \cdot \mathrm{FFNN}_a(\boldsymbol{x}_d), \tag{10}$$

$$\gamma_{k,d} = \frac{\exp(\gamma_d)}{\sum_{e=start(k)}^{end(k)} \exp(\gamma_k)}, \tag{11}$$

$$\hat{\boldsymbol{x}}_k = \sum_{d=start(k)}^{end(k)} \gamma_{k,d} \cdot \boldsymbol{x}_d, \tag{12}$$

模型使用了 Lee[9]的方法，利用前馈神经网络 FFNN_a 为每个 Token 分别新的权重 γ_d，在 softmax 时只对候选元素进行计算并加权求和得到向量 $\hat{\boldsymbol{x}}_k$，最终候选元素的向量表征 $\boldsymbol{p}_k$ 由特征向量 $\boldsymbol{\phi}(k)$，$\hat{\boldsymbol{x}}_k$ 以及隐层输出 $\boldsymbol{x}_d$ 共同构成，其中 $\boldsymbol{\phi}(k)$ 表示每个候选元素的长度大小。

$$\boldsymbol{p}_k = [\boldsymbol{x}_{start(k)}, \boldsymbol{x}_{end(k)}, \hat{\boldsymbol{x}}_k, \boldsymbol{\phi}(k)], \tag{13}$$

2.2.2 事件描述对同指判断

模型通过一个标准的前馈神经网络 FFNN_m 对每个候选元素打分，打分函数为 $\mathrm{S}_m(k)$，并设定阈值 $\lambda(\lambda = 0.2)$ 来保留得分最高的候选元素作为事件描述参与后期的指代消解。模型按照文本出现的先后顺序对事件描述排列，每个事件描述都要与其先行词对构成事件描述对，再利用一个前馈神经网络 FFNN_d 对每个事件描述对打分，打分函数为 $\mathrm{S}_d(u,k)(1 \le u \le k-1)$。

$$\mathrm{S_m}(k) = W_m \cdot \mathrm{FFNN}_m(\boldsymbol{p}_k), \tag{14}$$

$$\mathrm{S}_d(u,k) = \boldsymbol{W}_d \cdot \mathrm{FFNN}_d([\boldsymbol{p}_u, \boldsymbol{p}_k, \boldsymbol{p}_u \circ \boldsymbol{p}_k, \boldsymbol{\phi}(u,k)]), \tag{15}$$

如图 4 所示，模型在对事件描述对打分的过程中加入了文本类型(Genre)、事件类型(Type)以及时态属性(Realis)特征，其中时态特征和事件类型特征采用二进制编码方式表示两个事件描述对的时态属性和事件类型是否一致，$\boldsymbol{\phi}(u,\ k)$ 表示这三个特征向量的拼接。最终每个事件描述对的得分是由 $\mathrm{S}_m(k)$，$\mathrm{S}_m(u)$，$\mathrm{S}_d(u,k)$ 三个打分函数来共同决定的。模型最终的学习目标是使得预测的同指分布情况越接近标准的先行词元素集合 $\mathrm{GOLD}(i)$，及损失函数 $\mathrm{Loss}_{cluster}$ 的结果最小化。

$$\mathrm{S}(u,k) = \mathrm{S}_m(u) + \mathrm{S}_m(k) + \mathrm{S}_d(u,k), \tag{16}$$

$$\mathrm{Loss}_{\mathrm{cluster}} = -\sum_{k=1}^{K} \log \frac{\sum_{u' \in GOLD(k)} \exp(S(u',k))}{\sum_{u=\epsilon,0,\ldots,j-1} \exp(S(i,j))}, \tag{17}$$

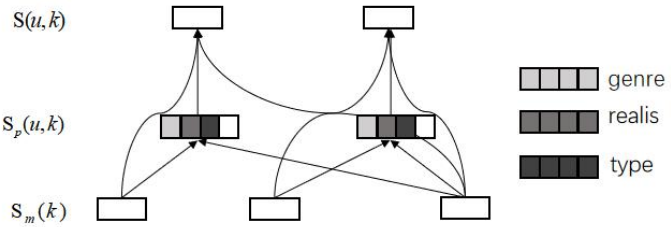


图 4　事件描述对打分

Fig.4　Event mention pair Score

2.3　非事件句，单链和同指链的区分

针对篇章中事件分布稀疏，单链数量多的问题，模型在 Bi-GRU 层上加入 CRF（Conditional random field algorithm）序列化标注，采用 BIO 的标注方式对篇章进行分类，如图 2 所示。

CRF 层的参数是由一个转移矩阵 $\boldsymbol{A}$ 构成，$\boldsymbol{A}_{x.y}$ 表示从第 x 个标签到第个 y 标签的转移得分，假设篇章的标签序列为 $\mathrm{Y} = \{y_1, y_2, y_3, y_4, ..., y_N\}$，模型对标签序列 Y 的打分函数为 $\mathrm{score(Y)}$。整个篇章的打分等于各个位置的打分之和，而每个位置的打分一部分是由 Bi-GRU 的输出 $\boldsymbol{x}_d$ 决定，另一个部分是由 CRF 的转移矩阵 $\boldsymbol{A}$ 决定，模型使用函数 softmax 得到归一化后的概率，并计算其损失得分 Loss_{crf}。

$$\text{score}(Y)=\sum_{d=1}^{N}\boldsymbol{x}_{d,y_d}+\sum_{d=1}^{N+1}\boldsymbol{A}_{y_{d-1},y_d}, \tag{18}$$

$$\text{Loss}_{\text{crf}}=-\log\frac{\exp(score(x,y))}{\sum_{y'}\exp(score(x,y'))}, \tag{20}$$

最终，模型利用门控机制控制各项损失的权重得到损失函数 Loss 的值，其中 $a,b,c,d\in[0,1]$。

$$\text{Loss}=a\text{Loss}_{\text{crf}}+b\,\text{Loss}_{cluster}+c\,\text{Loss}_{h}+d\,\text{Loss}_{s}, \tag{21}$$

$$a+b+c+d=1, \tag{22}$$

3 实验结果分析

3.1 实验设置

实验采用了 Lu[7]在英文语料上的数据集配置，包括选自 LDC2015E29，E68，E73，E94 的 648 篇作为训练语料，其中 509 篇作为训练集，139 篇作为验证集，LDC2016E72 中的 168 篇作为测试语料。并使用 MUC[16]，B^3[17]，CEAF_e[18]，BLANC[19]评测指标以及它们的均值 CoNLL，AVG-F 来对模型性能进行评测。

$$\text{CoNLL}=(\text{MUC}+\text{CEAF}_e+\text{B}^3)/3, \tag{23}$$

$$\text{AVG-F}=(\text{MUC}+\text{CEAF}_e+\text{B}^3+\text{BLANC})/4, \tag{24}$$

模型在对 Token 进行向量表征时使用的 Word Embedding 和 Characters Embedding 分别是已训练的 300 维 GloVe，以及通道数为 100，卷积核大小为 5 的卷积神经网络训练得到的 100 维向量。在训练篇章表征时，模型分别使用了 Bi-LSTM 和 Bi-GRU，我们对这两层的输出分别进行线性变换，变换矩阵 $\boldsymbol{W}_w$ 和 $\boldsymbol{W}_s$ 的输出维度为 600 和 400。此外，考虑到计算成本，模型在构建事件对时，将最大先行词匹配数量设置为 $M=150$，筛选后的候选元素个数为 Q，先行候选元素的搜索空间设为 $\min\{Q,M\}$。表 1 给出了本实验详细的参数配置，在 Features 一栏中，Span size 表示候选元素的长度大小，Genre 指 KBP 语料的文本类型，分别有新闻类文本（newswire）和论坛类文本（discussion fourms）。

表 1 实验参数配置

Table 1 Configuration of experiment parameters

Word	Bi-LSTM	Hidden dim:300	Number layers:1	Dropout: 0.5
Sentence	Bi-GRU	Hidden dim:200	Number layers:1	Dropout: 0.5
Document	Bi-GRU	Hidden dim:200	Number layers:1	Dropout: 0.2
$\text{FFNN}_{m,p}$		Hidden dim:150	Number layers:2	Dropout: 0.2
Feature	Pos Span size Type Realis Genre			Embedding dim: 20
Optimizer			Adam	
Learning			0.0001	

3.2 实验结果及分析

实验结果如表 2 所示，本文做了以下几组对比实验，其中 Baseline 是搭建的一个基准平台，它是没有加入 Doc embedding 和 CRF 序列化标注等工作的实验结果，同时我们用 Bi-LSTM 代替了 Bi-GRU 学习篇章的上下文信息。Doc embedding 和 CRF 两组实验是指分别在 Baseline 的基础上加入了篇章表征 $\boldsymbol{d}$ 和 CRF 序列化标注，最后一组实验 Doc+CRF 就是我们最终的事件指代消解模型，从表中可以看到。

（1）与 Start-of-the-art 的 Lu 联合学习模型相比，基准模型 Baseline 虽然在部分评测指标下，例如 B^3

和$CEAF_e$的结果要略逊一筹，但是从综合评测CoNLL和AVG-F来看，Baseline模型的性能都高出4个点。联合学习模型的核心思想是通过多任务间的特征共享，以及建立任务任务间的相互约束来提高各项任务的性能。在Lu的联合学习模型中不仅对每个单任进行特征抽取，同时两两任务之间也构建了大量的特征，这些特征工程费时费力，我们提出的模型无须构建复杂的人工特征，只需几种常见的词向量和词性向量就能够达到很好的效果；

（2）由于事件包含的语义信息远远多于实体，且文本中的同指事件句分布大多跨句子，因此篇章级别的语义信息有助于模型理解整篇文本的语义内容，同时分层attention能够帮助模型更好的捕捉事件句和触发词。通过对比Baseline与Doc embedding的实验结果，验证了篇章语义信息对本任务的作用，在各项评测结果下性能都得到了提升，其中综合评分提升了1.6个点左右；

（3）通过之前的分析可以知道，相较于实体，触发词分布稀疏且单链的数量较多成事件指代消解任务研究的一个难点，并且在本任务中，我们并不关注于事件的具体类型，只关注与同指链之间的关系，且单链不参与后期的消解过程，因此我们利用CRF对模型进行优化。在对比Baseine与CRF的实验结果，则证实了有效区分篇章中的非事件句、单链及同指链能够提升事件指代消解任务性能的猜想，在综合评分下提升了2.4个点左右；

（4）最后观察Doc+CRF这组实验，它的性能比Doc embedding和CRF两组都有所提升，这证明两种方法之间是相互兼容的，都能够同时对本任务起到一定的作用。同时观察最终模型Doc+CRF在各项评测指标下的性能都超越了Lu，这证明了我们模型在事件指代消解任务上的有效性

表2 事件指代消解实验结果

Table 2 Experimental results of event resolution

	MUC	B^3	$CEAF_e$	BLANC	CoNLL	AVG-F
Lu 等	27.41	40.90	39.00	25.00	35.77	33.08
Baseline	43.47	38.93	37.30	30.56	39.90	37.57
Doc embedding	45.15	40.33	39.17	31.97	41.55	39.16
CRF	45.30	41.16	40.46	33.68	42.31	40.15
Doc+CRF	46.90	42.09	40.22	34.60	43.07	40.95

4 总结

我们提出了一种融合篇章表征的事件指代消解平台，该平台将端到端的基于神经网络的实体指代消解模型很好的迁移到事件指代消解任务中，并通过比较事件与实体的区别对模型进行改进。通过分析语料发现，事件承载的语义内容要远远比实体复杂，并且篇章中的事件数量有效，且大多只被提及一次，不会与其他事件句构成同指关系，针对这些问题，我们首先加入篇章特征来丰富语义表征，并且在加入篇章特征的过程中，利用分级attention机制，抓住句子中的触发词语义和篇章中的事件句语义信息，同时利用CRF序列化标注区分哪些是不构成指代关系的事件，从而更多的保留同指事件描述参与后期的指代消解任务。经过各项评测表明，我们对事件指代消解任务的分析以及模型的改进是正确和有效的。然而模型本身还存在着不足和需要改进的地方，例如本模型在后期判断与先行词的同指关系时采用了贪心组合的方式，即每个候选元素都要与其前面的所有先行词两两组队，这种方式所带来的弊端是导致事件对的数目以平方数量级增多，并且事件对中负例的数量要远远多于正例，在后期的研究中，我们可以通过设置规则约束减少一些不必要的事件配对，从而使模型达到更好的性能。

参考文献

[1] Vendler Z. Verbs and Times. Philosophical Review, 1957, 66(2):143-160

[2] Liu Z, Araki J, Mitamura T, et al. CMU-LTI at KBP 2016 Event Nugget Track. TAC, 2016(11): 85

[3] Lu J, Ng V. UTD's Event Nugget Detection and Coreference System at KBP 2016. TAC, 2016

[4] Nguyen T H, Meyers A, Grishman R. New York University 2016 System for KBP Event Nugget: A Deep Learning Approach. TAC, 2016

[5] Chen Z, Ji H, Haralick R. A Pairwise Event Coreference Model, Feature Impact and Evaluation for Event Coreference Resolution. Association for Computational Linguistics, 2009:17-22

[6] 陆震寰,孔芳,周国栋. 面向多语料库的通用事件指代消解. 中文信息学报,2018,32(1):51-58

[7] Lu J, Ng V. Joint Learning for Event Coreference Resolution. Meeting of the Association for Computational Linguistics, 2017: 90-101

[8] Lee H, Recasens M, Chang A, et al. Joint entity and event coreference resolution across documents. Association for Computational Linguistics, 2012: 489-500

[9] Lee K, He L, Lewis M, et al. End-to-end neural coreference resolution. arXiv preprint arXiv:1707.07045, 2017

[10] Zhang R, Santos C N, Yasunaga M, et al. Neural coreference resolution with deep biaffine attention by joint mention detection and mention clustering. arXiv preprint arXiv:1805.04893, 2018

[11] Luo H, Glass J. Learning Word Representations with Cross-Sentence Dependency for End-to-End Co-reference Resolution. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018: 4829-4833

[12] Lee K, He L, Zettlemoyer L. Higher-order coreference resolution with coarse-to-fine inference. arXiv preprint arXiv:1804.05392, 2018

[13] Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate. Computer Science, 2014

[14] Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014: 1532-1543

[15] Turian J, Ratinov L, Bengio Y. Word representations: a simple and general method for semi-supervised learning. Proceedings of the 48th annual meeting of the association for computational linguistics. Association for Computational Linguistics, 2010: 384-394

[16] Vilain M, Burger J, Aberdeen J, et al. A model-theoretic coreference scoring scheme. Proceedings of the 6th conference on Message understanding. Association for Computational Linguistics, 1995: 45-52

[17] Bagga A, Baldwin B. Algorithms for scoring coreference chains. The first international conference on language resources and evaluation workshop on linguistics coreference, 1998: 563-566

[18] Luo X. On coreference resolution performance metrics. Proceedings of the conference on human language technology and empirical methods in natural language processing. Association for Computational Linguistics, 2005: 25-32

[19] Larta R, Eduard H, BLANC: Implementing the Rand Index for coreference evaluation. Natural Language Engineering, 17(4):485– 510