

基于卷积双向长短期记忆网络的事件触发词抽取

陈 斌¹, 周 勇¹, 刘 兵^{1,2}

(1. 中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116; 2. 中国科学院电子研究所, 北京 100094)

摘 要: 传统事件触发词抽取方法在特征提取过程中过分依赖自然语言处理工具, 容易造成误差累积。为解决该问题, 在卷积双向长短期记忆网络的基础上, 提出一种事件触发词抽取方法。通过卷积操作提取单词上下文语境信息, 同时利用长短期记忆网络保留句子级别特征, 从而提高事件触发词的抽取性能。在 ACE2005 英文语料上的实验结果表明, 该方法在事件触发词识别与分类阶段的 F 值达到 69.5%, 具有较好的抽取性能。

关键词: 事件抽取; 触发词; 卷积神经网络; 循环神经网络; 自然语言处理; 特征提取

中文引用格式: 陈斌, 周勇, 刘兵. 基于卷积双向长短期记忆网络的事件触发词抽取[J]. 计算机工程, 2019, 45(1): 153-158.

英文引用格式: CHEN Bin, ZHOU Yong, LIU Bing. Event trigger word extraction based on convolutional bidirectional long short term memory network[J]. Computer Engineering, 2019, 45(1): 153-158.

Event Trigger Word Extraction Based on Convolutional Bidirectional Long Short Term Memory Network

CHEN Bin¹, ZHOU Yong¹, LIU Bing^{1,2}

(1. College of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China;
2. Institute of Electronics, Chinese Academy of Sciences, Beijing 100094, China)

[Abstract] The traditional event trigger word extraction methods over rely on Natural Language Processing (NLP) tools in the feature extraction process, which is easy to cause error accumulation. To solve this problem, a new event trigger extraction based on convolutional bidirectional Long Short Term Memory (LSTM) network is proposed. This method utilizes a convolution operation to extract word contextual information and retains sentence-level features through LSTM unit to improve the performance of event trigger word extraction. Experimental results on the ACE2005 English corpus show that the method achieves an F value of 69.5% on the event trigger word recognition and classification stage, which has a good extraction effect.

[Key words] event extraction; trigger word; Convolutional Neural Network (CNN); Recurrent Neural Network (RNN); Natural Language Processing (NLP); feature extraction

DOI: 10.19678/j.issn.1000-3428.0049801

0 概述

事件抽取是信息抽取技术的重要环节,也是信息抽取领域的难点问题。它旨在从非结构化文本中抽取用户感兴趣的事件信息并以结构化的形式呈现出来^[1]。事件抽取技术对于计算机理解文本内容具有重要作用,能够将文本分析从语言层面提升到内容层面,对语义搜索、自动问答、舆情监测等具有重要意义。因此,事件抽取近年来得到了广泛的关注。

自动内容抽取 (Automatic Content Extraction, ACE) 测评会议是事件抽取领域较有影响的评测会

议。ACE 会议定义事件抽取任务是利用自动抽取技术从非结构化的文本数据中抽取用户关注的事件信息,并转化为结构化形式呈现给用户,从而方便阅读或浏览^[2]。ACE 会议定义了事件抽取 2 个核心子任务,即触发词抽取和事件要素抽取。触发词是最能够清晰表示事件发生的词语,事件类型由触发词类型决定,因此触发词抽取是事件抽取的关键,触发词抽取结果好坏将直接影响到事件要素的抽取。

本文通过将卷积神经网络 (Convolutional Neural Network, CNN)^[3] 与长短期记忆 (Long Short Term Memory, LSTM) 网络相结合,提出一种基于卷积双

基金项目: 国家自然科学基金青年基金“面向高维数据的稀疏非参核学习方法研究”(61403394); 国家自然科学基金面上项目“多目标低秩非参核学习模型与优化方法研究”(61572505)。

作者简介: 陈 斌 (1993—), 男, 硕士研究生, 主研方向为机器学习、数据挖掘; 周 勇 (通信作者), 教授、博士生导师; 刘 兵, 副教授。

收稿日期: 2017-12-21 **修回日期:** 2018-01-23 **E-mail:** chenb@cumt.edu.cn

向长短期记忆(Conv-BiLSTM)网络的事件触发词抽取模型。使用卷积层来抽取单词的词级别特征。采用长短期记忆网络提取句子级别特征。将词级别特征与句子级别特征相结合,通过分类器进行触发词抽取。

1 相关工作

事件抽取任务是自然语言处理领域的重要课题之一,相关的研究工作已经展开。目前事件抽取的研究大致可以分为3类:基于特征的方法,基于结构的方法和基于神经网络的方法^[4]。

传统事件抽取方法通常是基于特征或基于结构的方法,文献[5]使用一组传统的特征表示(如词性)来抽取事件。为了从文本中获取更多的线索,文献[6-7]分别提出交叉事件和跨实体推理的事件抽取方法。为了更好地表示特征对事件的影响,文献[8]提出一个特征加权方法对事件的各项特征进行权重重新分配。在基于结构的方法中,文献[1]设计一种基于依存句法分析与分类器融合的触发词抽取方法。文献[9]提出一种基于结构预测的事件抽取联合学习框架,并取得了较好的结果。然而,这些方法的抽取性能依赖于自然语言处理(Natural Language Processing, NLP)工具是否能够提取高质量的特征,且事件结构是建立在语义而非语法的基础上,这导致了事件抽取系统性能无法进一步提升。

近年来,基于神经网络的方法被应用到事件抽取中。文献[4]通过对池化层的改进提出一种动态多池化卷积神经网络模型进行事件抽取,并取得了较好的效果。文献[10]实现了一种双向循环神经网络(Recurrent Neural Network, RNN)的联合框架,通过提取有效特征进行事件抽取。在基于神经网络的方法中,研究者们通常采用从文本中能够直接获取的特征(如词向量、位置信息等),而不使用复杂的自然语言处理工具进行特征提取,这样有效避免了自然语言处理工具的错误累积对事件抽取的影响。

2 卷积双向长短期记忆网络模型及算法

在触发词抽取过程中,通常会面临2个主要问题:1)相同的词汇在不同的语境中表示不同的事件。2)一个事件句中包含多个事件。

S_1 : She lost her seat in the 1997 election.

S_2 : Hundreds of years, a lot of people have lost their lives.

S_3 : In Baghdad, a cameraman died when an American tank fired on the Palestine hotel.

在句子 S_1 和 S_2 中,“lost”作为触发词分别代表选 Personnel_Elect 事件和 Life_Die 事件。相同的词汇在不同的语境中表现出不同的含义。而在 S_3 中,“died”和“fired”分别表示 Life_Die 事件和 Conflict_

Attack 事件。同一个事件句中包含2个不同类型的触发词。

为了解决上述问题,当前方法通常依赖于提取的特征,这些特征大致可分为词级别特征和句子级别特征^[11]。词级别特征主要包含词汇的语义信息及其上下文语境信息。句子级别特征主要指整个句子包含的隐藏信息。本文模型分别采用卷积操作提取词级别特征和长短期记忆网络提取句子级别特征,从而避免了传统特征提取存在的数据稀疏、无法表示词语间隐含关系问题^[12]。

2.1 文本向量化

文本向量化是指将文本数据表示为计算机能够理解的方式。一个句子是由多个单词组成的,其公式可以表示为 $L = \{w_1, w_2, \dots, w_n\}$, 其中, n 是句子长度, w_i 是句子中的第 i 个单词。本文将每个单词 w_i 转换成实值向量 x_i , 其中 x_i 由以下2部分构成:

1) 词向量

单词 w_i 的词向量表示为 $e(w_i)$ 。词向量具有良好的语义特性,是表示词语特征的常用方式,词向量的每一维的值代表一个具有一定语义和语法解释的特征。本文采用 Word2vec^[13] 对大量文本数据进行训练获得词向量。Word2vec 是 Google 公司在2013年开放的一款用于训练词向量的软件工具,它根据给定的语料库,通过优化后的训练模型快速有效地将一个词语表达成向量形式。本文使用大量无标签语料训练词向量,训练方式采用 Skip-gram 模型。

2) 位置特征

位置特征是指当前单词在句子中的相对位置。本文采用一个5维的向量 $e(pe_i)$ 来表示位置特征。与词向量相似,位置向量通过随机初始化和反向传播优化而成。

因此,单词 w_i 可以采用如下公式表示:

$$x_i = e(w_i) \oplus e(pe_i) \quad (1)$$

其中, $x_i \in \mathbb{R}^d$, \oplus 表示将2个向量通过拼接的方式合并在一起。

一个长度为 n 的句子可以用如下方式表示:

$$X = \{x_1, x_2, \dots, x_i, \dots, x_n\} \quad (2)$$

其中, $X \in \mathbb{R}^{n \times d}$, x_i 是第 i 个单词 w_i 的特征向量。

由于卷积神经网络通常需要固定长度的输入向量,因此本文设置句子最大长度为 $maxlen$ 。在训练集和测试集中,小于最大长度的句子采用补零的方式,大于最大长度的句子截取前 $maxlen$ 个单词作为该句子。

2.2 卷积双向长短期记忆网络模型

卷积双向长短期记忆网络模型主要由3个部分构成:1)文本的向量表示部分;2)特征提取部分;3)输出层。文本的向量表示部分如第2.1节文本向量化所述,单词的特征表示由词向量和位置特征构成。特征提取部分由构成,分别是基于卷积的词级

别特征提取和基于长短期记忆网络的句子级别特征提取。输出层是将特征提取部分提取出的事件相关特征通过分类器识别候选触发词所属类别。

2.2.1 基于卷积的词级别特征提取

在事件触发词抽取任务中,卷积操作采用一个卷积核在序列上滑动以获取特征图(feature map)。卷积层的主要目的是获取单词的语义及其上下文特征,最后以 feature map 的形式表现出来。

卷积操作提取词级别特征的过程如图 1 所示(图中不同灰度代表不同单词提取的特征),给定句子的向量表示为 $X = \{x_1, x_2, \dots, x_i, \dots, x_n\}$, 其中 $x_i \in \mathbb{R}^d$ 。

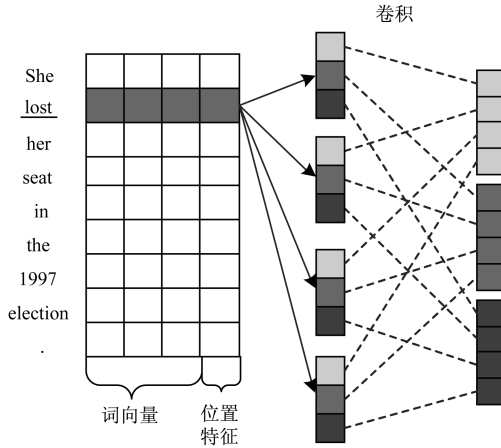


图 1 基于卷积的词级别特征提取过程

为获取一个单词语义及其周围单词的相关上下文特征,设计一个矩阵 $u \in \mathbb{R}^{k \times d}$ 作为卷积操作的卷积核。因此,单词 w_i 通过卷积操作获取的 feature map 可用如下公式表示:

$$x_i = f(u \cdot x_{i-k/2:i+k/2} + b) \quad (3)$$

其中, $x_{i-k/2:i+k/2}$ 表示单词 w_i 周围 k 个单词特征, $b \in \mathbb{R}^d$ 是一个偏置, $f(\cdot)$ 是一个非线性激活函数。如果单词 w_i 左边或右边没有其他单词(如句子的第 1 个单词或最后 1 个单词),则采用卷积神经网络中常用的补零方式将其补齐。将卷积核扫描句子中每一个单词,获取整个句子的 feature map:

$$g = [c_1, c_2, \dots, c_n] \quad (4)$$

其中, $g \in \mathbb{R}^n$, n 表示句子长度。

通常,在卷积神经网络中会采用多个不同卷积核来获取不同的特征。假定本文使用 m_1 个卷积核,那么整个句子的 feature map 可以表示为:

$$C = [g_1, g_2, \dots, g_{m_1}] \quad (5)$$

其中, $C \in \mathbb{R}^{n \times m_1}$, C 中的每一行代表一个单词经过卷积操作提取到的词级别特征。

在卷积神经网络中,通常在卷积之后会采用最大池化降低输出结果的维度,保留显著的特征。但在序列处理任务上,池化操作会丢失序列中单词的顺序信息,打破序列的组织,因此本文在模型构建中

舍弃最大池化操作,保留卷积提取的原始词级别特征。

2.2.2 基于长短期记忆网络的句子级别特征提取

长短期记忆网络是一种特殊结构的循环神经网络。LSTM 由 HOCHREITER 等^[14]于 1997 年提出,并由 GRAVES^[15]于 2012 年进行了改进和推广。LSTM 通过输入门、遗忘门和输出门控制记录序列中的长期信息,解决了 RNN 存在的长期依赖问题。在自然语言处理中,一个语句可以看作是一个由不同单词组成的序列,因此 LSTM 被广泛应用于机器翻译、命名体识别等自然语言处理任务中。

LSTM 单元结构如图 2 所示,LSTM 单元允许信息在没有相互作用的情况下通过。为改变存储单元中的信息,LSTM 单元设置 3 个逻辑门:输入门 i_t ,遗忘门 f_t 和输出门 o_t 。每个门提供一个介于 0 ~ 1 之间的数值,表示各操作传递或遗忘多少信息,并且每个门提供的数值都由输入 x_t 和前一隐藏状态 h_{t-1} 共同确定。

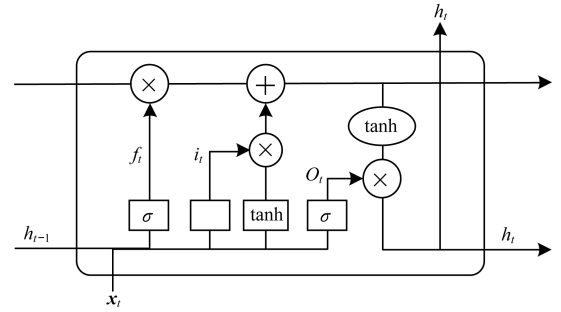


图 2 LSTM 单元结构

在事件触发词抽取这一具体任务中,给定句子的向量表示为 $X = \{x_1, x_2, \dots, x_i, \dots, x_n\}$, 其中 $x_i \in \mathbb{R}^d$ 是单词 w_i 的向量。句子级别特征提取过程如图 3 所示。

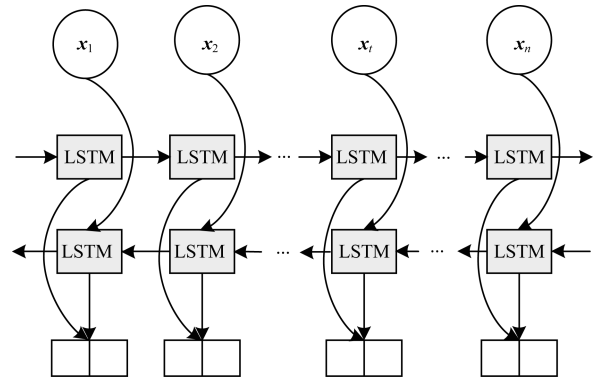


图 3 基于长短期记忆网络的句子级别特征提取过程

本文采用一个双向长短期记忆(BiLSTM)网络进行句子级别特征的提取。一个前向的 LSTM 网络计算 t 时刻单词 w_t 左边部分的状态 \vec{h}_t , 同时采用一个反向 LSTM 网络计算 t 时刻单词 w_t 右边部分的状

态 \overleftarrow{h}_t , 则 $h_t = [\overrightarrow{h}_t; \overleftarrow{h}_t]$ 就是 BiLSTM 在 t 时刻的输出。假定在 BiLSTM 训练过程中选取隐藏层数量为 m_2 , 则 BiLSTM 的输出结果可以表示为:

$$H = [h_1, h_2, \dots, h_n] \quad (6)$$

其中, $H \in \mathbb{R}^{n \times (2 \times m_2)}$, H 中的每一行代表一个单词通过 BiLSTM 提取得到的句子级别特征。

2.2.3 输出层

输出层的主要作用是将本文模型提取得到的特征利用分类器识别出句子中的事件触发词。通过卷积提取得到的词级别特征为 $C \in \mathbb{R}^{n \times m_1}$, 通过长短期记忆网络提取得到的句子级别特征为 $H \in \mathbb{R}^{n \times (2 \times m_2)}$, 将词级别特征和句子级别特征拼接在一起得到:

$$y = C \oplus H \quad (7)$$

其中, $y \in \mathbb{R}^{n \times (m_1 + 2 \times m_2)}$, \oplus 表示行向量拼接。将模型提取的特征放入 softmax 层进行事件触发词预测:

$$O = \text{softmax}(W_2 y + b_2) \quad (8)$$

其中, $W_2 \in \mathbb{R}^{(m_1 + 2 \times m_2) \times n_1}$ 表示连接权重矩阵, $b_2 \in \mathbb{R}^{n_1}$ 表示偏置, n_1 为事件触发词类型数目。

2.3 算法描述

2.3.1 训练阶段

在事件触发词抽取任务中, 将每一个单词都作为候选触发词, 通过模型提取事件相关特征进行识别分类。具体过程如下: 将处理过的训练样本按句划分, 每句截取前 60 个单词, 不足的句子予以补足, 然后采用词向量(300 维)和位置向量(5 维)作为特征向量表示每一个单词。每次训练随机选择 100 个句子, 则输入矩阵是形状为 $100 \times 60 \times 305$ 的三维矩阵。

在训练过程中, 将输入矩阵通过卷积操作和 BiLSTM 网络分别提取词级别特征和句子级别特征, 采用 softmax 函数作为分类器, 使用交叉熵代价(cross-entropy error)函数作为损失函数, 梯度下降方法利用 Adam 优化算法^[16]。Adam 算法根据损失函数对每个参数的梯度一阶矩估计和二阶矩估计动态调整针对于每个参数的学习速率。迭代以上过程, 直到结果收敛, 则认为训练过程完成。

2.3.2 测试阶段

在测试阶段, 数据预处理过程与训练阶段相同。将测试样本放入训练好的模型, 再通过 argmax 函数预测候选触发词的识别分类结果。整个算法流程具体如下:

输入 待识别事件句

输出 事件触发词类型

步骤 1 利用训练样本进行模型训练。

步骤 2 将待识别事件句通过数据预处理得到句子的向量表示(60×305 的特征矩阵 X , 句子长度为 60)。

步骤 3 对特征矩阵 X 通过卷积操作提取词级别特征, 设置卷积核大小 $k = 5$, 卷积核个数 $m_1 = 200$, 得到 60×200 的词级别特征矩阵 C 。

步骤 4 对特征矩阵 X 通过 BiLSTM 网络提取句子级别特征, 设置隐藏层的数量 $m_2 = 100$, 得到 60×200 的句子级别特征矩阵 H 。

步骤 5 将词级别特征矩阵 C 与句子级别特征矩阵 H 拼接, 放入 softmax 函数得到触发词类型预测概率矩阵, 通过 argmax 函数得到触发词类型。

步骤 6 输出事件触发词类型。

3 实验结果与分析

3.1 实验设置

本文实验数据选用事件抽取任务中被广泛使用的 ACE2005 英文语料, 该语料包含 599 篇文档, 并预先定义了 8 大类、33 子类的事件^[2]。与已有的事件抽取研究一样, 在进行触发词抽取任务时, 不考虑大类与子类的层次结构, 对 33 子类进行分类, 把识别子类型作为本文的任务。与文献[4-7, 9-11, 17]的实验设置相同, 本文选取 599 篇文档中固定的 40 篇作为测试集, 30 篇作为校验集以及剩下 529 篇作为训练集。

事件触发词抽取结果的性能评价标准与文献[4-7, 9-11, 17]相同:

1) 触发词识别: 若一个触发词实例的位置和标注文档一致, 则该触发词实例识别正确。

2) 触发词分类: 若一个触发词实例的位置和类型与标注文档一致, 则该触发词实例的类型分配正确。

本文采用通用的评价指标: 准确率, 召回率以及 F 值对事件触发词抽取性能进行评价。

3.2 实验结果对比

实验对本文提出 Conv-BiLSTM 方法与其他事件抽取方法进行性能对比。所对比的方法有 MaxEnt^[5]、Hong's cross-entity^[7]、Li's structure^[9]、DMCNN^[11]、S-CNNs^[17]、JRNN^[4], 对比结果如表 1 所示。

表 1 事件抽取方法性能对比 %

方法	触发词识别			触发词分类		
	准确率	召回率	F 值	准确率	召回率	F 值
MaxEnt	76.2	60.5	67.4	74.5	59.1	65.9
Hong's cross-entity	—	—	—	72.9	64.3	68.3
Li's structure	76.9	65.0	70.4	73.7	62.3	67.5
DMCNN	80.4	67.7	73.5	75.6	63.6	69.1
S-CNNs	78.1	71.8	74.8	74.1	64.8	69.1
JRNN	68.5	65.7	71.9	66.0	73.0	69.3
Conv-BiLSTM	79.8	70.1	74.3	74.7	64.9	69.5

文献[5]提出的 MaxEnt 方法采用最大熵模型进行事件抽取,是最早应用于事件抽取的机器学习算法之一。文献[7]提出的 Hong's cross-entity 方法是采用跨实体推理的方法进行事件触发词的抽取。在文献[9]中,Li's structure 方法是基于结构预测的事件抽取联合学习框架。但以上方法均依赖于额外的 NLP 工具和设计的特征,NLP 工具的误差会给事件触发词抽取的性能带来较大影响。文献[10,11,17]提出的 JRNN、DMCNN、S-CNNs 等模型都是基于神经网络的方法。其中,文献[11,17]是基于卷积神经网络的事件触发词抽取方法,文献[10]基于循环神经网络,这说明 CNN 和 RNN 在事件抽取任务是有效的。从表 1 可以看出,基于神经网络的事件触发词抽取方法的 F 值相比于传统方法普遍高出 1% 以上。这是因为神经网络能够提取出更具代表性的相关事件特征,避免了传统方法使用 NLP 工具提取诸如词性、命名体识别等特征时出现的错误传播。

在触发词分类的结果中,Conv-BiLSTM 获得了最好的结果,相比其他方法中最好的结果提高了约 0.2%。JRNN 通过联合框架的事件抽取方式,极大提高了召回率,但准确率却相对较低,因此没有达到更好的结果。在触发词识别的结果中,S-CNNs 获得了最好的触发词识别结果 74.8%,但在触发词分类结果上却低了 0.4%,这说明 Conv-BiLSTM 结合语境特征后提高了触发词的分类准确率,从而达到了更好的结果。最终的对比结果显示,Conv-BiLSTM 结合了 CNN 和 RNN,有效提取出词级别特征和句子级别特征,具有较好的事件触发词抽取性能。除了触发词抽取性能的提升,Conv-BiLSTM 还表现出更好的收敛特性,BiLSTM 与 Conv-BiLSTM 收敛情况对比如图 4 所示。

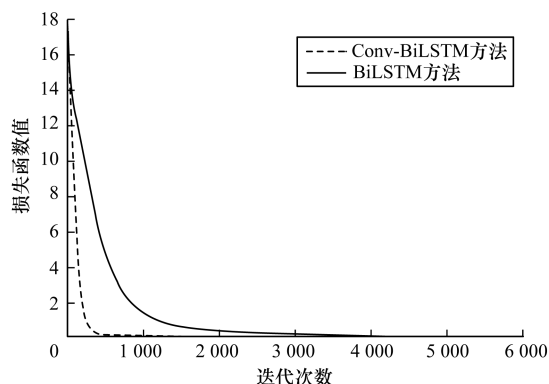


图4 BiLSTM 与 Conv-BiLSTM 收敛情况对比

从图 4 可以看出,BiLSTM 在训练到 4 000 次左右时收敛,而 Conv-BiLSTM 在 1 000 次左右就达到收敛。这是因为 BiLSTM 在训练过程中在一定程度上忽略了单词本身的语义信息,所以需要更多次迭

代才能达到收敛。而卷积操作能够捕获单词及其上下文语境信息,弥补了 BiLSTM 在触发词抽取任务上的不足。

3.3 不同结构的 RNN 触发词抽取结果

为了验证不同结构 RNN 对事件触发词抽取任务的影响,本文选择了 RNN、LSTM、BiLSTM 3 种 RNN 结构进行对比,结果如表 2 所示。

表 2 不同结构的 RNN 触发词抽取结果对比 %

方法	准确率	召回率	F 值
RNN	66.1	57.6	61.6
LSTM	71.6	56.6	63.2
BiLSTM	72.5	59.3	65.2
Conv-BiLSTM	74.7	64.9	69.5

在表 2 中,LSTM 只采用一个前向 LSTM 对事件句触发词进行抽取,BiLSTM 同时采用前向 LSTM 和反向 LSTM 进行事件特征的提取。从结果可以看出,LSTM 在事件触发词分类任务上比传统 RNN 结构要更加有效,这说明 LSTM 能够保留句子远端的信息。而 BiLSTM 相比于 LSTM 提高了 2%,这说明单一方向的句子远端信息对触发词识别的影响较小,触发词前后两端的句子信息能够更好地代表句子级别特征。

本文提出的 Conv-BiLSTM 方法相比于 BiLSTM 准确率、召回率和 F 值都有提高,其中 F 值提高了 4%。这说明结合卷积对于候选触发词周围语境的特征提取起到了很好的作用,在很大程度上提高了事件触发词的抽取性能。

3.4 模糊触发词抽取结果

词级别特征的提取旨在解决事件触发词抽取中出现的一个单词可能在不同上下文情景中表示不同事件的现象。本文定义模糊触发词为:在 ACE 事件抽取任务中,根据自身语义信息及其上下文语境能够表示 2 个或 2 个以上事件的触发词。比如,在句子 S_1 和 S_2 中,“lost”分别表示 Personnel_Elect 事件和 Life_Die 事件,是一个模糊触发词。模糊触发词抽取结果如表 3 所示。

表 3 模糊触发词抽取结果对比 %

方法	准确率	召回率	F 值
BiLSTM	71.1	59.3	64.7
Conv-BiLSTM	73.2	63.1	67.8

从表 3 可以看出,相比于 BiLSTM 方法,Conv-BiLSTM 方法在模糊触发词抽取的准确率、召回率和 F 值都有提高,其中 F 值提高了约 3%。这说明通过卷积操作选取合适的窗口,提取触发词周围的语境特征及其上下文信息对模糊触发词的抽取是有效的,能够在一定程度上解决事件抽取中存在的一词多义现象。

4 结束语

由于传统事件触发词抽取方法过分依赖自然语言处理工具,存在误差累积以及特征提取不足的问题,因此本文提出基于卷积双向长短期记忆网络的事件触发词抽取方法。通过结合卷积操作与双向长短期记忆网络,充分利用词级别特征与句子级别特征,提高触发词的抽取效果。同时简化特征的提取方式,避免使用自然语言处理工具带来的误差累积问题。实验结果表明,该方法在触发词分类阶段的 F 值最终达到 69.5%。下一步将研究文档层面信息对事件的影响,提取出更有效的分类特征,提高触发词的抽取效果。

参考文献

- [1] CHEN C, NG V. Joint modeling for Chinese event extraction with rich linguistic features[C]//Proceedings of COLING'12. Washington D. C., USA: IEEE Press, 2012:529-544.
- [2] DODDINGTON G R, MITCHELL A, PRZYBOCKI M A, et al. The Automatic Content Extraction (ACE) program-tasks, data, and evaluation [EB/OL]. [2017-11-03]. <https://www.researchgate.net/publication/228770793>.
- [3] KIM Y. Convolutional neural networks for sentence classification[EB/OL]. [2017-11-03]. <https://arxiv.org/abs/1510.038202>.
- [4] CHEN Y, LIU S, HE S, et al. Event extraction via bidirectional long short-term memory tensor neural networks [C]//Proceedings of China National Conference on Chinese Computational Linguistics. Berlin, Germany: Springer, 2016:190-203.
- [5] AHN D. The stages of event extraction [C]//Proceedings of the Workshop on Annotating and Reasoning About Time and Events. Cambridge, USA: MIT Press, 2006:1-8.
- [6] LIAO S, GRISHMAN R. Using document level cross-event inference to improve event extraction [C]//Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Cambridge, USA: MIT Press, 2010:789-797.
- [7] HONG Y, ZHANG J, MA B, et al. Using cross-entity inference to improve event extraction[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics; Human Language Technologies. Cambridge, USA: MIT Press, 2011:1127-1136.
- [8] 付剑锋, 刘宗田, 刘炜, 等. 基于特征加权的事件要素识别[J]. 计算机科学, 2010, 37(3):239-241.
- [9] 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. Cambridge, USA: MIT Press, 2013:73-82.
- [10] 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. Cambridge, USA: MIT Press, 2016:300-309.
- [11] 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. Cambridge, USA: MIT Press, 2015:167-176.
- [12] TURIAN J, RATINOV L, BENGIO Y. Word representations: a simple and general method for semi-supervised learning [C]//Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Cambridge, USA: MIT Press, 2010:384-394.
- [13] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [EB/OL]. [2017-11-03]. <http://cn.arxiv.org/abs/1301.3781>.
- [14] HOCHREITE R S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 2012, 9(8):1735-1750.
- [15] GRAVES A. Supervised sequence labelling with recurrent neural networks [M]. Berlin, Germany: Springer, 2012.
- [16] KINGMA D, BA J. Adam: a method for stochastic optimization[EB/OL]. [2017-11-03]. <http://cn.arxiv.org/abs/1412.6980>.
- [17] ZHANG Z, XU W, CHEN Q. Joint event extraction based on skip-window convolutional neural networks [C]//Proceedings of International Conference on Computer Processing of Oriental Languages. Berlin, Germany: Springer, 2016:324-334.

编辑 陆燕菲