

文章编号: 1003-0077(2019)08-0077-07

基于混合神经网络的实体和事件联合抽取方法

吴文涛^{1,2}, 李培峰^{1,2}, 朱巧明^{1,2}

- (1. 苏州大学 计算机科学与技术学院, 江苏 苏州 215006;
2. 江苏省计算机信息技术处理重点实验室, 江苏 苏州 215006)

摘要: 实体和事件抽取旨在从文本中识别出实体和事件信息并以结构化形式予以呈现。现有工作通常将实体抽取和事件抽取作为两个单独任务, 忽略了这两个任务之间的紧密关系。实际上, 事件和实体密切相关, 实体往往在事件中充当参与者。该文提出了一种混合神经网络模型, 同时对实体和事件进行抽取, 挖掘两者之间的依赖关系。模型采用双向 LSTM 识别实体, 并将在双向 LSTM 中获得的实体上下文信息进一步传递到结合了自注意力和门控卷积的神经网络来抽取事件。在英文 ACE 2005 语料库上的实验结果证明了该文方法优于目前最好的基准系统。

关键词: 事件抽取; 实体抽取; 自注意力; 门控卷积神经网络

中图分类号: TP391 **文献标识码:** A

Joint Extraction of Entities and Events by a Hybrid Neural Network

WU Wentao^{1,2}, LI Peifeng^{1,2}, ZHU Qiaoming^{1,2}

- (1. School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China;
2. Provincial Key Laboratory for Computer Information Processing Technology, Suzhou, Jiangsu 215006, China)

Abstract: Entity and event extraction aim at detecting entities and events from text, respectively. Previous studies in information extraction usually took entity extraction and event extraction as two separate tasks without capturing the close relationship between the two tasks. This paper proposes a hybrid neural network to simultaneously extract the entity and the event, and exploit the dependencies between them. This network first uses encoder-decoder bidirectional LSTM module to identify entities, and then introduces the entity context information from the above bidirectional LSTM module to a neural network, which combines self-attention and gated convolution to facilitate event extraction. Experimental results on the ACE 2005 English corpus show that our model outperforms the state-of-the-art methods.

Keywords: event extraction; entity extraction; self-attention; gated convolutional neural networks

0 引言

实体(Entity)和事件(Event)抽取是信息抽取(Information Extraction)的两个子任务, 这两个子任务密切相关。事件抽取的任务是从文本中抽取预先定义好的各种类型事件实例(Event Mention)及其论元(Argument)^①。其中, 实体是事件的核心组成部分, 实体抽取任务有助于准确抽取事件。

例1给出了一个事件实例, 其中, “reaching”是

事件的触发词, 所触发的事件类型为 Attack(攻击)。“Scud missiles”是其中的一个实体, 其实体类型是 WEA(武器), 充当的角色是 Instrument(工具)。这个实体的识别明显有助于识别该事件为 Attack(攻击)类型事件。因为多数攻击类型事件都有攻击的武器。

例1 Iraq believed to have Scud missiles capable of **reaching** Baghdad.

现有的工作大多数将实体抽取和事件抽取作为

^① 事件的参与者和属性, 由实体实例组成。

两个单独任务,很少关注两个子任务的相关性。在事件抽取中,绝大多数现有工作都假设文本中的实体已知^[1-5],但这在实际应用中并不成立。Li 等^[6]先使用命名实体识别工具识别出人物、组织和地点等实体实例,然后再将这些抽取的实体作为输入来抽取事件,这往往会导致错误传递。在上述例子中,若命名实体识别工具识别不出“Scud missiles”是武器类实体,或错误地将其识别为人物,就可能导致事件抽取系统无法正确抽取该 Attack 事件。另外,少数研究为实体抽取和事件抽取建立联合学习模型,但这些模型往往基于特征工程,依赖复杂人工特征。另外,这些模型严重依赖于其他任务(如句法分析和依存分析等),这也会导致级联错误。

为了解决上述问题,本文提出了一个联合实体抽取和事件抽取的混合神经网络模型(Hybrid Neural Networks for Entity and Event Extraction, HNN-EE)。该模型的核心是实体抽取和事件抽取共享一个双向 LSTM 层,通过底层共享参数,互相促进学习,获得实体和事件之间丰富的关联信息。此外,该模型还采用线性(Conditional Random Field, CRF)层结构来模拟标签之间的交互关系来解码整个句子的标签,并引入自注意力(Self Attention)机制和门控卷积神经网络(Gated Convolutional Neural Networks)来捕获任意词之间的关系和提取局部信息,实现信息的多通道融合。在英文 ACE 2005 语料库上的实验结果表明,本文的方法明显优于目前最先进的基准系统。

1 相关工作

目前,事件抽取的相关研究工作虽然很多,但为了降低任务复杂性,绝大多数工作假设文档中的实体已经被识别。例如,Ahn^[1]使用词汇、句法特征以及外部知识库来抽取事件。Hong 等^[2]充分利用实体类型的一致性特征,提出利用跨实体推理进行事件抽取的方法。Chen^[3]等将触发词抽取和论元抽取作为两个整体任务来抽取中文事件,进而防止错误传递。Li 等^[6]采用基于结构化感知机的联合模型,将触发词抽取和论元抽取看作一个整体的序列标注任务。Liu 等^[7]利用概率软逻辑模型来编码全局信息,进一步提升了事件抽取性能。

近年来,随着深度学习的发展,更多的神经网络模型被引入到事件抽取。Nguyen、Grishman^[4]、以及 Chen 等^[5]运用卷积神经网络来避免复杂的特征

工程。Nguyen 等^[8]提出基于双向循环神经网络模型,同时抽取触发词和论元。Sha 等^[9]针对双向循环神经网络模型没有有效地利用依存句法信息,提出了依赖桥的双向 LSTM 模型,充分利用句法信息来抽取事件,性能得到明显提升。Liu 等^[10]利用已经标注的论元信息,提出了基于有监督的注意力机制事件抽取方法。

目前,仅有少数的工作从生文本(实体信息未知)中抽取事件。Li 等^[11]提出一个基于结构化预测模型,同时抽取实体,关系和事件。Yang 等^[12]提出篇章内的事件和实体抽取联合模型,通过采用联合因子图模型来联合学习每个事件内部的结构化信息、篇章内不同事件间的关系和实体信息,明显提高了实体和事件的抽取性能。不同于 Yang 等^[12]的工作依赖大量的人工特征,本文的一个特点是采用混合神经网络模型通过底层共享来挖掘实体和事件抽取两个任务间的依赖关系。

2 基于混合神经网络的实体和事件抽取模型(HNN-EE)

为了解决管道模型中的级联错误和传统方法中依赖人工特征的问题,本文提出了一个混合神经网络模型 HNN-EE,其架构如图 1 所示。

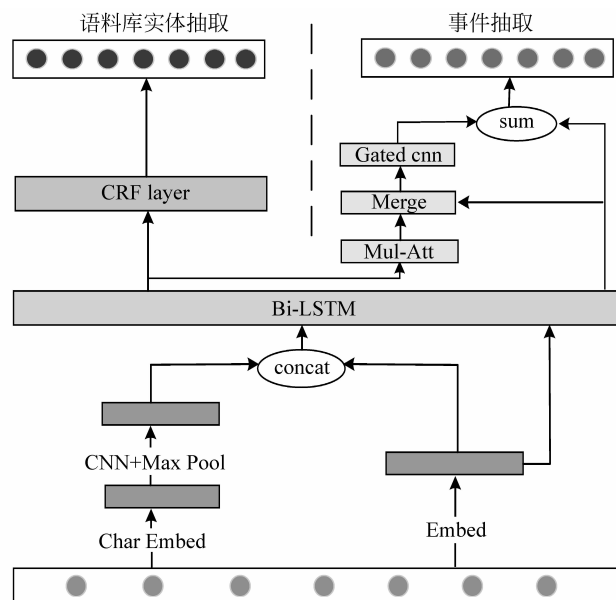


图 1 HNN-EE 模型的架构

HNN-EE 主要通过实体和事件抽取两个任务在浅层共享参数,协同训练,互相促进,提高性能。该模型主要包含以下 3 个部分:①输入和编码层;

②实体抽取模块；③事件抽取模块。首先，在输入和编码层部分，输入包含了词向量和经过卷积、池化后的字符向量。编码层双向 LSTM 可以将序列中每个词的上下文信息从前向和后向两个方向很好地保留并传递下去，提取序列中的全局信息。实体抽取任务和事件抽取任务共享双向 LSTM 编码层，两个任务通过底层共享参数的形式共同学习，在训练时两个任务通过后向传播算法更新共享参数来实现两个任务之间的依赖。编码层后有两个通道，一个连接到实体抽取模块，由线性 CRF 层来解码，获得最佳的实体标签序列；另一个送入到自注意力层和门控卷积层，自注意力层捕获序列内部的联系，门控卷积神经网络可以控制信息的流动，捕获到重要的局部信息，与经过双向 LSTM 编码后的句子向量结合，提取出句子序列的全局和局部信息，最后进入到 softmax 层进行触发词识别和分类。

2.1 输入和编码层

在输入层中，对于句中每个词，本文使用了预训练的词向量^[13]进行初始化，作为事件抽取模块的输入。Zadrozny^[14]的研究已经证明卷积神经网络能够从词的字符表示中有效提取出形态学信息（如词的前缀或后缀）。所以本文对字符向量进行卷积和全局最大池化操作后获得每个单词的字符表示向量，然后字符表示向量与词向量拼接在一起传入到下一层双向 LSTM 中，作为实体抽取模块的输入。在训练过程中，本文不再更新词向量的参数。

双向 LSTM 编码层包括前向 LSTM 层、后向 LSTM 层和它们的拼接层。双向 LSTM 中的一个输入序列可以表示为 $x = \{x_1, \dots, x_t, \dots, x_n\}$ ，其中 $x_t \in \mathbb{R}^d$ 是句子中第 t 个词的 d 维向量， n 是给定句子的长度。然后，经过两个平行的 LSTM 层（前向 LSTM 层和后向 LSTM 层），分别从相反的方向对输入进行编码，捕获序列中词之间的长远依赖关系。对于每个词 x_t ，前向 LSTM 层和后向 LSTM 层各自产生一个隐藏层的单元输出 $\vec{h}_t = \text{LSTM}_F(x_t, \vec{h}_{t-1})$ 和 $\overleftarrow{h}_t = \text{LSTM}_B(x_t, \overleftarrow{h}_{t+1})$ ，两个向量中包含了句子中任意词的上下文信息。最后将两个向量拼接在一起，得到最终的隐藏层向量 $h_t = [\vec{h}_t; \overleftarrow{h}_t]$ 。双向 LSTM 共享层提取到两个任务共同的依赖特征，然后进入到不同的网络层中，提取对应任务的独立特征。

2.2 实体抽取模块

对于实体抽取，本文使用 BIO 标签模式（Begin：实体开始单词，Inside：实体其余单词，Outside：非实体单词）为每个词赋予一个实体标签，每个标签包含了实体中单词的位置信息。虽然双向 LSTM 能够捕获到长距离的依赖信息，但是对于序列标注（或一般结构化预测）任务，有必要考虑相邻标签之间的依赖性。比如，I-ORG（类型为组织的实体非开头词）后不能跟随 I-PER（类型为人物的实体非开头词）。因此本文对标签序列使用线性 CRF 建模，而不是对每个标签独立解码。

本文使用 $y = \{y_1, \dots, y_n\}$ 来表示标签序列， $Y(x)$ 表示 y 的可能标记序列集合。线性 CRF 的概率模型定义为 $p(y|x; W, b)$ ，如式(1)所示。

$$p(y|x; W, b) = \frac{\prod_{i=1}^n \varphi_i(y_{i-1}, y_i, X)}{\sum_{y' \in Y(x)} \prod_{i=1}^n \varphi_i(y'_{i-1}, y'_i, X)} \quad (1)$$

其中， $\varphi_i(y', y, x) = \exp(W_{y', y}^T z_i + b_{y', y})$ 是隐含函数， $W_{y', y}^T$ 和 $b_{y', y}$ 对应着标签对 (y', y) 的权重矩阵和偏置值。

解码时搜索序列标签 y^* 的最高条件概率，如式(2)所示。

$$y^* = \operatorname{argmax}_{y \in Y(x)} P(y|x; W, b) \quad (2)$$

对于线性 CRF 模型（仅考虑两个连续标签之间的交互），采用维特比算法可以有效进行解码。

2.3 事件抽取模块

事件抽取模块包括自注意力层和门控卷积层，自注意力层主要学习序列内部的词依赖关系，捕获序列的内部结构；门控卷积层主要控制信息流动，提取更高层的特征。具体如下。

2.3.1 自注意力层

当实体抽取模型识别出实体及其类别，本文将实体的编码信息送入到自注意力层，它只需要序列本身就可以计算其表示，也是一个序列编码层，寻找序列内部的联系。Attention 的如式(3)所示。

$$\text{Attention}(Q, K, V) = \operatorname{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (3)$$

其中，矩阵 $Q \in \mathbb{R}^{n \times d}$ 、矩阵 $K \in \mathbb{R}^{n \times d}$ 和矩阵 $V \in \mathbb{R}^{n \times d}$ ， d 是网络中隐藏神经元的数量，该层主要在序列内部计算每个词与其他词的相似度，寻找任意两个词之间的联系。

本文引入了 Vaswani 等^[15]提出的多头注意力 (Multi-Head Attention) 结构, 将输入向量矩阵 $\mathbf{H} \in \mathbb{R}^{n \times d}$ 通过不同的线性变换映射成矩阵 $\mathbf{Q}, \mathbf{K}, \mathbf{V}$, 然后再做计算, 这个过程重复 h 次, 将结果拼接起来, 得到编码向量, 如式(4)~式(6)所示。

$$\mathbf{M}_i = \text{Attention}(\mathbf{QW}_i^Q, \mathbf{KW}_i^K, \mathbf{VW}_i^V) \quad (4)$$

$$\mathbf{M} = \text{Concat}(\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_h) \quad (5)$$

$$\mathbf{Y} = \mathbf{MW} \quad (6)$$

其中, h 为平行头的数量。 $\mathbf{W}_i^Q \in \mathbb{R}^{n \times d/h}$ 、 $\mathbf{W}_i^K \in \mathbb{R}^{n \times d/h}$ 、 $\mathbf{W}_i^V \in \mathbb{R}^{n \times d/h}$ 和 $\mathbf{W} \in \mathbb{R}^{d \times d}$ 是对应线性变换的权重矩阵。

2.3.2 门控卷积层

Liu 等^[16]的研究表明, 卷积神经网络善于从序列对象中捕获显著的特征。本文将经过自注意力层编码后的实体语义向量和双向 LSTM 编码后的句子隐藏层向量结合在一起, 送入到卷积神经网络中, 来获取序列的局部信息, 实现信息的多通道融合。对于卷积子层, 本文采用 Dauphin 等^[17]提出的 GLU(Gated Linear Unit), 与标准卷积相比, 门控卷积神经网络在每层卷积操作后都加上一个输出门限, 控制信息的流动, 不仅有效地降低梯度弥散的可能性, 而且还保留了非线性变换的能力。因此, 它在语言建模和机器翻译等自然语言处理任务中都取得了较优的效果。给定两个卷积核 $\mathbf{W} \in \mathbb{R}^{k \times m \times n}$ 和 $\mathbf{V} \in \mathbb{R}^{k \times m \times n}$, GLU 的输出计算如式(7)所示。

$$\text{GLU}(\mathbf{X}) = (\mathbf{X} * \mathbf{W}) \odot \sigma(\mathbf{X} * \mathbf{V}) \quad (7)$$

其中, m 和 n 分别代表着输入和输出的特征图数量, k 是卷积核宽度, \odot 表示矩阵元素之间的点积, σ 是 sigmoid 激活函数。

本文使用多个卷积核来捕获实体序列中的局部特征, 它能够捕获各种粒度的 N -gram 局部语义信息, 修复模型中一些由于词歧义造成的错误, 为触发词识别和分类提供了非常重要的信息。本文使用了宽度为 k 的卷积核生成整个实体序列的局部特征 $\mathbf{X} \in \mathbb{R}^{n \times d}$, 其中 n 是给定句子的长度, d 是每个词的维度。然后与双向 LSTM 生成的隐藏层向量 \mathbf{H} 结合在一起, 作为最终的特征向量, 送入到 softmax 层中生成所有事件类型的概率分布, 选取概率最大的事件类型作为最终的结果。如式(8)、式(9)所示。

$$\mathbf{F} = \text{concat}(\mathbf{X} + \mathbf{H}) \quad (8)$$

$$\mathbf{O} = \text{softmax}(\mathbf{W}_s \mathbf{F} + \mathbf{b}_s) \quad (9)$$

其中, \mathbf{H} 是包含向量 \mathbf{h}_i 的矩阵, \mathbf{W}_s 和 \mathbf{b}_s 是 softmax 函数的权重矩阵和偏置值。为了防止过拟

合, 本文在 softmax 层前使用了 dropout 策略。

本文通过采用随机梯度下降算法来最小化负对数似然函数的方式来进行模型训练, 并采用 Adam^[18] 优化器算法来优化模型参数。

3 实验

3.1 语料和实验设置

为了评估本文方法的有效性, 本文主要在英文 ACE 2005 语料库上进行实验。英文 ACE 2005 语料库标注了 33 种事件类型(加上“NONE”类别, 本文采用 34 个类型作为预定义标签)。本文的数据划分和 Li 等^[6]一致, 选取 529 篇文档作为训练集, 40 篇新闻报道作为测试集, 剩下的 30 篇作为开发集。另外, 本文也采用 Li 等^[6]定义的事件评估指标。事件触发词抽取任务要求触发词不仅被正确识别, 并且要求触发词被赋予正确的事件类型。对于实体, 只考虑 PER(人物)、ORG(组织)、GPE(政治)、LOC(地点)、FAC(设施)、VEH(交通工具)、WEA(武器)和 ACE 的 TIME(时间)以及 VALUE(数值)表达式共 9 种实体类型, 若实体被正确识别并且实体类型一致, 则认为实体抽取正确。

为了进一步验证本文方法的有效性, 也汇报了在 TAC KBP 2015 数据集上的结果。该数据集是 KBP 2015 Event Nugget 评测人物提供的训练和测试数据, 有 38 个子类型和一个“NONE”类别。训练数据为 158 篇文档, 测试数据为 202 篇文档。

另外, 遵循前人的工作, 本文也采用准确率 P (Precision)、召回率 R (Recall) 以及 F_1 值作为实验性能指标。

本文的实验参数是在英文 ACE 2005 语料库的开发集中进行调整。为了防止模型过拟合, dropout 的比例是 0.5, 批次大小为 50。字符向量随机初始化生成, 维度为 30。卷积核宽度为 3, 数量分别为 30 和 300。本文使用预训练的词向量^[13]来初始化词向量, 维度为 300 维, 双向 LSTM 的隐藏层神经元数量为 300, 多头注意力的数量 10。

3.2 实验结果

为了验证本文的 HNN-EE 模型在实体和事件抽取上的性能, 将它和三个基准系统做比较: ①Li: Li 等^[11]提出的结构化预测框架, 该框架同时解决实体抽取、关系抽取和事件抽取三个信息抽取任务;

②Yang: Yang 等^[12]提出的实体和事件联合抽取模型,对文档中事件、实体以及它们之间的关系进行联合推理;③HNN-EE(w/o entity): HNN-EE 模型不利用任何实体信息,对触发词进行抽取。表 1 是四个系统在英文 ACE 2005 语料库上在实体和事件抽取方面的性能对比。

表 1 ACE 2005 语料库上系统性能对比(%)

系统	实体抽取			事件抽取		
	<i>P</i>	<i>R</i>	<i>F₁</i>	<i>P</i>	<i>R</i>	<i>F₁</i>
Li	85.3	76.5	80.7	70.3	60.9	65.2
Yang	82.4	79.2	80.7	75.1	63.3	68.7
HNN-EE (w/o entity)	82.8	81.4	82.1	77.2	60.7	68.0
HNN-EE	84.0	82.5	83.2	74.4	67.3	70.6

从表 1 中实体抽取的结果中可以看出:

① HNN-EE 模型优于传统模型,性能相较于 Li 和 Yang 均提高了 2.5 个百分点,这充分说明了本文方法在实体抽取方面的有效性。

② 相较于 HNN-EE(w/o entity),HNN-EE 的实体抽取性能也取得了明显提升。这表明两个任务之间存在互补性,本文的模型抓住了它们间的联系,使这两个任务互相促进。

③ HNN-EE(w/o entity)的实体抽取性能相较于 Li 和 Yang 均提高了 1.4 个百分点,这表明基于字符-词向量的双向 LSTM-CRF 模型,不但能够有效地捕获句子的全局信息,而且还考虑到相邻标签的约束信息,取得不错性能。

从表 1 中事件抽取的结果同样可以看出:

① HNN-EE(w/o entity)模型在触发词抽取中的性能较 Li 提升了 2.8 个百分点,这是因为双向 LSTM 能够捕获到句子的全局信息,对事件抽取性能提升明显;比 Yang 降低了 0.7 个百分点,这是由于在触发词抽取过程中没有充分利用到实体语义信息,忽略了实体对触发词的影响,导致性能略低。

② 本文的 HNN-EE 模型在触发词抽取中的性能均优于传统的联合推理模型。与 Li 相比 F_1 值提高了 5.4 个百分点;与 Yang 相比 F_1 值提高 1.9 个百分点。这个实验结果证明了本文方法与基于人工设计的特征联合推理方法相比,具有挖掘实体和触发词之间隐含的深层语义信息的优越性。

③ 和 HNN-EE(w/o entity)模型相比,HNN-EE 模型的 F_1 值提升了 2.6 个百分点。主要是因

为 HNN-EE 模型中的两个任务之间通过双向 LSTM 层共享参数,协同训练,利用了实体语义的编码信息,考虑到触发词和实体标签的所有组合,捕获了触发词和实体之间的依赖关系。另外,共享 LSTM 层允许信息在事件和实体之间传播,提供了更多的语义一致性信息,减少了错误传递。而在 HNN-EE(w/o entity)模型中没有获取实体的编码信息,从而无法获取到任何实体特征,导致性能下降。例如,例 2 和例 3 中的触发词都是“leave”,但例 2 中事件类型是 Transport(运输),例 3 中事件类型是 End-Position(离职),如果只单独考虑触发词“leave”,很难识别出触发词的事件类型。由事件类型和实体类型对应性分布特点可知,Transport 事件中出现的实体类型主要是 GPE(政治),End-Position 事件中出现的实体类型主要是 ORG(组织),HNN-EE 模型联合两个任务,挖掘触发词和实体之间的隐含依赖关系,两者之间相互促进,提高事件抽取性能。

例 2 Bush gave Saddam 48 hours to **leave** Baghdad.

例 3 Greenspan wants to **leave** Federal Reserve.

3.3 实验分析

为了进一步验证 HNN-EE 模型的有效性,本文还设计了 4 个对比实验。具体如表 2 所示。其中,① Pred_Entity: HNN-EE 模型使用 Stanford Named Entity Recognition 工具来识别句子中的实体信息;② No_Share: HNN-EE 模型未共享双向 LSTM 编码层;③ No_Gcnn: HNN-EE 模型未使用门控卷积神经网络;④ No_MulAtt: HNN-EE 未使用多头注意力机制。

表 2 ACE 2005 数据集上对比实验(%)

系统	事件抽取		
	<i>P</i>	<i>R</i>	<i>F₁</i>
Pred_Entity	74.2	64.6	69.1
No_Share	73.2	63.6	68.1
No_Gcnn	79.3	60.3	68.5
No_MulAtt	79.6	60.9	69.0
HNN-EE	74.4	67.3	70.6

Pred_Entity 使用工具识别出句子中的实体,导致性能下降了 1.5 个百分点。这主要是因为命名实

体识别的性能为 51.5 个百分点,远低于 HNN-EE 模型中的命名实体识别性能,许多人称代词的实体类型无法识别出。例 4 中“He”和“his”的实体类型都是 PER(人物),HNN-EE 模型均能正确识别出实体类型,使得句中包含了丰富的实体信息,事件抽取模型提取出触发词和实体之间的关系,从而正确识别出触发词“appeal”的事件类型为 Appeal(上诉)。

例 4 He lost an **appeal** case on his theft sentence on April 18.

No_Share 中未共享双向 LSTM 层,可以看出性能下降了 2.5 个百分点。这是由于模型没有通过双向 LSTM 层共享参数,两个任务单独训练,无法通过更新共享参数来实现两个子任务之间的依赖,使得事件抽取任务中无法包含两个任务的共同特征,导致包含丰富实体信息的句子中也无法抽取事件。

No_Gcnn 中删除门控卷积神经网络,未能从序列中提取局部信息,导致性能下降明显, F_1 值下降了 2.1 个百分点。

No_MulAtt 删除多头注意力机制,没有捕获序列内部词与词的依赖关系,未从序列的不同表示空间里学习到更多信息,损失了部分全局信息,导致性能下降了 1.6 个百分点。较 No_Gcnn 模型性能下降略低,原因是双向 LSTM 中的记忆模块能够充分学习到整个序列的长远依赖关系,捕获到全局信息,所以删除多头注意力机制对获取全局信息影响不大。

3.4 KBP 2015 实验性能

为了验证本文方法的有效性,在另外一个语料库 TAC KBP 2015 上做了测试。需要说明的是,KBP 语料没有标注实体,所以无法进行实体抽取任务,本文仅进行事件抽取的性能对比。为了保证一致性,本文使用的训练模型以及超参数保持不变。本文引入两个基准系统,①TAC-KBP: Hong 等^[19]提出的事件抽取模型;②GCN-ED: Nguyen 等^[20]提出的基于实体池化机制的图卷积模型,取得了目前最佳性能。实验结果如表 3 所示。

表 3 KBP 2015 数据集上系统性能对比(%)

系统	事件抽取		
	P	R	$F1$
TAC-KBP	75.23	47.74	58.41
GCN-ED	70.30	50.60	58.80
HNN-EE	67.68	51.13	58.25

从表 3 中可以看出,本文的模型与 TAC-KBP 系统相比,取得了相当的性能(-0.16 个百分点),因为本文模型中没有扩展外部数据,而 TAC-KBP 利用了外部语料资源来扩充训练数据,额外获得了更多的同质样本来辅助提高事件抽取性能。HNN-EE 模型与目前性能最好的 GCN-ED 模型相比,也取得了相当的性能(-0.55 个百分点)。这是因为本文模型较为简单,没有编码句法信息,而 GCN-ED 利用了多层图卷积网络来挖掘深层的句法语义信息,相对较为复杂。

4 总结与展望

本文提出了一个抽取句子中实体和事件的联合方法,通过模型中双向 LSTM 层共享参数,获取实体和事件之间的关系,互相学习,互相促进,捕获到各自任务的共有和私有特征。模型再通过引入自注意力机制和门控卷积神经网络来获取序列内部依赖关系和局部信息,实现信息的多层融合。在 ACE 2005 语料库上性能得到提升,并且在 TAC KBP 2015 语料库上取得了相当的性能,证明了本文方法的有效性。今后的工作重点将考虑实体、触发词和论元三者之间的联系,去挖掘事件之间的更复杂关系。

参考文献

- [1] Ahn D. The stages of event extraction[C]//Proceedings of the Workshop on Annotating & Reasoning about Time & Events. Association for Computational Linguistics, 2006: 1-8.
- [2] Hong Y, Zhang J, Ma B, et al. Using cross-entity inference to improve event extraction[C]//Proceedings of Meeting of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, 2011: 1127-1136.
- [3] Chen C, Ng V. Joint modeling for Chinese event extraction with rich linguistic features[C]//Proceedings of COLING, 2012: 529-544.
- [4] Nguyen T H, Grishman R. Event detection and domain adaptation with convolutional neural networks[C]//Proceedings of ACL-IJCNLP, 2015: 365-371.
- [5] Chen Y, Xu L, Liu K, et al. Event extraction via dynamic multi-pooling convolutional neural networks [C]//Proceedings of Meeting of the Association for Computational Linguistics, 2015: 167-176.

- [6] Li Q, Ji H, Huang L. Joint event extraction via structured prediction with global features[C]//Proceedings of Meeting of the Association for Computational Linguistics, 2013: 73-82.
- [7] Liu S, Liu K, He S, et al. A probabilistic soft logic based approach to exploiting latent and global information in event classification[C]//Proceedings of the 13th AAAI Conference on Artificial Intelligence, 2016: 2993-2999.
- [8] Nguyen T H, Cho K, Grishman R. Joint event extraction via recurrent neural networks[C]//Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016: 300-309.
- [9] L Sha, Feng Q, Chang B B, et al. Jointly extraction event trigger and arguments by dependency-bridge RNN and tensor-based argument interaction[C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 2018: 5916-5923.
- [10] Liu S, Chen Y, Liu K, et al. Exploiting argument information to improve event detection via supervised attention mechanisms[C]//Proceedings of Meeting of the Association for Computational Linguistics, 2017: 1789-1798.
- [11] Li Q, Ji H, Hong Y, et al. Constructing information networks using one single model[C]//Proceedings of Conference on Empirical Methods in Natural Language Processing, 2014: 1846-1851.
- [12] Yang B, Mitchell T. Joint extraction of events and entities within a document context[C]//Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016: 289-299.
- [13] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//Proceedings of Advances in Neural Information Processing Systems, 2013, 26: 3111-3119.
- [14] Zdrozny B. Learning character-level representations for part-of-speech tagging[C]//Proceedings of International Conference on International Conference on Machine Learning. JMLR.org, 2014: II-1818.
- [15] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of Advances in Neural Information Processing Systems, 2017: 5998-6008.
- [16] Liu Y, Wei F, Li S, et al. A dependency-based neural network for relation classification[J]. arXiv: 1507.04646v1, 2015: 285-290.
- [17] Dauphin Y N, Fan A, Auli M, et al. Language modeling with gated convolutional networks[J]. arXiv preprint arXiv, 2017.
- [18] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv:1412.6980v9, 2017.
- [19] Hong Y, Lu D, Pan X M, et al. RPI BLENDER TAC-KBP2015 System Description[C]//Proceedings of the TAC. 2015.
- [20] Nguyen T H, Grishman R. Graph convolutional networks with argument-aware pooling for event detection[C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 2018.



吴文涛(1993—), 硕士研究生, 主要研究领域为自然语言处理。

E-mail: wtwu163@stu.suda.edu.cn



朱巧明(1963—), 博士, 教授, 主要研究领域为中文信息处理、Web 信息处理。

E-mail: qmzhu@suda.edu.cn



李培峰(1971—), 通信作者, 博士, 教授, 主要研究领域为自然语言处理、机器学习。

E-mail: pfli@suda.edu.cn