

基于位置编码与实体交互信息的关系抽取方法^①



厉晓妍, 张德平

(南京航空航天大学 计算机科学与技术学院, 南京 211106)

通信作者: 张德平, E-mail: depingzhang@163.com

摘要: 关系抽取作为信息抽取领域的重要研究课题, 其主要目的是抽取句子中已标记实体对之间的语义关系, 对句子语义理解及知识库构建有着重要作用. 针对现有抽取方法中未能充分利用单词位置信息和实体间的交互信息导致重要特征丢失的问题, 本工作提出一种基于位置编码与实体交互信息的关系抽取方法 (BPI-BERT). 首先将新型位置编码融入 BERT 预训练语言模型生成的词向量中后使用平均池化技术得到实体和句子向量, 再利用哈达玛乘积构造实体交互信息, 最后将实体向量、句子向量及交互信息向量拼接得到关系向量并输入到 Softmax 分类器进行关系分类. 实验结果表明 BPI-BERT 在精准率和 $F1$ 上较现有方法有提高, 证明了 BPI-BERT 的有效性.

关键词: 位置编码; 实体交互; 预训练语言模型; 关系抽取; 监督学习; 深度学习; 特征融合

引用格式: 厉晓妍, 张德平. 基于位置编码与实体交互信息的关系抽取方法. 计算机系统应用, 2022, 31(6): 238-244. <http://www.c-s-a.org.cn/1003-3254/8528.html>

Relation Extraction Method Based on Entity Interaction and Position Encoding

LI Xiao-Yan, ZHANG De-Ping

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: As an important research program in the field of information extraction, relation extraction is mainly aimed at extracting semantic relations between labeled entity pairs in sentences, which plays an important role in sentence semantic understanding and knowledge base construction. The existing extraction methods fail to make full use of the word position information and the interaction information between entities, which leads to the loss of effective features in relation extraction. To solve this problem, this study proposes a relation extraction method BPI-BERT based on the interaction information between position encoding and entities. The novel position coding is integrated into the word vector generated by the BERT pre-trained language model, and the entity and sentence vectors are obtained through the average pooling technology. The Hadamard product is used to construct the entity interaction information. Finally, the entity vector, sentence vector, and interaction information vector are stitched together to obtain the relation vector which is then input to the Softmax classifier for relation classification. The experimental results show that the precision and $F1$ of BPI-BERT are significantly improved compared with those of the existing methods, and thus the effectiveness of BPI-BERT is proved.

Key words: position encoding; entity interaction; pre-trained language model; relation extraction; supervised learning; deep learning; feature fusion

1 引言

关系抽取的任务是预测给定非结构化文本序列中

一个实体对之间的关系, 预测的结果通常以三元组的形式呈现. 关系抽取作为知识图谱、智能问答、推荐

^① 基金项目: 十四五装备预研项目 (JCKY2020605C003)

收稿时间: 2021-09-10; 修改时间: 2021-10-14; 采用时间: 2021-10-15; csa 在线出版时间: 2022-05-26

系统等上层应用的基础支撑,是一项重要的自然语言处理任务。随着互联网高速发展,非结构化文本激增,如何使用关系抽取技术准确抽取结构化知识成为现阶段的研究热点和难点。

目前主流的方法大多是基于监督学习的,即给出预定义的关系集合,将关系抽取转换为关系分类任务。这些方法通常以 LSTM^[1]、CNN^[2]、GCN^[3] 等网络为基础编码器,融合多种特征后进行关系分类。Xu 等人^[4] 提出基于 LSTM 的融合句法依存分析树的最短路径 (SDP) 以及词向量特征、词性特征、WordNet 特征和句法类型特征进行关系抽取相关工作,并取得了不错的效果,但是过多地依赖一些已有的自然语言处理工具包,可能出现错误的积累传播;于是 Zeng 等人^[5] 提出使用 CNN 进行关系抽取,利用卷积深度神经网络 (CDNN) 来提取词汇和句子层次的特征,将所有的单词标记作为输入,而无须复杂的预处理,解决了从预处理系统中提取特征可能会导致错误传播并阻碍系统性能问题。但 CNN 不擅长抽取全局特征,因此为了弥补 CNN 在这方面的缺陷;亢晓勉等人^[6] 在文档级神经翻译任务中设计了 5 种位置编码对篇章结构编码以组织全局特征,提升翻译质量,这 5 种编码分别是绝对位置编码、相对位置编码、绝对深度编码、相对深度编码和路径位置编码。同样的思想在 Transformer 系列的模型中有所体现,如 Vaswani 等人^[7] 在 Transformer 模型中提出使用正余弦变换表示单词的绝对位置以弥补自注意力机制中顺序丢失的问题。虽然绝对位置能反映单词的位置,但是不能反映单词与单词的内在位置关系,因此 Dai 等人^[8] 在 Transformer-XL 中提出了相对位置编码,将单词之间的距离差作为词向量,不仅标记位置而且反映单词之间的远近。此外, Zhu 等人^[9] 在图神经网络中提出另一种位置标记,将主体和客体标记为 0 和 1,其余标记为 2,目的是突出实体的位置,但没有突出实体和其他单词的位置关系。上述方法的位置编码大部分只关注了绝对位置或相对位置中的一种,却忽略了关系抽取任务的特殊性。

另一方面注意力机制作为捕捉重点信息的手段被引入关系抽取任务中,其实现方法是将单词之间的相似度分数作为权重与词向量相乘。后来的学者在此基础上开展了相关工作,Transformer^[7] 系列的预训练语言模型如 BERT^[10]、GPT3^[11]、RoBERTa^[12] 等在自然语言处理任务中表现优异,其内部的多头注意力机制

可以捕获丰富的上下文关系。在上述预训练语言模型中, BERT 的应用最为广泛,如 Eberts 等人^[13] 使用 BERT 生成词向对文本进行片段划分,然后再去判断每个片段所属的实体类型;同时 Rajpurkar 等人^[14] 还将 BERT 应用于 SQuAD 问答任务中,其目的是找到答案片段的开始和结束位置。与上述方法类似,本工作中也使用 BERT 生成词向量。

单词建模后通常要进行关系向量的构造,现有的构造方法大多是基于三段式池化拼接方法来展开的。起初, Zeng 等人^[15] 在 PCNN 中提出以实体为分界点,将句子分成 3 段,分别进行最大池化后拼接形成关系向量;后来, Guo 等人^[16] 提出在利用 Mask 机制将主体、客体和句子向量剥离后池化拼接;而 Zhu 等人^[9] 在 GP_GCN 模型中构造关系向量时,直接将三者池化拼接得到关系向量,句子向量没有剥离主体和客体向量。Lee 等人^[17] 提出使用注意力机制生成关系向量,这些方法虽然从不同角度考虑了句子向量的组成部分,但都忽略了实体之间的交互信息,导致分类器的表现一般。

针对上述两个问题,本文提出了新的解决方法 BPI-BERT (method based on position encoding and entity interaction with BERT pretrained model),其内容和贡献如下:

(1) 对于位置编码没有针对性的问题,提出了一种新的位置编码策略,从多种角度考虑位置信息,使得模型在不过多依赖外来工具包的情况下,依然可以捕捉局部及全局的信息,避免了引入外来工具导致的误差累积传播;

(2) 对于关系向量缺乏实体交互信息的问题,提出了一种实体交互策略,使用哈达玛乘积构造主客体之间的交互信息以增强关系向量,有益于分类器进行关系选择;

(3) 实验表明, BPI-BERT 在 SemEval-2010 Task 8 数据集上的精准率和 F1 均有提高,证明了本文所提方法是有效的。

2 BPI-BERT

给定一个包含一对实体 (e_1, e_2) 的句子 T , 其中 $T = [CLS, T_1, \dots, T_i, \dots, T_n]$, 模型的任务是对给定句子中的一对实体进行关系分类, 预测结果为关系 $r \in R$, 其中 R 是预定义的关系集合。图 1 是 BPI-BERT 的模型结

构图,模型共分为3个模块,分别是BERT编码器、位置编码器和关系分类器。

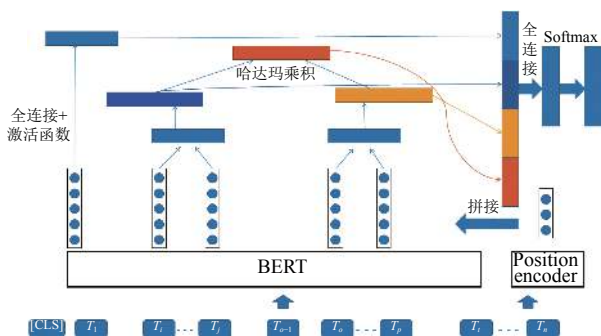


图1 BPI-BERT 模型概览图

如上图所示, BERT 编码器将输入的单词序列转换为词向量; 位置编码器根据实体标记得到的位置后对每个单词进行位置编码, 随后将二者拼接后得到最终的词向量; 关系分类器首先根据实体在句子中的位置分别提取词向量中的主体向量、客体向量及句子向量后再使用平均池化处理成同样尺寸的向量, 对主客体进行哈达玛乘积得到实体交互向量, 最后将四个向量拼接输入到 Softmax 函数中进行关系分类。

2.1 位置编码

本文所提位置编码记录了单词的绝对位置, 单词与实体之间的相对距离, 实体之间的相对距离. 这种记录方法能够清晰地记录实体、单词的位置, 相对关系. 其具体步骤如下:

(1) 读入句子, 根据实体标记得得到两个实体的开始和结束位置 $e_1^{start}, e_1^{end}, e_2^{start}, e_2^{end}$, 除此之外还需要记录下句子长度 l .

(2) 位置向量的初始维度设置为 4, 当前单词的位置用 i 表示, 单词的位置向量 p_i 表示为 $[i/l, (i - e_1^{start})/l, (i - e_2^{start})/l, (e_2^{start} - e_1^{start})/l]$.

(3) 使用单层前馈神经网络对步骤 (2) 中的向量进行转换 $p'_i = w_p \cdot p_i$, 其中, w_p 是参数矩阵, p'_i 是变换后的位置向量。

p_i 中第 1 个维度表示当前单词在句子中的位置, 第 2 个维度表示当前单词和实体 1 的相对位置, 第 3 个维度表示当前单词和实体 2 的相对位置, 第 4 个维度表示主体和客体的相对位置. 为了避免句子长度对位置编码的影响, 除以句子长度 l 的目的是消除句子长度对位置编码的影响, 起到归一化的作用. 本文所提位置编码不仅记录了单词位置, 突出了单词与实体之间

的相对位置, 实体之间的相对位置, 还记录了单词在句中的位置, 使得模型可以捕捉句子间的内在关系, 起到数据增强的作用. 图 2 是位置编码的一个实例。

如图 2 所示, 给定的句子是“俗话说得好, 失败是成功的一部分”, 其中, “失败”和“成功”是标记的两个实体, “部分”是关系, 那么对于“好”字的位置编码是 $[5/14, -1/14, -4/14, 2/14]$. 为了使位置编码更有表现能力, 本方法使用前馈神经网络对初始位置编码进行了变换。

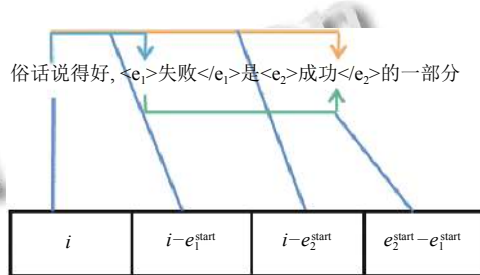


图2 位置编码实例

2.2 BERT 编码器

BERT 自身的多头注意力机制可以捕获丰富上下文信息, 因此本工作使用 BERT 生成词向量, 供后续分类任务使用. 另外 BERT 内部已经实现自注意力机制, 所以本方法没有再添加其他注意力的计算, 以减少模型的参数数量和训练时间. BPI-BERT 词向量生成方法与 Wu 等人^[18]类似, 都是将实体标记字符转换为特殊字符后与句子中的其他单词共同输入到 BERT 中进行编码. BERT 中的自注意力机制过程如图 3 所示。

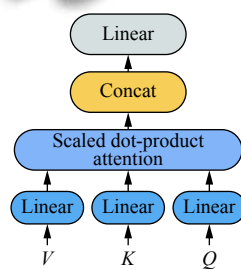


图3 注意力机制

将句子序列 $T = [CLS, T_1, \dots, T_i, \dots, T_n]$ 输入到 BERT 中, 输出为 H :

$$H = [h_0, h_1, \dots, h_i, \dots, h_n]$$

H 是序列的词向量表示, 计算公式如下。

$$Attention = \text{Softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (1)$$

$$H = \text{MultiHead}(T) \quad (2)$$

式(1)是单头注意力的计算公式,式(2)中的 *MultiHead* 是 BERT 多头函数,是多个单头注意力拼接的结果,目的是利用多头注意力挖掘词向量在不同空间维度上的含义,从而获得带有丰富上下文语义的词向量.将 BERT 生成的词向量和第 2.1 节得出的位置向量进行拼接,得到最终的词向量表示:

$$H' = [h_i; p_i'] \quad (3)$$

其中,“;”表示拼接操作.

2.3 关系分类

传统关系向量的构造方法将主体向量、客体向量和句子向量池化后拼接后进行关系分类.向量拼接的方法虽然保证信息的全面性但忽略实体之间的交互,因此本文提出利用哈达玛乘积获得实体交互向量.

哈达玛乘积是向量对位相乘的一种计算方法,在此处可以看作是一种特征交叉的手段,用于学习实体之间高级且复杂的特征.这种交叉特征不仅保留了原有的特征维度,减少了信息损失,而且在实际编码过程中也易于操作和实现.

关系向量的构造流程如图 4 所示.

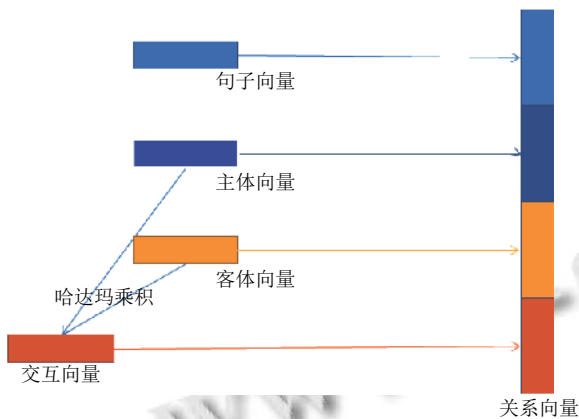


图 4 关系向量的构造

如图 4 所示,关系向量主要由 4 个部分组成,分别是主体向量、客体向量、句子向量(不包含主体和客体向量)及主客体的交互向量.主体和客体可能由多个单词或字组成,所以使用平均池化来获得相同尺寸的向量,句子向量也是如此;交互向量是将池化后的实体向量进行哈达玛乘积操作;最后将 4 种向量以上图的顺序进行拼接,得到关系向量.

$$e_1 = \text{avgpooling}(e_1^{\text{start}}, e_1^{\text{end}}) \quad (4)$$

$$e_2 = \text{avgpooling}(e_2^{\text{start}}, e_2^{\text{end}}) \quad (5)$$

$$s = \text{avgpooling}(H') \quad (6)$$

$$so = e_1 \cdot e_2 \quad (7)$$

$$r = [e_1; e_2; s; so] \quad (8)$$

式(4)和式(5)中的 start 和 end 分别代表实体的开始和结束位置; *avgpooling* 是平均池化函数,使用 *avgpooling* 的目的是保留句子和实体的整体特征,若使用最大池化可能会有信息丢失的问题.最后将处理后的结果使用单层神经网络 *FFNN* 和 *Softmax* 函数进行关系分类.

$$p_r = \text{Softmax}(\text{FFNN}(r)) \quad (9)$$

本工作中的损失函数为交叉熵损失函数.另外为了防止过拟合现象的出现,在模型的全连接层及 BERT 注意力模块中均会加入 *dropout*.

3 实验

3.1 数据集与评价标准

本任务使用的 SemEval-2010 Task 8 数据集,共包含 10 717 例子(其中 8 000 条用于训练,2 717 条用于测试),关系类别的数量是 19,其中 9 种关系类型是有向的,1 种关系类型是不属于 9 种有向关系的其他关系类型.这 9 种有向的关系类型分别是 Cause-Effect, Component-Whole, Content-Container, Entity-Destination, Entity-Origin, Instrument-Agency, Member-Collection, Message-Topic 和 Product-Producer,其他关系类型是 Others.

本方法针对的是基于监督学习的关系分类任务,首先对于模型抽取结果可以分成以下 4 种: *TP* (true positive),指的是样本是正例且被预测为正例; *FP* (false positive),指的是样本是负例但被预测为正例; *FN* (false negative)指的是样本是正例但被预测为负例; *TN* (true negative),指的是样本是负例且被预测为负例.针对分类问题本文使用精准率 Precision 和 *F1* 作为评价指标.以下是其计算公式:

$$P = \frac{TP}{TP + FP} \quad (10)$$

$$R = \frac{TP}{TP + FN} \quad (11)$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \quad (12)$$

精准率 P 是从查准率的角度评估结果,指的是预测为正例正确的数量占原本正例数量的比例;召回率是从查全率的角度评估结果,指的是预测为正例正确的数量占所有预测为正例数量的比例;分类优化的目标是将 Precision 和 Recall 都尽可能地提高,但有些极端的情况下二者不一定能正确地衡量关系分类的性能,二者会相互影响相互制约,一般来说,精准率高时召回率会低,召回率高时精准率会低,因此使用 $F1$ -Measure 将二者综合考虑,作为主要的评价指标

3.2 实验参数设置

本文使用 Uncased_L-12_H-768_A-12 版本的 BERT 预训练语言模型,层数是 12,词向量的维度是 768,模型中的其他参数设置如表 1 所示。

表 1 模型参数

Parameters	Values
Layers of BERT	12
Hidden size	768
Conversed pos dim	30
Optimizer	Adam
Learning rate	2E-5
Batch size	8
Dropout rate	0.1
Batch size	8
Epoch	10

3.3 实验结果分析

为了验证本文所提方法的有效性,将本文所提的 BPI-BERT 与在 SemEval-2010 Task 8 数据集的其他方法如 CRCNN^[19]、TRE^[20]、Entity Attention Bi-LSTM^[17] 及 Att-Pooling-CNN^[21] 在同一实验环境下进行了对比,表 2 给出了本文所提方法和上述方法在测试集上的评测结果,表 3 给出本方法消融实验的结果,表 4 给出了本文所提方法在所有关系类型上的 P 、 R 、 $F1$ 的评价结果。

表 2 各个模型的对比结果

方法	P	$F1$
CRCNN ^[19]	0.6919	0.75
TRE ^[20]	0.7312	0.7931
Att-Pooling-CNN ^[21]	—	0.859
Entity Attention Bi-LSTM ^[17]	—	0.88
本文模型	0.8487	0.8854

CRCNN 模型是以 CNN 作为特征抽取器,加入相对位置编码,池化单词向量后进行关系分类,并使用新的成对排序损失函数进行关系分类;TRE 是以 Trans-

former 为编码器,通过无监督的预训练学得句子的隐含语义特征;Entity Attention Bi-LSTM 以 Bi-LSTM 为编码器,利用注意力机制构造关系向量,Att-Pooling-CNN 在构造关系向量时与 Attention Bi-LSTM 类似,但使用的编码器是 CNN. 本方法主要是在编码器的选择、位置编码方式、实体交互信息层面与这些方法进行对比. 上述方法均在同一实验环境下运行,会有个别实验结果与原论文不一致的情况出现,与其他方法的实验对比结果如表 2 所示。

表 3 3 种模型配置的结果

配置	P	$F1$
基线 (baseline)	0.8432	0.8799
增加位置信息	0.8498	0.8835
增加交互信息	0.8498	0.8827
增加位置和交互信息	0.8487	0.8854

由表 2 数据可知,本文所提的 BPI-BERT 模型在精准率上达到 84.87%, $F1$ 得分达到 88.54%,较先前的最优结果在 $F1$ 得分上提升了 0.5%。由实验结果可知,BERT 的精准率和 $F1$ 上比 CNN 和 LSTM 高出许多,表现优异,相较于 CNN 和 RNN 等网络结构捕获了更加丰富的下文信息。

为了探索实体交互信息和位置信息对实验结果的影响,本文在原有 BERT 模型的基础上加入了不同配置以验证本文所提方法的有效性. 下表是不同配置实验结果。

本方法的 baseline 配置是未加入位置编码和实体交互信息,只使用 BERT 进行编码. 另外本文还提供了 3 种配置:一是加入位置编码,二是加入实体交互向量,三是在加入位置编码的基础上增加实体交互向量. 从实验结果上来看,加入位置编码的配置比 baseline 在 $F1$ 上提升了近 0.3%,提高的原因是位置编码中蕴含了单词与实体的内在联系,突出了实体信息,帮助模型定位实体的位置以及捕捉与实体相关的句子片段;加入实体交互向量提升模型效果的原因是对实体交互信息有了向量上的定义,并把这种向量表示融入到原有的关系向量中. 最后一种配置是同时加入位置编码和交互信息,比前两种的单独配置在 $F1$ 上提高了近 0.2%. 3 种配置在 $F1$ 上的结果,较 baseline 提升明显;在精准率上的差别不大并且同时加入位置编码和实体交互信息的模型比只加其中一种信息的模型在精准率上甚至还降低了一些. 综上所述,本文所提出的位置编码和

交互信息方法是有效的。

为了探索模型在不同关系类型上的性能表现, 本文将模型在所有关系类型上的实验结果进行了统计, 如表4所示。

表4 BPI-BERT 在各关系类型上的评价结果 (%)

类型	P	R	F1
Cause-Effect	91.62	93.29	92.45
Component-Whole	84.24	83.97	84.11
Content-Container	87.37	90.10	88.72
Entity-Destination	92.33	94.86	93.58
Entity-Origin	88.03	88.37	88.20
Instrument-Agency	82.78	80.13	81.43
Member-Collection	86.38	87.12	86.75
Message-Topic	86.16	95.40	90.55
Product-Producer	85.00	88.31	86.62
Other	69.25	61.01	64.87

由表4数据可知, 本文所提的 BPI-BERT 模型在 Cause-Effect、Entity-Destination、Content-Container 等关系类型上表现较好, P 、 R 、 $F1$ 的指标都达到近 90%。而在 Other 类型上表现较差, 比其他关系类型在评价结果上降低近 15%, 原因可能是 Others 类型的样本数量偏多, 出现样本不平衡的问题, 后续工作会在数据类别不平衡问题上研究和改进。

4 结论与展望

本研究从位置编码出发进行关系分类, 在不依赖外部工具包的情况下, 加入句子本身包含的先验知识, 从而避免了外部工具包带来的错误传播; 此外, 本方法在传统关系向量构造方法上做出了改进, 加入了实体交互信息, 使得关系向量不再是简单拼接实体和句子信息。从实验结果上来看 BERT 预训练语言模型表现确实优异, 是今后工作生成词向量的有效工具。

本文所提方法针对的是基于监督学习的关系抽取, 需要大量且优质的标注数据, 然而标注数据通常不易获得且标注困难, 所以今后的研究方向会偏向于数据生成和数据标注, 以此来扩展关系抽取的应用领域和应用场景。另外关系类别分布不平衡的问题在现实中十分常见, 如何解决数据不平衡的问题也是之后的研究内容。

参考文献

1 Hochreiter S, Schmidhuber J. Long short-term memory.

Neural Computation, 1997, 9(8): 1735–1780. [doi: 10.1162/neco.1997.9.8.1735]

2 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe Nevada: Curran Associates Inc., 2012. 1097–1105.

3 Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv: 1609.02907, 2017.

4 Xu Y, Mou LL, Li G, *et al.* Classifying relations via long short term memory networks along shortest dependency paths. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: ACL, 2015. 1785–1794.

5 Zeng DJ, Liu K, Lai SW, *et al.* Relation classification via convolutional deep neural network. Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. Dublin: Dublin City University and Association for Computational Linguistics, 2014. 2335–2344.

6 亢晓勉, 宗成庆. 融合篇章结构位置编码的神经机器翻译. 智能科学与技术学报, 2020, 2(2): 144–152.

7 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.

8 Dai ZH, Yang ZL, Yang YM, *et al.* Transformer-XL: Attentive language models beyond a fixed-length context. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: ACL, 2019. 2978–2988.

9 Zhu H, Lin YK, Liu ZY, *et al.* Graph neural networks with generated parameters for relation extraction. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: ACL, 2019. 1331–1339.

10 Devlin J, Chang MW, Lee K, *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis: ACL, 2019. 4171–4186.

11 Radford A, Narasimhan K, Salimans T, *et al.* Improving language understanding by generative pretraining. 2018.

12 Liu YH, Ott M, Goyal N, *et al.* RoBERTa: A robustly optimized BERT pretraining approach. arXiv: 1907.11692,

- 2019.
- 13 Eberts M, Ulges A. Span-based joint entity and relation extraction with transformer pre-training. *Frontiers in Artificial Intelligence and Applications*, 2020, 325: 2006–2013.
 - 14 Rajpurkar P, Zhang J, Lopyrev K, *et al.* SQuAD: 100, 000+ questions for machine comprehension of text. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin: Association for Computational Linguistics, 2016. 2383–2392.
 - 15 Zeng DJ, Liu K, Chen YB, *et al.* Distant supervision for relation extraction via piecewise convolutional neural networks. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon: ACL, 2015. 1753–1762.
 - 16 Guo ZJ, Zhang Y, Lu W. Attention guided graph convolutional networks for relation extraction. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence: ACL, 2019. 241–251.
 - 17 Lee J, Seo S, Choi YS. Semantic relation classification via bidirectional LSTM networks with entity-aware attention using latent entity typing. *Symmetry*, 2019, 11(6): 785. [doi: [10.3390/sym11060785](https://doi.org/10.3390/sym11060785)]
 - 18 Wu SC, He YF. Enriching pre-trained language model with entity information for relation classification. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. Beijing: ACM, 2019. 2361–2364.
 - 19 dos Santos C, Xiang B, Zhou BW. Classifying relations by ranking with convolutional neural networks. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Beijing: ACL, 2015. 626–634.
 - 20 Alt C, Hübner M, Hennig L. Improving relation extraction by pre-trained language representations. *arXiv: 1906.03088*, 2019.
 - 21 Shen YT, Huang XJ. Attention-based convolutional neural network for semantic relation extraction. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. Osaka: The COLING 2016 Organizing Committee, 2016. 2526–2536.