

文章编号: 1003-0077(2019)10-0064-09

基于卷积循环神经网络的关系抽取

宋睿, 陈鑫, 洪宇, 张民

(苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

摘要: 关系抽取是信息抽取领域一项十分具有挑战性的任务, 用于将非结构化文本转化为结构化数据。近年来, 卷积神经网络和循环神经网络等深度学习模型, 被广泛应用于关系抽取的任务中, 且取得了不错的效果。卷积神经网络和循环神经网络在该任务上各有优势, 且存在一定的差异性。其中, 卷积神经网络擅长局部特征提取, 循环神经网络能够捕获序列整体信息。针对该现象, 该文综合卷积神经网络抽取局部特征的优势和循环神经网络在时序依赖中的建模能力, 提出了卷积循环神经网络(convolutional recurrent neural network, CRNN)。该模型分为三层: 首先针对关系实例抽取多粒度局部特征, 然后通过聚合层融合不同粒度的特征, 最后利用循环神经网络提取特征序列的整体信息。此外, 该文还探究多种聚合策略对信息融合的增益, 发现注意力机制对多粒度特征的融合能力最为突出。实验结果显示, CRNN 优于主流的卷积神经网络和循环神经网络, 在 SemEval 2010 Task 8 数据集上取得了 86.52% 的 F_1 值。

关键词: 关系抽取; 卷积神经网络; 循环神经网络; 聚合策略; 注意力机制

中图分类号: TP391

文献标识码: A

Combination of Convolutional Recurrent Neural Network for Relation Extraction

SONG Rui, CHEN Xin, HONG Yu, ZHANG Min

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

Abstract: Relation extraction is a challenging task in information extraction, which is used to transform unstructured text into structured data. In recent years, deep learning models such as Convolutional Neural Network and Recurrent Neural Network have been widely used in relation extraction tasks and have achieved good results. To combine the advantages of CNN to extract local features and RNN to model in time series dependence, this paper proposes a convolutional recurrent neural network (CRNN) to extract phrase-level features and multi-granularity phrases for relation instances. The model is divided into three layers. Firstly, multi-granularity local features are extracted for the relation instance, and then the different granularity features are merged through the aggregation layer. Finally, the overall information of the feature sequence is extracted by RNN. In addition, this paper also explores the gains of various aggregation strategies for information fusion, and finds that the attention mechanism is the most prominent for the fusion of different granularity features. The experimental results show that CRNN is superior to state of the art CNN and RNN models with 86.52% of F_1 scores on the SemEval 2010 Task 8 dataset.

Keywords: relation extraction; convolutional neural network; recurrent neural network; aggregation strategy; attention mechanism

0 引言

信息抽取(information extraction, IE)研究如何将非结构化描述的自然文本转化为结构化数据。

实体关系抽取(relation extraction, RE)则是该领域的一项重要子课题,其目的是从自然文本中挖掘给定实体对之间的语义关系。如例 1 所示,句中的实体 relics 和 culture 需被判定为 Entity-Origin 关系类型。主流的关系抽取方法将该任务当作分类问题

收稿日期: 2019-01-28 定稿日期: 2019-05-21

基金项目: 国家自然科学基金(61672367, 61672368); 国家重点研发计划(2017YFB1002104)

处理,通过设计分类模型学习关系实例的特性,从而利用训练完备的分类器预测实体间关系类型。

例 1 *The archaeological site yields relics from long forgotten culture.*

<译>: 考古遗址发掘了被遗忘已久的文化古物。

实体 1 *relics*

实体 2 *culture*

关系类型: *Entity-Origin*

传统的关系抽取主要依靠手工特征^[1-2]或核函数^[3-6]构造分类模型,这类方法不仅难以捕获句子的全部信息,且容易产生误差,从而影响了最终分类器的性能。而深度学习方法不依赖手工构造的特征,因而缓解了传统方法的错误传递问题。但是,深度学习方法无法兼顾局部信息和整体信息。其中,卷积网络主要依靠卷积核抽取局部特征,因此在长句建模中效果不佳,循环网络则依靠序列单元提取序列中的上下文关系,对局部特征关注度不够。因此,本文提出一种兼顾局部与整体信息的卷积循环神经网络(convolutional recurrent neural network, CRNN),旨在更好地表征关系实例。模型主要包括以下三个部分:

① 在卷积层通过多尺度卷积核抽取多粒度局部特征;

② 在聚合层采用三种不同的聚合策略对多粒度特征进行融合;

③ 在循环层通过 Bi-LSTM 建模局部特征间的上下文依赖,获取全局特征。

本文的研究意义在于,提出了一种融合局部与整体信息的网络架构,并以此对关系实例建模,在 SemEval-2010 评测数据集^[7]上取得了优于主流模型的性能。其贡献主要在于:

① 以紧凑的短语级别表示输入 RNN,替代原有松散的词级别输入;

② 利用 RNN 来处理特征序列,从而弥补 CNN 在应对序列数据上的不足;

③ 探究了多种聚合策略对信息融合的效果。

本文组织结构如下:第 1 节概述相关工作,第 2 节介绍模型的框架和内部策略,第 3 节给出实验设置、结果和分析,最后为总结和展望。

1 相关工作

关系抽取是信息抽取领域一个重要研究方向,

通常做法是将关系抽取当作分类问题处理。在识别出关系实例中的实体对的基础上,利用分类器来判定关系实例的类型。传统的关系抽取方法主要包括基于特征的方法和基于核函数的方法。近年来随着深度学习的发展,关系抽取也逐渐从传统方法过渡到深度学习方法,下面将具体介绍这几类方法的发展过程。

基于特征的方法通过着重提取文本中的重要特征来描述实体间的关系,如实体类型、依存树^[1]和词块^[2]特征等,并将其组织成向量形式,利用机器学习算法(如支持向量机、最大熵等)对关系特征进行分类。该方法依赖特征的设计和自然语言处理工具(如命名实体识别、词性标注、短语提取等)的准确率,因此设计耗时且容易造成错误累积,进而影响最终的分类性能。

基于核函数的方法通过设计核函数计算对象在高维空间的相似度,以获取对象的结构化特征,并以此构建分类模型。Zelenko^[3]等通过浅层句法分析设计树核函数;Culotta^[4]等通过依存句法捕获对象间的相似性来构造树核函数;Bunescu^[5]等将句法最短路径与树核函数融合;Zhang^[6]等利用卷积树核探究句法特征的效果。这类方法依赖核函数的设计,且同样受到自然语言处理工具损失的影响,因此通用性不强。

基于神经网络的方法能自动学习文本特征,不需要手动构建特征,因此避免了传统方法带来的损失。神经网络的方法主要包括基于卷积神经网络的方法和基于循环神经网络的方法。其中卷积神经网络的方法主要包括:Zeng^[8]等提出利用卷积神经网络对关系实例建模,并引入了位置向量;Nguyen^[9]等通过设计不同尺度的卷积核对语义关系进行表征;Santos^[10]等提出利用合页损失函数替代原有的交叉熵,以更好地区分不同类型的关系实例;Wang^[11]等则在卷积网络中尝试了多层注意力机制,以着重突出句子成分对关系标签的贡献。循环神经网络的方法主要包括:Zhang^[12]等提出用 RNN 替代 CNN 建模关系实例,并用简单的位置标签替代位置向量;Li^[13]等通过循环神经网络对不同的句法结构进行建模;Miwa^[14]等利用 Tree-LSTM 建模关系实例,并考虑不同的句法结构类型;Zhou^[15]等则将双向循环神经网络和注意力机制应用于关系抽取任务中;Xiao^[16]等将长句切割,并利用两层 RNN 和注意力机制分别编码。

从建模方法的角度而言,目前神经关系抽取方

法主要包括基于卷积神经网络的方法和基于循环神经网络的方法,两类神经网络方法各有优劣。其中,基于卷积神经网络的方法擅长提取关系实例的局部特征,而基于循环神经网络的方法则能够很好地捕获序列的整体特征信息。但是,卷积神经网络过于关注局部特征而忽略整体的序列化信息。相比而言,循环神经网络则侧重考虑序列的整体信息,对局部特征的和表示和应用略显欠缺。针对二者在表示学习层面上互为补充的特点,本文提出一种融合局部和整体信息的卷积循环神经网络,通过聚合策略聚合卷积产生的多粒度局部特征,并通过循环神经网络对聚合特征抽取整体信息。

2 基于 CRNN 的关系抽取

2.1 框架概述

本文的关系抽取模型主要包括三部分:卷积层抽取多粒度局部特征、聚合层融合多粒度特征和循环层对特征上下文依赖建模。模型的具体框架如图 1 所示,其中,输入序列表示由 d_w 维的词向量(word embedding)和 d_p 维的位置向量(position embedding)拼接而成。模型首先将序列表示输入卷积层,利用多卷积窗口抽得多粒度的 n -gram 特征(此处由于卷积操作的平移不变性,使得特征的排列顺序与原始词序列一致,因此不同窗口大

小的卷积特征即相当于对应的 n -gram 特征);然后在聚合层将抽得的特征输入聚合层,将不同粒度的卷积特征进行融合,并将其按时序组织,形成以原词为中心的多粒度融合特征。最后经过循环层,利用 LSTM 捕获序列长期依赖的能力,对序列特征进行建模,针对高阶的 n -gram 表示,提取其中的上下文依赖信息。

2.2 局部和整体信息建模

为了同时考虑关系实例的整体和局部信息,本节主要介绍如何综合利用卷积神经网络和循环神经网络对关系实例建模,以建立局部特征和整体特征间的联系,从而更好地获得关系表征。

2.2.1 局部特征抽取

关系实例的重要信息通常体现在序列的局部区域,但粒度却并不绝对。为了使模型能够捕捉不同粒度的局部特征(local feature),本文采用不同尺度的卷积核捕捉文本中不同粒度的 n -gram 特征,以获取更全面的局部特征信息。

设输入的句子矩阵为 C ,其中每个词的维度为 d (包括 d_w 维的词向量表示和 d_p 维的位置向量)。为了得到输入句子的特征表示,模型初始化窗口大小为 k 的滤波器用于卷积操作,其宽度与词向量维度一致。卷积操作如式(1)所示。

$$c_i^k = \sigma(\sum(C[i:i+k] \cdot H_k) + b) \quad (1)$$

其中,“ \cdot ”代表点乘, σ 表示 sigmoid 激活函数, $C[i:i+k]$ 表示第 i 到 $i+k$ 的词向量序列, H_k 为宽度为 k 的卷积核。为了保证不同的卷积核抽取的特征图长度一致,本文采用了 SAME 的 padding 策略,以使不同粒度特征图和输入序列长度一致。

2.2.2 局部信息融合

为了更好地建模序列的整体信息,本文通过聚合层对多粒度局部特征进行融合,并将聚合后的特征按时序进行组织,从而得到融合多粒度局部信息的特征矩阵 $[c_1, c_2, \dots, c_n]$ 。其中, c_i 为第 i 组特征向量,表示以原序列第 i 个词为中心的多粒度卷积结果的聚合特征。为了探究不同的特征聚合策略,本文将在 2.3 节详细介绍三种聚合策略。

2.2.3 整体上下文建模

为了更好地捕捉特征序列间的整体信息,本文将循环层置于卷积层之上,旨在通过循环神经网络捕获特征序列的上下文关系。在具体技术环节,本文采用双向长短期记忆网络(bidirectional long

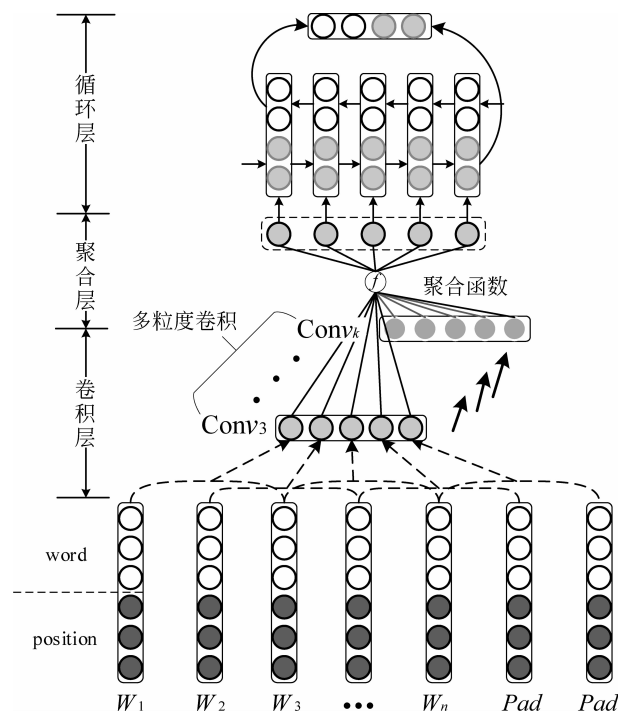


图 1 CRNN 模型框架

short term memory network, BiLSTM) 构建循环层。该网络通过前向、后向两组链式网络结构分别对特征序列进行建模, 并将两组链式网络最后时刻的输出进行拼接作为序列的整体信息。具体来看, BiLSTM 将 LSTM 神经元作为基本的处理单元, 其内部通过多组门控机制实现对整体序列信息的控制。其借助三组数学公式实现信息的处理与计算, 如式(2)~式(4)所示。

$$\begin{bmatrix} o_t \\ g_t \\ i_t \\ f_t \end{bmatrix} = \begin{bmatrix} \sigma \\ \tanh \\ \sigma \\ \sigma \end{bmatrix} \left(\mathbf{W} \begin{bmatrix} c_t \\ h_{t-1} \end{bmatrix} + b \right) \quad (2)$$

$$l_t = f_t \odot l_{t-1} + i_t \odot g_t \quad (3)$$

$$h_t = o_t \odot \tanh(l_t) \quad (4)$$

其中, c_t 代表聚合层在 t 时刻的输出, h_{t-1} 代表上一神经元的输出, σ 表示 sigmoid 激活函数, \mathbf{W} 为权重矩阵, f_t , i_t 和 o_t 分别代表 LSTM 内部的遗忘门、输入门和输出门。

2.3 融合多粒度 n-gram 的聚合策略

卷积层获取的 n -gram 特征, 由于卷积核尺寸的不同, 其抽取的特征也呈现多粒度的分布。为了更好地组织不同粒度的特征, 以适应循环层建模, 本文尝试了三种聚合策略将不同粒度特征融合, 分别是全拼接策略、逐元素最大池化策略以及自注意力机制策略。下面将具体阐述三种聚合策略。

2.3.1 全拼接策略

采用全拼接策略(full concentrate), 对不同粒度的特征直接进行拼接。如图 2(a)所示, 将特征按序列展开, 其中每个时刻的聚合特征由该时刻不同粒度的特征词拼接而得, 结果如式(5)所示。

$$c_i = [c_i^1; c_i^2; \dots; c_i^k] \quad (5)$$

其中, c_i^k 代表第 i 个词的卷积窗口为 k 的特征表示。

2.3.2 逐元素最大池化策略

采用逐元素最大池化(element-wise maxpooling)的策略, 对不同粒度的特征进行筛选。如图 2(b)所示, 对多组 n -gram 特征逐位最大池化, 即逐个取特征位上多组特征的最大值, 形成向量与原始特征位数一致, 具体操作如式(6)~式(7)所示。

$$c_{ij} = \max(c_i^1[j], c_i^2[j], \dots, c_i^k[j]) \quad (6)$$

$$c_i = [c_{i1}; c_{i2}; \dots; c_{in}] \quad (7)$$

其中, $c_i^k[j]$ 表示第 i 个词的 k -gram 卷积特征的第 j 维特征, n 代表 n -gram 特征的宽度, 即卷积核的数目。

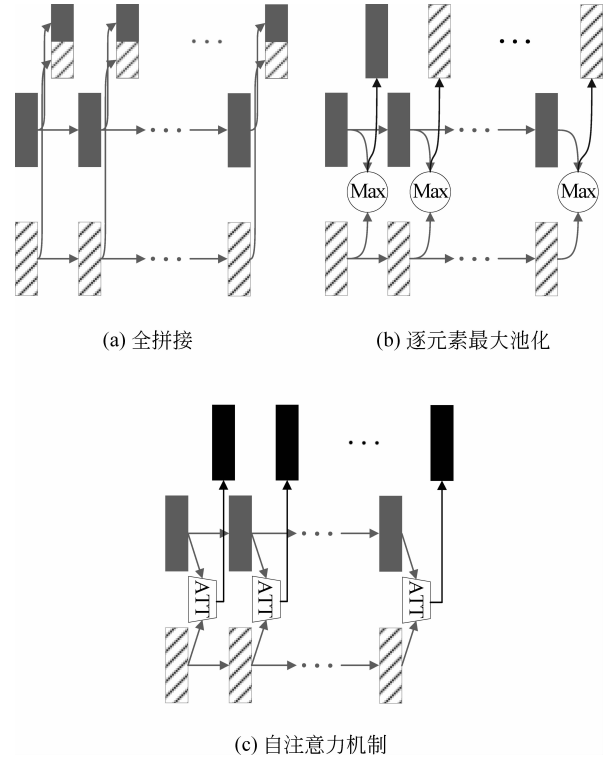


图 2 聚合策略

2.3.3 自注意力机制策略

采用自注意力机制^[17](self-attention), 综合考虑不同粒度特征的信息。如图 2(c)所示, 通过计算不同粒度 n -gram 特征之间的相关性, 将不同粒度特征相互关联, 从而赋予不同粒度特征相应的权重, 通过加权求和获取其表征。具体计算如式(8)~式(10)所示。

$$U_i^t = \tanh(\mathbf{W}_w c_i^t + b_w) \quad (8)$$

$$\alpha_i^t = \frac{\exp(U_i^{tT} u_w)}{\sum_t \exp(U_i^{tT} u_w)} \quad (9)$$

$$c_i = \sum_1^k \alpha_i^t c_i^t \quad (10)$$

其中, c_i^t 表示第 t 种特征的第 i 个词表征。

本文将以上三种聚合策略都作用于多粒度的 n -gram 特征, 并且将聚合的结果 $[c_1, c_2, \dots, c_n]$ 按时间步 t 输入循环层, 以建模多粒度短语之间的上下文依赖。

2.4 模型训练

本节主要介绍模型的训练过程。在循环层获取关系表征之后, 模型设置 Softmax 作为分类器, 用于接收隐层的输出, 并据此生成每个关系类别的概率分布, 然后最小化这一概率分布与关系实例的实际

类别的交叉熵来训练模型。

具体来讲,在给定的训练实例 X^i 及其标签 y^i 上,模型预估每个类别下的概率 $\tilde{y}_j^i \in [0,1]$,其中 $y^i \in \{1,2,\dots,k\}$, k 为关系类别的数量。通过 Adam 优化器来最小化分类结果和真实类别之间的交叉熵损失进行参数的学习。损失函数的公式如式(11)所示。

$$L(X^i, y^i) = \sum_j^k l(y^i = j) \log(\tilde{y}_j^i) \quad (11)$$

其中, l 为指示函数,当 $y^i = j$ 为真时 $l=1$,否则为 0。

在学习参数的过程中,深度学习模型往往会由于参数过多而导致过拟合现象,为了防止过拟合现象的发生,本文尝试两种不同策略来解决这一问题:

① Dropout 策略是通过在训练过程中忽略某些节点,即不让数据经过部分神经元。这种随机丢弃的策略相当于组合了多个网络模型,且模型参数数量都较原先模型少,因此有更好的泛化能力。本节在模型的输入层和 RNN 之后都添加了 dropout 策略。

② 正则化则是给模型的代价函数添加罚项,以降低模型本身的复杂度。本文使用 L_2 正则化项来约束模型参数,如式(12)所示。

$$L = L_0 + \frac{\lambda}{2n} \sum_w w^2 \quad (12)$$

其中, L 为总代价函数, L_0 代表原始的代价函数, w 代表所有的模型参数, n 为训练集大小, λ 则为正则项系数,用于权衡正则项和原始代价函数的比重。

3 实验与分析

为了评估提出的 CRNN 模型在关系抽取任务上的有效性,本文在关系抽取的基准数据集上进行了实验,旨在验证:

① 融合局部和整体信息的 CRNN 模型在关系抽取任务上是否较主流模型有明显优势;

② 不同的聚合策略对模型会产生哪些影响。

下面,首先在 3.1 节介绍实验数据集及其划分,然后在 3.2 节介绍具体评价指标及参数设置,最后详细阐述模型的实验性能以及与其他方法的对比。

3.1 实验数据及划分

对于关系抽取的实验,本文采用语义评测会议 SemEval 的关系抽取数据集 SemEval-2010 Task 8^[7],该语料集包含 8 000 个训练样本和 2 717 个测试样本,其中共 9 大类关系和一个其他类型,关系类型具体如表 1 所示。

表 1 SemEval-2010 Task 8 数据集关系类型

关系类型	描述
Cause-Effect(C-E)	因果关系
Component-Whole(C-W)	整体与部分
Entity-Destination(E-D)	实体和目的
Entity-Origin(E-O)	实体和来源
Product-Producer(P-P)	产品和生产者
Member-Collection(M-C)	成员和组织
Message-Topic(M-T)	信息和主题
Content-Container(C-C)	内容和包含
Instrument-Agency(I-A)	工具和机构
Other(O)	其他

此外,依据测试样本的分布,本文从训练集中抽取了 1 600 个样例充当开发集,用于调整参数及选择模型,其各个类别的语料分布如图 3 所示。

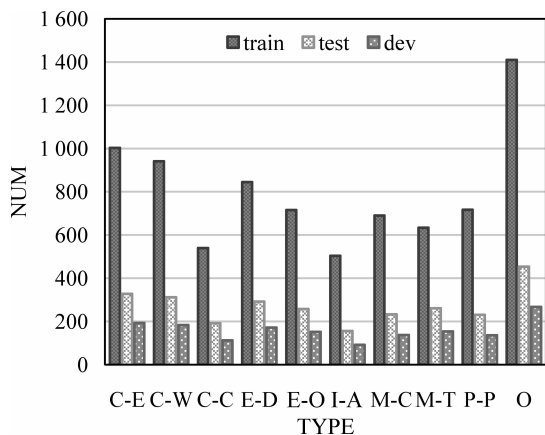


图 3 实验语料集分布

3.2 实验设置及评价指标

本文实验的词向量采用的是预训练的 Glove 词向量^①,维度为 300 维。模型超参设置上,本文尝试了多组模型参数,并通过网格搜索在开发集上选择

① <https://nlp.stanford.edu/projects/glove/>

了最优的模型。其中,最优结果在 20~25 个迭代轮获取。模型具体参数实验范围和最终设置如表 2 所示。

表 2 模型超参设置

参数	最终值	实验范围
Position embed size	20	10~50
CNN filter size	3, 4	2~4 to 5~9
CNN filter num	75	60~150
GRU unit num	120	60~150
LSTM unit num	100	60~150
learning rate	0.01	
RNN layer num	1	1~2

本文采用准确率 P (precision)、召回率 R (recall)和 F_1 值作为衡量模型性能的评价指标,首先通过混淆矩阵求得每个类别下的 P 、 R 、 F_1 值,然后通过宏平均(Macro-average)求得所有类别的算术平均值作为整体性能的指标。

3.3 实验结果及分析

本文的实验以 MultiCNN^[9]为基线系统(Baseline),本文复现的该模型在未引入额外知识的前提下,分类性能能够达到 82.63%。

3.3.1 局部与整体建模

在第一组实验中,本文在 MultiCNN 获取的多粒度卷积特征基础上,通过添加循环层捕获序列之间的上下文信息,替代了原先的池化层。得到的实验结果如表 3 所示,其中聚合层采用直接拼接的策略。

表 3 添加循环层的实验结果
(实验结果省略%, F_1 值为宏平均)

方法	F_1 值
Baseline(MultiCNN)	82.63
+ BiLSTM	86.47

从实验结果可以看出,在 MultiCNN 的基础上添加循环层,提高了模型的整体分类效果。其原因在于,单独的 CNN 只能关注序列的局部区域,通过循环层能更好地捕捉特征间的联系,从而获取关系实例的整体特征。此外,以 LSTM 作为基本的处理单元,其遗忘门和记忆门在一定程度上起到了特征

选择的作用,效果比简单的池化更优。该实验证明,通过 RNN 能有效弥补 CNN 在整体信息建模上的不足,从而提升关系抽取的性能。

3.3.2 聚合策略

在第二组实验中,本文旨在验证三种不同的聚合策略对信息融合的影响。在上节模型的基础上,分别尝试逐元素最大池化和自注意力机制的聚合策略,实验结果图 4 所示。

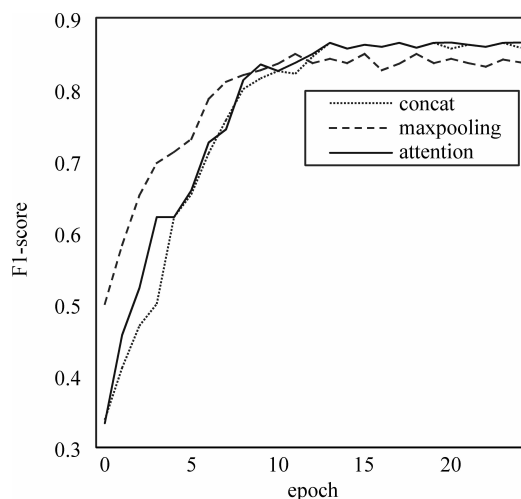


图 4 聚合策略实验结果

从实验结果可以看出,自注意力机制和直接拼接的聚合策略结果近似,且都优于逐元素最大池化策略。其原因在于:

- ① 模型选取 LSTM Cell 作为基本处理单元,其中,遗忘门和记忆门在一定程度上起到了特征选择的作用,因此不需要再利用池化操作对特征进行筛选;
- ② 自注意力学习到的不同粒度的权重信息,直接拼接也能够获取,不过是将权重学习的负担传递到循环层。

综合以上两组实验结果,本文在基线系统的基础上,组合利用自注意力机制的聚合策略和基于 BiLSTM 的整体建模思路,在关系抽取语料集上进行实验,并详细比较了基线系统和本文的模型在各个类别下的指标,实验结果如表 4 所示。从表中可以看出,相比基线系统的多通道 CNN,本文的模型在各个类别均有性能优势。其中,召回率提升明显,说明原先被误判为 other 类别的样本都能被本文的模型正确识别。并且,观测发现,本文的模型在长句预测上的性能要显著优于基线系统。如例 2 中关系实例,基线系统判定为 other 类别,而本文的模型则正确标识为 Member-Collection 类。

表 4 最优模型性能结果
(实验结果省略%, F_1 值为宏平均)

类型	Baseline			Ours		
	P	R	F_1	P	R	F_1
C-E	86.80	94.21	90.35	94.79	94.21	94.5
C-W	79.03	83.33	81.12	82.76	84.62	83.68
C-C	79.26	89.58	84.11	79.55	91.15	84.95
E-D	84.54	91.78	88.01	88.06	93.49	90.70
E-O	82.95	84.88	83.91	86.13	91.47	88.72
I-A	70.19	72.44	71.29	76.54	79.49	77.99
M-C	81.97	85.84	83.86	78.81	90.99	84.46
M-T	80.49	88.51	84.31	84.41	95.40	89.57
P-P	76.39	77.06	76.72	82.23	86.15	84.14
ALL	80.18	85.29	82.63	83.69	89.66	86.52

例 2 *Skype, a free software, allows a hookup of multiple computer users to join in an online conference call without incurring any telephone costs.*

<译>: Skype 是一款免费软件,允许多个计算机用户连接加入在线电话会议,且不会产生任何电话费用。

实体 1 *hookup*

实体 2 *users*

关系 *Member-Collection*

3.3.3 与现有方法对比

最后本文选取了当前主流的关系抽取模型与本文提出的 CRNN 模型进行对比,分别是:

① SVM: 融合多种词法、句法、框架语义等特征;

② Multi-CNN: 利用多尺度卷积核抽取特征+最大池化;

③ CR-CNN: 模型为基本的 CNN,损失函数利用合叶损失函数替代交叉熵损失;

④ SDP-LSTM: 利用 LSTM 建模实体间的最短依存路径;

⑤ BiLSTM+ATT: 融合双向循环神经网络和注意力机制;

⑥ BiLSTM+CNN: 组合 BiLSTM 和 CNN,其中 BiLSTM 在前,CNN 在后。

如表 5 所示,本文将主流的关系抽取模型分三大类列出: 基于特征的方法、基于 CNN 的方法和基于 RNN 的方法。

表 5 CRNN 和其他模型进行对比
(实验结果省略%, F_1 值为宏平均)

类型	模型	F_1 值
Feature-based	SVM	82.2
CNN	Multi-CNN	82.8
	CR-CNN	84.1
RNN	SDP-LSTM	83.7
	BiLSTM+ATT	84.0
RCNN	BiLSTM+CNN	79.5
CRNN	CNN+ATT+LSTM	86.5

从表 5 中可以看出,相较于传统基于特征的方法以及基于 CNN 和基于 RNN 的模型,本文提出的 CRNN 模型在不依赖外部特征的情况下,取得了优于三类模型的效果。此外,本文还对比了 BiLSTM 和 CNN 的另一种结合方法 RCNN^[18],该方法首先通过 BiLSTM 对每个词进行编码,然后通过卷积层抽取局部特征。实验显示,本文的 CRNN 模型要明显优于 RCNN 模型。原因在于本文提出的 CRNN 模型通过聚合策略对特征进行融合,该方法在保留词序信息的基础上,起到了融合多粒度局部信息的效果,在结合循环神经网络和卷积神经网络中起到了重要作用。

4 总结与展望

本文针对现有的基于神经网络的关系抽取模型

无法融合局部和整体信息的问题,提出了一种融合局部特征和整体序列信息的卷积循环神经网络。该模型通过卷积层抽取多粒度局部特征、聚合层融合多粒度信息、循环层捕获特征之间的上下文联系,从而将关系实例的局部特征和整体信息进行了融合,并在聚合层探究了多种特征融合策略,最终得到了融合局部和整体信息的关系表征。在 SemEval-2010 Task8 数据上的实验证明:

① 融合了局部和整体信息的模型架构比单独建模局部特征和序列关系的模型在关系抽取上更有效;

② 自注意力和直接拼接聚合策略在特征融合上效果类似,并且都优于逐元素最大池化策略。

后续工作将在本文的基础上,继续探究如何更好地利用实体词和关系之间的联系,以凸显重要的局部特征,如采用注意力机制捕获不同特征的重要性。

参考文献

- [1] Guodong Z, Jian S, Jie Z, et al. Exploring various knowledge in relation extraction[C]//Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, 2005: 427-434.
- [2] Kambhatla N. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations[C]//Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions. Association for Computational Linguistics, 2004: 22.
- [3] Zelenko D, Aone C, Richardella A. Kernel methods for relation extraction[J]. Journal of Machine Learning Research, 2003, 3(2): 1083-1106.
- [4] Culotta A, Sorensen J. Dependency tree kernels for relation extraction[C]//Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2004: 423.
- [5] Bunescu R C, Mooney R J. A shortest path dependency kernel for relation extraction[C]//Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2005: 724-731.
- [6] Zhang M, Zhang J, Su J. Exploring syntactic features for relation extraction using a convolution tree kernel [C]//Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, 2006: 288-295.
- [7] Hendrickx I, Kim S N, Kozareva Z, et al. Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals[C]//Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions, 2009: 94-99.
- [8] Zeng D, Liu K, Lai S, et al. Relation classification via convolutional deep neural network[C]//Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics; Technical Papers, 2014: 2335-2344.
- [9] Nguyen T H, Grishman R. Relation extraction: Perspective from convolutional neural networks[C]//Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing, 2015: 39-48.
- [10] Santos C N, Xiang B, Zhou B. Classifying relations by ranking with convolutional neural networks[J]. arXiv preprint arXiv: 1504.06580, 2015.
- [11] Linlin Wang, Zhu Cao, Gerard de Melo, et al. Relation classification via multi-level attention CNNs [C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016: 1298-1307.
- [12] Zhang D, Wang D. Relation classification via recurrent neural network[J]. arXiv preprint arXiv: 1508.01006, 2015.
- [13] Li J, Luong M T, Jurafsky D, et al. When are tree structures necessary for deep learning of representations[J]. arXiv preprint arXiv: 1503.00185, 2015.
- [14] Miwa M, Bansal M. End-to-end relation extraction using lstms on sequences and tree structures[J]. arXiv preprint arXiv: 1601.00770, 2016.
- [15] Zhou P, Shi W, Tian J, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016, 2: 207-212.
- [16] Xiao M, Liu C. Semantic relation classification via hierarchical recurrent neural network with attention [C]//Proceedings of COLING, 2016: 1254-1263.
- [17] 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.
- [18] Feng X, Qin B, Liu T. A language-independent neural network for event detection[J]. Science China(Information Sciences), 2018, 61(09): 81-92.



宋睿(1993—), 硕士研究生, 主要研究领域为关系抽取、文本匹配。

E-mail: cnsr27@gmail.com



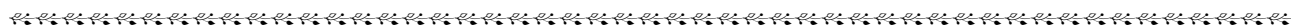
陈鑫(1993—), 硕士研究生, 主要研究领域为自动问答、文本匹配。

E-mail: chenxinkuaile@gmail.com



洪宇(1978—), 通信作者, 教授, 研究生导师, 主要研究领域为话题检测、信息抽取、信息检索。

E-mail: tianxianer@gmail.com



(上接第 56 页)



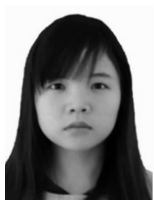
李卫疆(1969—), 博士, 副教授, 硕士生导师, 主要研究领域为信息检索、自然语言处理。

E-mail: hrbrichard@126.com



李涛(1993—), 硕士研究生, 主要研究领域为自然语言处理、实体关系抽取。

E-mail: 754563116@qq.com



漆芳(1994—), 硕士研究生, 主要研究领域为自然语言处理、情感分析。

E-mail: 363869360@qq.com