

基于深度神经网络和注意力机制的实体关系抽取方法研究

陈泽峰, 赵占芳

河北地质大学信息工程学院, 河北 石家庄

收稿日期: 2022年9月24日; 录用日期: 2022年10月20日; 发布日期: 2022年10月27日

摘要

关系抽取是自动构建知识图谱的关键技术之一, 其根本目标是抽取实体间的语义关联关系。针对非结构化文本实体关系抽取方法中存在的上下文环境信息难以准确表征、句子间的实体关系特征未被充分利用等问题, 本文提出了一种新的基于神经网络和注意力机制相结合的关系抽取模型PCNN-ATT-BiLSTM。该模型采用“RNN + CNN”网络框架, 其中的RNN利用双向长短期记忆神经网络(Bi-LSTM)来捕获文本语句的上下文信息和浅层语义特征, 利用分段卷积神经网络(PCNN)捕获文本语句的局部短语特征, 并结合注意力机制捕获文本语句的关键信息进行关系预测。该模型在公开数据集SemEval-2010 Task8上取得了82.92%的F1值, 实验结果表明, 该方法在非结构化文本的实体关系抽取方面表现出了较好的性能, 为实体关系的自动获取提供了新的方法支持。

关键词

关系抽取, 知识图谱, 双向长短期记忆神经网络, 分段卷积神经网络, 注意力机制

Entity Relation Extraction Based on Deep Neural Network and Attention Mechanism

Zefeng Chen, Zhanfang Zhao

College of Information and Engineering, Hebei GEO University, Shijiazhuang Hebei

Received: Sep. 24th, 2022; accepted: Oct. 20th, 2022; published: Oct. 27th, 2022

Abstract

The relation extraction is one of the key technologies to automatically construct knowledge graph, and its fundamental goal is to extract semantic relation between entities. The current methods for

extracting entity relation from unstructured text are difficult to accurately describe the context information, and the entity relation between sentences is not fully utilized. This paper proposes a new relation extraction model PCNN-ATT-BiLSTM based on the combination of neural network structure and attention mechanism. The model adopts the "RNN + CNN" framework, in which RNN uses the Bi-directional Long Short-Term Memory network (Bi-LSTM) to capture the context information and shallow semantic features of text data statements, uses the Piecewise Convolutional Neural Network (PCNN) to capture the local phrase features of text statements, and combines the attention mechanism to capture the key information of text statements, so as to better capture the information in sentences for relationship prediction. The model achieves 82.92% F1 value on the public dataset SemEval-2010 Task8. The experiment shows that the method shows good performance and provides a new method support for the automatic acquisition of entity relation.

Keywords

Relation Extraction, Knowledge Graph, Bi-Directional Long Short-Term Memory Network, Piecewise Convolutional Neural Network, Attention Mechanism

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在大数据时代背景下,随着互联网技术的迅猛发展,网络数据信息呈现爆炸式增长的态势,从海量文本中自动提取有价值信息是自然语言研究领域的研究热点。信息抽取(Information Extraction)是自然语言处理(Natural Language Processing, NLP)领域一个重要的研究方向,关系抽取(Relation Extracion, RE)是其中的重要任务之一。关系抽取是为了识别出文本实体中的目标关系[1],实体关系抽取解决了原始文本中目标实体之间的关系分类问题,有效的关系抽取技术可以提升构建知识图谱的质量和效率。经过几十年的发展,实体关系提取的理论和技術,从早期的人工设计和规则提取模型到后期的基于机器学习和深度学习的模型,已经得到长足发展。随着各类模型的准确性和召回率的不断提高,模型提取关系的适应性比以前更强。

关系抽取是自动构建大规模知识图谱的关键。关系抽取就是识别出实体和它们之间的关系并提取出来,并将其加入到知识图谱中,最终构造出知识图谱[2]。目前,在实体关系抽取任务中主流的方法是有监督学习的关系抽取,特别是随着深度神经网络学习模型的深入应用,关系抽取效果显著提升。对于有监督的实体关系抽取方法[3],国内外目前主要研究集中在如何利用卷积神经网络[4] (Convolutional Neural Networks, CNN)、循环卷积神经网络[5] (Recurrent Neural Network, RNN)和长短时记忆网络[6] (Long Short-Term Memory, LSTM)等进行实体关系抽取。

CNN 在实体关系抽取任务中的应用很广泛。最初仅在图像的分类和识别[7]中用过 CNN,后来慢慢应用于文本分类相关任务中。2014年,Zeng [8]等人提出将 CNN 方法应用于实体关系抽取任务,以预训练的词向量为基础,通过 CNN 网络融合位置信息提取句子级别特征,将句子特征和语法特征结合进行关系分类,采用单个最大池化操作来确定语句中的最重要的特征,但是单个最大池化操作捕获不到两个实体之间的结构信息。因此,2015年,Zeng [9]等人在2014年的基础上提出了用于关系抽取任务的分段卷积神经网络模型,通过使用多实例学习范式,利用该神经网络模型构建了一个基于远程监控数据的关系提取器。采用卷积神经网络模型处理时,没有将文本距离较大的信息等获取到。因此,RNN 及其改进模

型(如 LSTM)被研究者们尝试用于关系抽取任务中。Socher [10]等提出的关系抽取方法使用到了递归神经网络(RNN), 该方法利用了句子的句法结构信息, 但是忽略了实体对的位置信息和语义信息。Zhang [11]等人提出通过双向 RNN 对远距离关系进行建模, 在数据集上进行实验验证, 缺点是在具体的迭代过程中不能避免梯度消失的问题, 能利用的信息有很大的局限性。近两年, 研究者开始考虑结合使用卷积神经网络和循环神经网络来进行实体关系抽取。Zhang [12]等人将卷积神经网络和双向长短时记忆网络融合在一起提取语义关系, 并利用两个神经网络的特点进一步提取信息的特征。随着注意力(Attention)机制的兴起, 一些研究者将注意力模型加入到实体关系抽取任务中。Zhou [13]等人在基于双向 LSTM 网络的实体关系抽取模型中加入注意力机制, 对经过 LSTM 层学习到的特征表示赋予权重值, 加权得到最终的句子表示向量, 但是该方法没有为多实例学习中的实例分配有偏的权重, 同时仅利用了词向量信息。

基于此, 本文提出了一种基于注意力机制的双向长短时记忆网络(Bi-LSTM)与分段卷积神经网络(Piecewise CNN, PCNN)相结合的方法。该模型采用“RNN + CNN”网络框架, 其中的 RNN 利用双向长短时记忆网络(Bi-LSTM)来捕获文本数据语句的上下文信息和浅层语义特征, 利用 PCNN 捕获文本语句的局部短语特征, 并结合注意力机制捕获文本语句的关键信息来进行关系预测。

2. 模型

本文采用卷积神经网络与循环神经网络结合的模型进行实体关系抽取任务, 模型结构如图 1 所示。首先对输入的文本句子进行向量化表示, 将得到的向量送入 PCNN-ATT-BiLSTM 网络模型中, 最后对于模型的输出进行分类, 得到关系抽取结果。

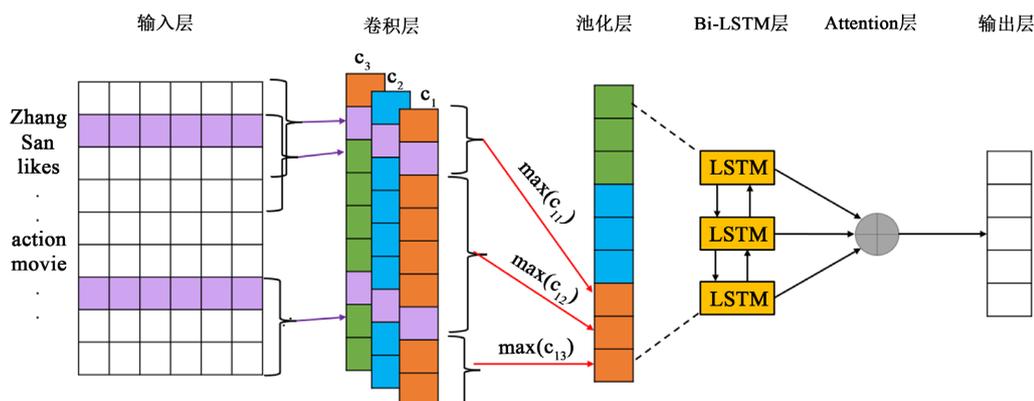


Figure 1. PCNN-ATT-BiLSTM module
图 1. PCNN-ATT-BiLSTM 模型

2.1. 向量表示

由于神经网络模型不能直接利用原始输入文本信息, 所以在将自然语言文本形式的句子输入神经网络模型之前, 首先必须对句子进行向量表示。本文先将词向量和词的位置向量信息结合, 同时把实体对的上下文信息用向量表示加入到输入层, 实现句子的向量表示。

2.1.1. 词向量表示

词向量表示是关系抽取的基础, 本文采用的是 GloVe 方法[14]。GloVe 是基于全局信息对全局词频统计构建共现矩阵来表示词向量化, 从而可以蕴含更多的词义、语义和语法等信息。如对一个包含 n 个单词的句子输入序列 $X = \{x_1, x_2, \dots, x_n\}$, 使用 GloVe 方法将每一个单词映射为唯一的 d^m 维度的向量表示。

2.1.2. 位置向量表示

本文使用 Zeng [8]等人提出了结合位置特征的想法, 利用实体对周围的单词来丰富信息特征。在向量表示阶段引入位置特征, 与词向量信息进行结合, 例如在句子“Xiaoming want to read books in the school.”中, 词“read”到实体“Xiaoming”和到实体“school”的相对距离分别是 2, -3, 然后用 d^p 维度的向量表示两个相对距离 r_1, r_2 , 并将两者结合得到位置特征向量信息 $P(x_i)=[r_1, r_2]$ 。

2.1.3. 实体对上下文

在实体关系抽取中, 距离能体现关键信息。本文提出将实体对的上下文信息的想法, 通过每个实体左右两侧各两个词汇来表示上下文信息, 然后将获取到的每个实体的两侧的两个词汇进行组合。计算方法如下(1)所示:

$$h = [x_1, x_2, x_3, x_4, y_1, y_2, y_3, y_4] \quad (1)$$

x_1, x_2 表示是第一个实体词左侧的两个词所对应的词向量(维度为 d^s), x_3, x_4 表示的是第一个实体词右侧的两个词所对应的词向量, h (维度是 $8 \times d^s$) 表示的是获取到两个实体的上下文信息。

最后, 合并句子中各个词的词向量信息、位置向量信息和实体对上下文信息输入到网络模型中, $T \in R^{s \times d}$ (s 表示句子长度, d 表示向量化的维度, $d = d^w + 2 \times d^p + 8 \times d^s$ 。

2.2. 分段卷积神经网络(PCNN)

2.2.1. 卷积层

卷积层可以从多个角度将数据中的不同特征提取出来, 在卷积操作中, 权重矩阵 $W \in R^{c \times d}$ (c 表示卷积核的大小, d 是词向量维度)与句子向量表示 $S = (s_1, s_2, \dots, s_n)$ (n 表示句子长度, $s_i \in R^d$) 之间的乘积求和运算。模型中一般会使用多个卷积核, 假设模型中使用了 k 个卷积核, 每个句子向量与每个卷积核的计算都会得到一个特征, 最终得到一个向量 c_k 。如公式(2):

$$c_k = f(W_k s_{i,j} + b_k) \quad (2)$$

其中, $s_{i,j}$ 表示的是 s_i 到 s_j 的特征向量的连接, W_k 和 b_k 分别表示的是第 k 个卷积核和第 k 个标量偏置, f 表示的是 ReLU 非线性激活函数。

通过一系列不同卷积核捕捉到句子中具有不同语义信息的特征, 最后得到卷积层的结果是: $C = \{c_1, c_2, \dots, c_k\}$ 。

2.2.2. 最大池化层

池化层是在卷积层之后, 可以对卷积后的特征进行降维采样。PCNN(Piecewise CNN)是在 CNN 基础上通过修改 pooling 的方式得到的, 在执行完卷积操作之后, 池化层会根据两个实体所在的位置将卷积结果分为三段, 再分别对每一段进行最大池化操作。

$$p_{ij} = \max(c_{ij}), 1 \leq i \leq k, 1 \leq j \leq 3 \quad (3)$$

接着每个卷积核都会得到的是三部分对应的池化向量表示进而整合成一个向量 $p_i = \{p_{i1}, p_{i2}, p_{i3}\}$, 再把所有卷积核得到的新的向量拼接在一起得到最后的结果矩阵。

2.3. Bi-LSTM 层

Bi-LSTM 是一种 RNN 特殊的类型。为了解决 RNN 模型中长期依赖、有用信息的间隔有大有小、长短不一的问题, 在 1997 年提出了被改进的 RNN 模型——LSTM [15] (Long Short-Term Memory 网络)。

LSTM 通过三个门结构(记忆门, 遗忘门, 输出门), 选择性地遗忘部分历史信息, 加入部分当前输入信息, 最终整合到当前状态并产生输出状态。LSTM 结构如图 2 所示:

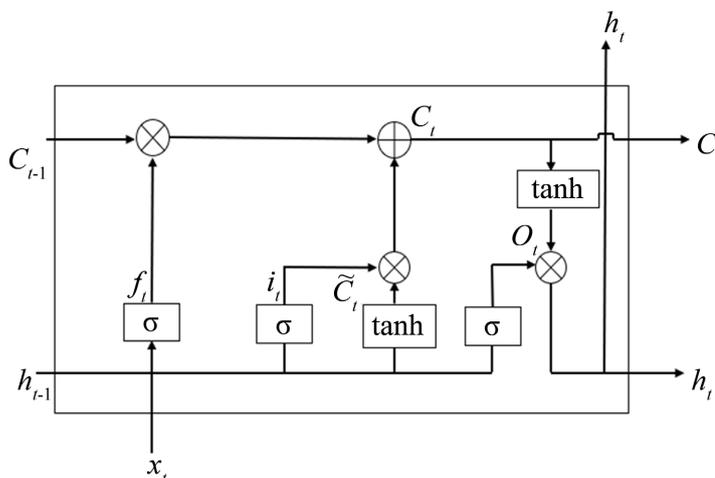


Figure 2. LSTM structure
图 2. LSTM 结构

遗忘门 f_t : 负责控制继续保存长期状态 C_t , 将在使用过程中把上一时刻细胞状态 C_{t-1} 中一些不必要的信息选择出来并“忘记”它们。

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

记忆门: 负责控制把即时状态输入到长期状态 C_t , 决定了新输入的信息 x_t 和上一时刻输出信号 h_{t-1} 中哪些信息将被保留到细胞状态 C_t 中。

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (5)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (6)$$

当前时刻的单元状态 C_t 是由上一次的单元状态 C_{t-1} 按元素和遗忘门 f_t 进行交互, 再用当前输入的单元状态 \tilde{C}_t 按元素与输入门 i_t 进行交互, 再将两个交互结果相加产生的。

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (7)$$

输出门 o_t : 负责控制是否把长期状态 C_t 作为当前的 LSTM 的输出, 将上一时刻的细胞状态 C_{t-1} 和输出信号 h_{t-1} 中的信息与当前输入信息 x_t 整合在一起进行输出 h_t 。

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (8)$$

$$h_t = o_t * \tanh(C_t) \quad (9)$$

LSTM 神经网络模型的门结构可以解决 RNN 无法处理长距离的依赖的问题, 学习到长期依赖信息, 而不需要专门为此付出很大代价。本文采用双向长短期记忆网络(Bi-LSTM)模型学习到更加齐全的上下文信息。将输入序列同时输入到正向 LSTM 和反向 LSTM 中, 最后将前向的隐向量 h_{t1} 和后向的隐向量 h_{t2} 进行拼接得到 Bi-LSTM 隐藏层的输出 h_t :

$$h_t = h_{t1} \oplus h_{t2} \quad (10)$$

最后双向长短期记忆网络的结果 $H = \{h_1, h_2, \dots, h_n\}$, 表示经过 Bi-LSTM 模型后学习到的特征。

2.4. 注意力层

注意力模型最近几年在深度学习各个领域被广泛使用, 无论是图像处理、语音识别还是自然语言处理等各种不同类型的任务中都应用到了注意力机制[16]。为了从学习到的特征挑选出对当前任务更为重要的特征, 丢掉不重要特征信息, 本文引入注意力机制对特征信息进行权重赋值, 筛选出于关系抽取任务影响较大的部分信息。

1) 注意力分布, 计算公式如(11)所示:

$$\alpha = \text{softmax}(w^T H) = \frac{\exp(w^T H)}{\sum \exp(w^T H)} \quad (11)$$

2) 加权平均为, 计算公式如(12)所示:

$$att = H\alpha^T \quad (12)$$

最后得到利用基于 Attention 机制的 Bi-LSTM 抽取文本特征向量的特征组合的结果 $H_1 = \tanh(att)$ 。

2.5. 输出层

该层是神经网络模型的最后一层, 是对神经网络数据的做最后处理得到的结果。将从注意力层得到输出向量 H_1 后, 将其作为输入向量进行关系分类得到最后的分类结果。

本文采用的是带 L2 惩罚项的交叉熵损失函数, 通过自适应的 Adam 优化器对其优化。交叉熵损失函数公式如下:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m t_i \log(y_i) + \lambda \|\theta\|_F^2 \quad (13)$$

m 表示的是关系分类中的关系分类数, λ 表示的是 L2 正则化系数。

采用解决多元分类问题的 Softmax 激活函数。Softmax 函数将多分类的结果以概率的形式展现出来。

$$p(y) = \text{softmax}(W_s H_1 + b_s) \quad (14)$$

$$y' = \frac{\arg \max p(y)}{y} \quad (15)$$

在最后选取输出结果的时候, 就选取概率最大(也就是值对应最大的)结果预测目标。

3. 实验

为了对本文所提出的模型应用于关系抽取任务中的有效性进行验证, 将模型在 SemEval-2010 Task8 数据集上测试实验, 并与其他几种方法进行对比。介绍模型涉及到的数据集, 参数设置以及实验结果。

3.1. 数据集

SemEval-2010 Task8 数据集是实体关系抽取任务常用的标准数据集。其包含训练集(Training Dataset)和测试集(Test Dataset)两部分, 比例分配如下表 1 所示。

Table 1. SemEval-2010 Task8Dataset

表 1. SemEval-2010 Task8 数据集

| 训练集/个数据样例 | 测试集/个数据样例 | 总数据/个数据样例 |
|-----------|-----------|-----------|
| 8000 | 2717 | 10,717 |

数据集中的每一条数据都包括一个完整的语句、标注的两个实体、一个关系以及注释, 且数据中的信息标注都是由人工完成, 不包含噪声, 具有极高的准确规范性。数据集中包括 9 中具体的关系类型和 1 种 other 类型(当句子中实体之前不满足前九种关系, 属于该关系类型)。

3.2. 超参数设置

词向量的嵌入阶段使用的是预训练的 GloVe100 维的词向量, 位置向量维度是 10 维, 实体上下文信息是 40 维, 卷积核窗口大小为 3, 卷积核为 150, 设置学习率为 0.5, L2 正则化系数为 0.0001, dropout 为 0.5。模型中的其他参数矩阵采用随机初始化的方式。

3.3. 实验结果与分析

在同一个环境下, 将本文提出的方法与所有对比方法在同一个语料库上进行对比实验, 实验结果均采用 *F1-score* 来进行评估。其模型的实验结果如表 2 所示。

Table 2. Resulting data of model experiment

表 2. 模型实验结果

| Model | Feature Sets | <i>F1</i> |
|-----------------|---|--------------|
| CNN [3] | word, position, word pair, words around word pair | 79.81 |
| PCNN [4] | word, position | 81.47 |
| ATT-BiLSTM [7] | word, position | 82.59 |
| PCNN-ATT-BiLSTM | word, position, context before and after words | 82.92 |

由表 1 可知, 通过对比分析, 在各个模型中获得的 *F1* 值的大小可以看出, 本文所提模型 PCNN-ATT-BiLSTM 的 *F1* 值最高。基于混合神经网络的模型的性能是优于单个网络模型的, 混合神经网络模型得益于各个模型在关系抽取中的作用互补, 可以从多个角度来进行特征提取训练, 从而更精确地提取关系特征。例如, 单个 PCNN 模型只能关注到序列中的局部特征, 而基于 PCNN 和 Bi-LSTM 两个神经网络模型的 PCNN-ATT-BiLSTM 混合模型可以通过 Bi-LSTM 来捕获文本数据语句的上下文信息和浅层语义特征以及 PCNN 更好地获取特征之间的联系, 并结合注意力机制捕获文本语句的关键信息, 最终提取到整体关键特征, 从而得到更好的关系抽取结果, 进而验证了所提模型的有效性。

为了能更直观的将各个模型的在数据集 SemEval-2010 Task8 的分类性能展示出来, 采用折线图将实验对比结果 *F1-score* 表示出来。对比结果如图 3 所示。

将本文模型与其他几种模型结果通过折线图展示出来进行对比, 可以更直观的看出来模型效果优于其他几种模型结果。

在实体关系抽取任务中, 利用的特征信息越多, 对分类效果会越好。考虑输入层的特征信息的因素, 通过实验来验证。实验结果如表 3 所示。

从上述表格结果分析可以得出, 对于本文所提出的 PCNN-ATT-BiLSTM 模型特征输入进行比较, 仅利用词向量信息的实验结果要比词向量和位置向量信息结合的情况以及词向量和实体对上下文信息 (context before and after words) 结合的情况都要差。在同样使用词向量信息, 但是结合词位置向量信息的模型效果要比结合实体对上下文信息的效果要好, 这就证明了实体的位置信息要比实体对上下文信息更重要。但是当结合词信息、位置信息以及实体对上下文信息作为特征输入时, *F1* 值最高, 所以向网络模型中输入的特征信息越多, 关系模型的分类效果越好。

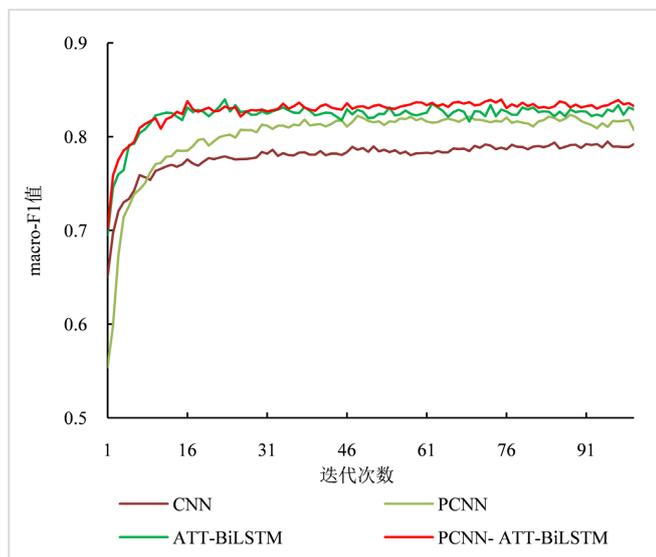


Figure 3. Model comparison results

图 3. 模型对比结果

Table 3. Resulting data of feature experiment

表 3. 特征实验结果

| Model | Feature Sets | F1 |
|-----------------|--|--------------|
| PCNN-ATT-BiLSTM | word | 78.89 |
| | word, position | 82.36 |
| | word, context before and after words | 81.94 |
| | word, position, context before and after words | 82.92 |

采用折线图将本文所提出来的模型基于输入特征的实验对比结果 $F1$ -score 表示出来。对比结果如图 4 所示: WV 表示的是词向量, POS 表示的是位置向量, CON 表示的是实体对上下文信息。

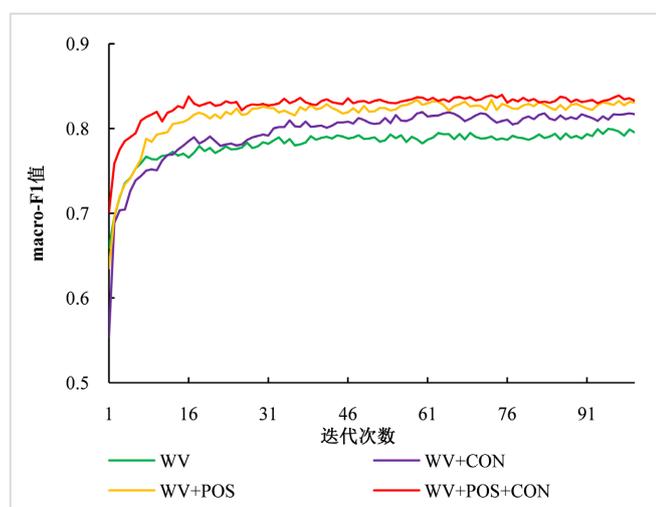


Figure 4. Feature comparison results

图 4. 特征对比结果

将本文所提出的模型利用输入特征的不同实验结果通过折线图展示出来, 可以更直观的看到当结合词信息、位置信息以及实体对上下文信息作为特征输入时, 关系抽取性能更好。

4. 总结

针对非结构化文本实体关系抽取方法中存在的上下文环境信息难以准确表征、句子间的实体关系特征未被充分利用等问题, 本文提出了一种新的基于神经网络结构和注意力机制结合关系抽取模型 PCNN-ATT-BiLSTM。该模型是基于 PCNN 和 Bi-LSTM 两种神经网络模型, 能够更好的挖掘出文本语句中包含的上下文语义信息, 可以较好的学习相关环境特征, 避免了使用复杂的特征工程, 最终通过在数据集 SemEval-2010 Task8 上进行实验, 取得了较好的实验效果。未来本文还可以从特征利用方面入手, 在实体关系抽取模型中加入语义角色、词性、语法结构等特征, 以争取获得更好的实体关系抽取性能。

基金项目

河北省社会科学基金项目: 网络叙词表的语义表示及知识服务机制研究(HB20TQ003)。

参考文献

- [1] 庄传志, 靳小龙, 李忠, 孙智. 基于注意力机制的混合神经网络关系分类方法[J]. 计算机工程与科学, 2020, 42(1): 64-70.
- [2] Hendrickx, I., Kim, S.N., Kozareva, Z., Nakov, P., Séaghdha, D.Ó., Padó, S., *et al.* (2009) SemEval-2010 Task 8: Multi-Way Classification of Semantic Relation between Pairs of Nominals. *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions*, Boulder, 4 July 2009, 94-99. <https://doi.org/10.3115/1621969.1621986>
- [3] 刘辉, 江千军, 桂前进, 张祺, 王梓豫, 王磊, 等. 实体关系抽取技术研究进展综述[J]. 计算机应用研究, 2020, 37(S2): 1-5.
- [4] Lin, Y., Shen, S., Liu, Z., Luan, H. and Sun, M. (2016) Neural Relation Extraction with Selective Attention over Instances. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Berlin, 7-12 August 2016, 2124-2133. <https://doi.org/10.18653/v1/P16-1200>
- [5] Xu, Y., Jia, R., Mou, L., Li, G., Chen, Y.C., Lu, Y.Y., *et al.* (2016) Improve Relation Classification by Deep Recurrent Neural Networks with Data Augmentation. *Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers*, Osaka, 11-16 December 2016, 1461-1470.
- [6] 王红, 史金钊, 张志伟. 基于注意力机制的 LSTM 的语义关系抽取[J]. 计算机应用研究, 2018, 35(5): 1417-1420+1440.
- [7] Kim, Y. (2014) Convolutional Neural Networks for Sentence Classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, 25-19 October 2014, 1746-1751. <https://doi.org/10.3115/v1/D14-1181>
- [8] Zeng, D., Liu, K., Lai, S., Zhou, G.Y. and Zhao, J. (2014) Relation Classification via Convolutional Deep Neural Network. *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, Dublin, 23-29 August 2014, 2335-2344.
- [9] Zeng, D., Liu, K., Chen, Y. and Zhao, J. (2015) Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, 17-21 September 2015, 1753-1762. <https://doi.org/10.18653/v1/D15-1203>
- [10] Socher, R., Huval, B., Manning, C.D. and Ng, A.Y. (2012) Semantic Compositionality through Recursive Matrix-Vector Spaces. *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Jeju Island, 12-14 July 2012, 1201-1211.
- [11] Zhang, D. and Wang, D. (2015) Relation Classification via Recurrent Neural Network. arXiv:1508.01006.
- [12] Zhang, X.B., Chen, F.C., Huang, R.Y., *et al.* (2018) Entity Relationship Extraction Based on the Fusion of CNN and Bidirectional LSTM. *Chinese Journal of Network and Information Security*, 4, 48-55.
- [13] Zhou, P., Shi, W., Tian, J., Qi, Z.Y., Li, B.C., Hao, H.W., *et al.* (2016) Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Berlin, 7-12 August 2016, 207-212.

<https://doi.org/10.18653/v1/P16-2034>

- [14] Pennington, J., Socher, R. and Manning, C.D. (2014) GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, 25-29 October 2014, 1532-1543. <https://doi.org/10.3115/v1/D14-1162>
- [15] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [16] 石磊, 王毅, 成颖, 魏瑞斌. 自然语言处理中的注意力机制研究综述[J]. *数据分析与知识发现*, 2020, 4(5): 1-14.