

文章编号: 2095-2163(2023)02-0204-06

中图分类号: TP391.1

文献标志码: A

# 基于 BERT 和注意力引导图卷积网络的关系抽取

徐瑞涓, 高建瓴

(贵州大学 大数据与信息工程学院, 贵阳 550025)

**摘要:** 针对现有图卷积网络在关系抽取任务中存在文本语义, 语法表征不准确和在不同树结构上并行化计算较难等问题, 本文提出一种基于 BERT 和注意力引导图卷积网络的关系抽取模型。首先, 在模型的输入层使用 BERT 和 Bi-LSTM 编码出适应于上下文语境的词向量; 其次, 对输入的树结构采用最短路径为中心的修剪方式, 减少树中的无关信息; 最后, 在模型中引入多头注意力机制, 自动学习不同子空间内对关系提取有用的相关子结构, 并在 TACRED 数据集上进行验证。实验结果表明, 相对于基线模型, 本文提出的模型显著提高了实体关系抽取的  $F_1$  值。

**关键词:** 关系抽取; BERT; 最短路径; 多头注意力机制; 图卷积

## Relation extraction based on BERT and attention-guided graph convolution networks

XU Ruijuan, GAO Jianling

(College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China)

**【Abstract】** To address the problems of text semantics, inaccurate syntactic representation and difficult parallelized computation on different tree structures in existing graphical convolution networks for relationship extraction task, a relation extraction model based on BERT and attention-guided graphical convolution networks is proposed. First, word vectors adapted to the context are encoded in the input layer of the model using BERT and Bi-LSTM. Then the shortest path-centered pruning is applied to the input tree structure to reduce irrelevant information in the tree. Finally, a multi-head attention mechanism is introduced in the model to automatically learn relevant substructure useful for relation extraction in different subspaces. The experimental results on TACRED dataset show that the proposed model in this paper significantly improves the  $F_1$  value of entity relationship extraction compared to the baseline model.

**【Key words】** relation extraction; BERT; shortest path; multiplex attention mechanism; graph convolution

## 0 引言

随着互联网技术的发展和日益激增的网络数据, 从大量数据中获取关键信息是当前自然语言处理(Natural Language Processing, NLP)任务中的一个研究热点, 良好的关系抽取(Relation Extraction, RE)表现为自然语言处理的下游任务奠定了基础。实体关系抽取的任务的目标是从非结构化文本中抽取出实体关系三元组(SPO)。传统的关系抽取方法主要有两种主流框架, 第一种是流水线的方法, 第二种是联合抽取的方法。流水线的方法首先识别句中的实体, 然后进行实体关系的抽取, 该方法形式简单, 操作灵活, 但缺点仍然很明显, 其级联特性在进行实体抽取时产生的误差会一直延续到关系抽取任

务中, 造成错误传播<sup>[1]</sup>。针对流水线方法存在的缺点, 联合抽取的方法将实体识别和关系抽取两个任务合并在一个模型中进行统一建模, 但是此方法模型的复杂性比较高。

虽然传统的流水线方法和联合抽取的方法已经在实体关系抽取任务中有出色的表现, 但都不能处理欧几里得空间数据, 并且在文本中存在重叠关系时, 往往只能预测出一种关系, 因此传统的关系抽取模型依旧面临挑战。

在通过图卷积网络进行关系抽取任务中, 依存分析图是常用的构图方式之一, 依存分析图展示了文本之间的依赖关系。基于依存关系的模型, 能够捕获字面形式上比较模糊的远距离句法关系(如从句比较长或者有着复杂的范围限定的情况)<sup>[2]</sup>。

**作者简介:** 徐瑞涓(1997-), 女, 硕士研究生, 主要研究方向: 深度学习、自然语言处理; 高建瓴(1969-), 女, 硕士, 副教授, 硕士生导师, 主要研究方向: 数据库系统、数据挖掘。

**通讯作者:** 高建瓴 Email: 454965711@qq.com

收稿日期: 2022-04-13

2018 年 Zhang 等人<sup>[3]</sup> 针对依存树中包含大量无关系信息, 提出一种基于修剪依存树的图卷积神经网络, 并用于实体关系抽取问题, 将修剪后的句法依存树引入图卷积网络进行实体关系抽取任务。2019 年, Guo 等人<sup>[4]</sup> 提出了注意力引导的图卷积网络模型 AGGCN, 将完整的依存树作为模型的输入, 结合多头注意力机制使模型自动学习不同特征空间有用的子结构, 再通过注意力引导层使用多头注意力机制构造  $N$  个注意力引导邻接矩阵, 将输入依存树转换为  $N$  个不同的全连接加权图后进行关系分类任务。Tian 等人<sup>[5]</sup> 提出一种依赖驱动的方法, 将基于图卷积网络的注意力机制应用于从现成的依赖解析器获得的依赖树中的不同上下文词汇, 以区分不同词汇依赖的重要性。本文提出一种基于 BERT 和注意力引导图卷积网络的关系抽取的方式。首先, 在模型表示层摒弃传统的词向量表示方法, 引入预训

练模型 BERT 结合 Bi-LSTM 网络, 训练出更适应上下文信息的词向量, 经过微调后的 BERT 模型将更适应于当前语料并具备更好的语义信息, 增强模型的泛化能力; 其次, 在构建的依存图中, 采用最短路径的修剪树的方式, 修剪掉图中的无关信息, 降低噪声词的影响, 增强其鲁棒性; 最后, 将多头注意力机制与图卷积网络融合, 使得模型能够学习到不同子空间的特征和不同重要程度的信息, 结合密集连接, 使模型能够更好地学习图形表示。

### 1 方法介绍

本文的模型 ASP-GCN (shortest path attention-guided graph convolutional network) 总体框架如图 1 所示。模型主要由 BERT 词向量表示模块, Bi-LSTM 网络模块, 融合最短路径注意力引导图卷积网络模块 (graph convolutional network, GCN) 构成。

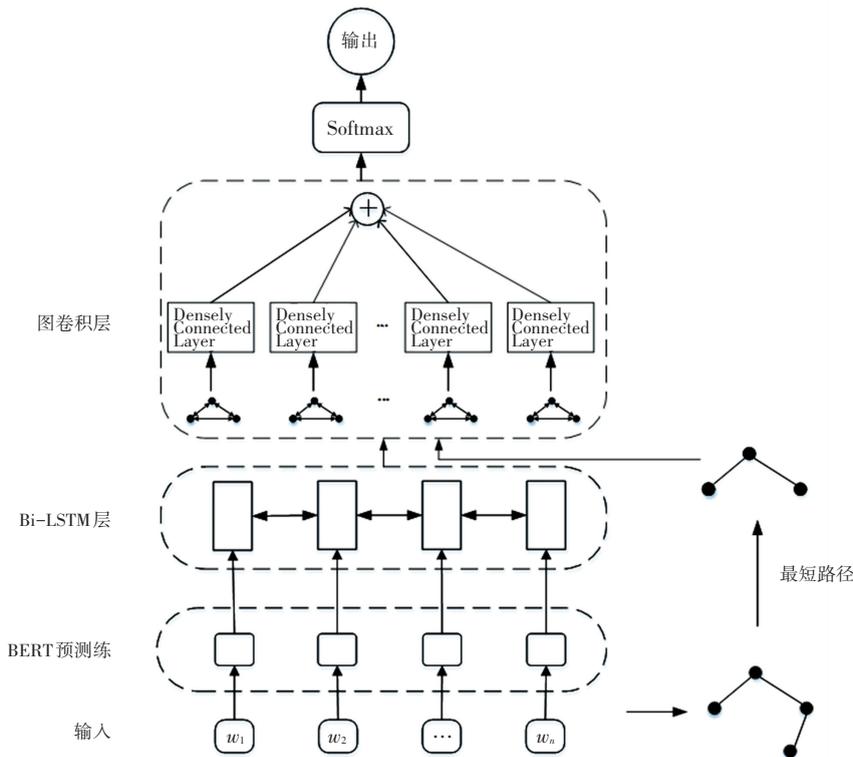


图 1 ASP-GCN 总体框架

Fig. 1 ASP-GCN general framework

#### 1.1 基于 BERT 的表示层

在自然语言处理任务中, 模型执行任务的第一步是将文本转化为计算机能够理解的表达形式, 输入到模型中进行接下来的操作。2018 年 Devlin 等<sup>[6]</sup> 提出 BERT 预训练模型, 以 Transformer 模型的编码器作为基模型和多头注意力机制构建的模型, 保存深层的特征信息, 具有很好的泛化能力。

例如输入句子  $L = \{w_1, w_2, \dots, w_n\}$  是带有

“[CLS]”和“[SEP]”的句子序列,  $n$  为模型设置的输入句子的最大长度, 其输入表示包含 3 个部分, 分别是词嵌入 (token embedding), 位置嵌入 (position embedding) 和分割嵌入 (segment embedding) 3 个嵌入特征的单位之和, 如图 2 所示。BERT 模型处理每一个  $w_i$  的输出向量  $e_i$  计算方式, 式(1):

$$e_i = E_{token}(w_i) + E_{seg}(w_i) + E_{pos}(w_i) \quad (1)$$

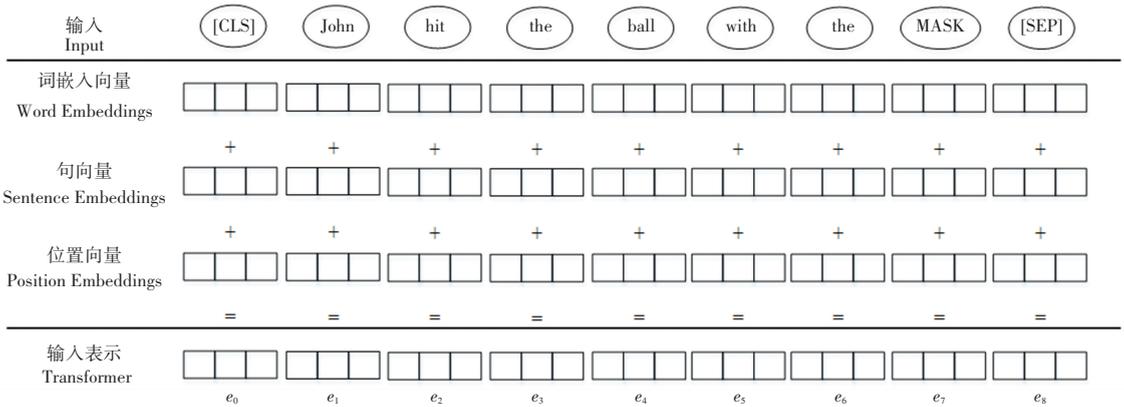


图 2 BERT 模型输入结构

Fig. 2 Input structure of BERT model

将模型输出的  $[e_1, e_2, \dots, e_n]$  向量作为 BERT 模型内部 Transformer 编码网络部分的输入表示,如图 3 所示。

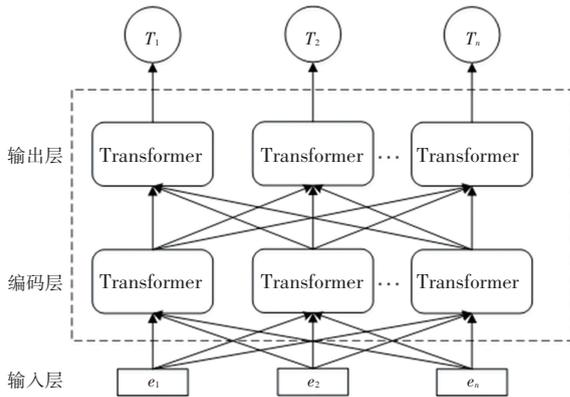


图 3 BERT 模型网络结构

Fig. 3 Network structure of BERT model

BERT 通过预训练完成模型参数的微调,得出输入序列的词嵌入表示为  $[T_1, T_2, \dots, T_n]$ 。相比传统的 Word2vec, Glove 等预训练模型来说, BERT 更能够充分考虑抽取文本的上下文信息,获得更为精确的词向量。

### 1.2 Bi-LSTM 层

长短期记忆神经网络 (Long Short Term Memory, LSTM) 是一种改进型的 RNN 模型,该模型有效地解决了传统 RNN 模型存在的梯度消失,从而限制网络学习长期依赖的问题<sup>[7]</sup>。Bi-LSTM 模型是由结构相同的前向 LSTM 和后向 LSTM 构成,如图 4 所示。

将前向 LSTM 的隐藏向量  $[h_{l1}, h_{l2}, \dots, h_{ln}]$  和后向 LSTM 的隐藏向量  $[h_{r1}, h_{r2}, \dots, h_{rn}]$  拼接得到 Bi-LSTM 网络的隐藏向量矩阵  $H = [h_1, h_2, \dots, h_n]$ ,  $H \in R^{n \times 2h}$ 。

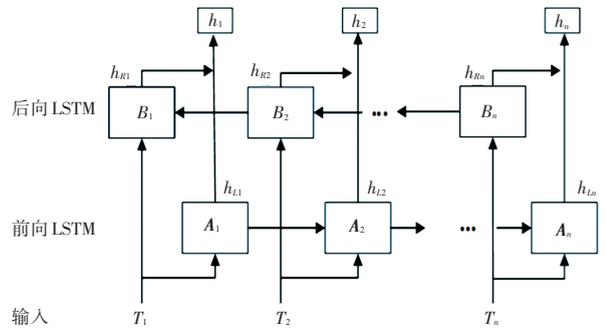


图 4 Bi-LSTM 网络结构

Fig. 4 Bi-LSTM network structure

现有 GCN 模型高度依赖从正确的解析树中提取关键信息,而现有的解析算法在许多情况下产生的并不是一棵完美的树,为了解决这些问题,本文在模型中加入 Bi-LSTM 层,通过 BERT 预训练后每个词表示为一个向量  $T_n$ ,将词向量表示  $T_n$  输入到 Bi-LSTM 中,得到句中每个单词基于上下文的特征向量  $h_n$  计算如式(2):

$$h_n = BiLSTM(T_n; \theta) \quad (2)$$

其中,  $T_n$  为 Bi-LSTM 的隐藏状态向量,  $\theta$  为网络参数。

通过加入 Bi-LSTM 层,模型中包含了除关键路径以外的其他信息,与仅仅只用最短路径的方法相比,其鲁棒性得到了很大的提升。

### 1.3 GCN 模块

此模块包含两个部分,第一部分是多头注意力机制,用以分别关注文本中来自不同子空间的特征,生成不同的全连通图;第二部分是将全连通图输入图卷积网中进行特征提取。

#### 1.3.1 依存句法分析

在图卷积网络进行关系抽取任务中,无用信息过多会引起噪声传播,导致句子中的有效信息很难被关系抽取模型有效地学习,依存句法分析是自然

语言处理中传统的分析方法之一,在关系抽取任务中已经取得不错的效果。

依存关系主要通过有向弧来体现句子成分支配与被支配关系,句中的动词往往被认为是支配语句的核心。因此,本文所构建的依存句法图是以句中单词作为节点,依存关系作为边的结构图,例如:对句子“John hit the ball with the bat”进行语义依存分析,得到语义依存图如图5所示,用7条从中心词指向依存词的有向边表示存在7对依存关系,依存边上标注的标签为依存关系类型,其中单词“hit”为根节点,“John”和“hit”具有依存关系,“hit”是中心词,“John”是依存词,二者之间的关系标签是“NSUBJ”(名词充当主语)。

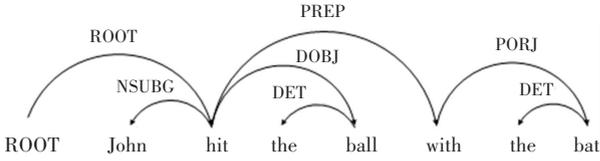


图5 语义依存树

Fig. 5 Semantic dependency tree

为了去除语法依存图中的冗余信息,消除噪声对后续关系抽取的影响,本文在依存句法图的基础上,获取句子最短依存路径(SDP)。大量研究证明最短路径可以有效表示实体之间的语义关系结构,路径中包含的词汇信息足够表征大多数文本的主要信息,对于句中存在否定关系的特殊情况,最短路径的修剪方法容易忽略句中的否定关系词,因此本文模型中融合了最短路径为中心的修剪树方式,最大限度保留有关内容的同时,删除树中的无关信息,提高模型的鲁棒性。

### 1.3.2 多头注意力机制引导层

多头注意力机制可以学习不同特征空间词之间的依赖关系和句中不同成分对词的重要程度<sup>[8]</sup>。本文将修剪好的树融合多头注意力机制分别关注文本中来自不同子空间的特征,并将不同特征的依赖树转换为完全连通的加权图,得到每个连通图的邻接矩阵 $\mathbf{A}$ ,图中每条边对应相应的权重。如: $\mathbf{A}(1)$ 为图 $G(1)$ 的邻接矩阵, $\mathbf{A}(n)$ 为图 $G(n)$ 的邻接矩阵。将任意邻接矩阵 $\mathbf{A}(t)$ 与Bi-LSTM网络的输出 $h_n$ 作为后面图卷积层的输入。通过注意力引导层使用注意力来诱导节点之间的关系,对于通过间接多跳路径连接的节点,可以通过模型中的可微分函数来捕捉这些关系。

在本文提出的模型中,通过使用多头注意力机制来计算属于邻接矩阵 $\mathbf{A}(t)$ 的信息, $\mathbf{A}(t)$ 表示通

过第 $t$ 个头注意力得到的邻接矩阵,这使得模型能够关注来自不同表示子空间的信息,计算公式(3):

$$\mathbf{A}(t) = \text{softmax}\left(\frac{Q\mathbf{W}_i^Q \times (K\mathbf{W}_i^K)^T}{\sqrt{d}}\right) \quad (3)$$

其中, $Q$ 和 $K$ 是模型中 $L-1$ 层的输出表示, $\mathbf{W}_i^Q$ 、 $\mathbf{W}_i^K$ 、 $\mathbf{W}_i^V$ 是参数矩阵。

通过注意力引导层,经过密集连接层得到一个更大的全连接图,可以更好地学习图表示。

### 1.3.3 图卷积网络

本文在特征提取和依存树后引入图卷积网络(Graph Convolutional Network, GCN)进一步对特征进行处理,通过构建依存句法图,获取最短路径为中心的依存图,结合多头注意力机制关注不同子空间的信息,生成多个带有权重的全连通图,并各自转化为相应的邻接矩阵 $\mathbf{A}(t)$ 。通过BERT预训练模型对输入文本进行编码,Bi-LSTM网络融合上下文信息,得到输入文本的词嵌入表示,将邻接矩阵 $\mathbf{A}(t)$ 作为输入词嵌入的权值,将加权结果输入到GCN中进行特征提取。

图卷积网络进行特征提取时,给定一个具有 $n$ 个节点的图,可以用 $n \times n$ 的邻接矩阵 $\mathbf{A}$ 来表示此图的结构,Marcheggiani等人<sup>[9]</sup>将边的方向性纳入模型来扩展用于编码依赖树的GCN,为树中的每个节点添加一个自循环,并且包括依赖弧的相反方向,这意味着在图中如果节点 $i$ 与节点 $j$ 之间存在一条边则 $A_{ij}$ 或 $A_{ji} = 1$ ,否则 $A_{ij} = 0$ 或 $A_{ji} = 0$ ,如果将 $L$ 层第 $i$ 个节点的输入向量表示为 $\mathbf{h}_i^{l-1}$ ,输出向量表示为 $\mathbf{h}_i^l$ ,则可以将图卷积操作写成式(4):

$$\mathbf{h}_i^l = \rho\left(\sum_{j=1}^n A_{ij} \mathbf{W}^{(l)} \mathbf{h}_j^{l-1} + \mathbf{b}^{(l)}\right) \quad (4)$$

其中, $\mathbf{W}^{(l)}$ 是权重矩阵; $\mathbf{b}^{(l)}$ 是偏置向量; $\rho$ 是激活函数; $\mathbf{h}_i^{(0)}$ 是 $x_i$ 的初始输入, $x_i \in R^d$ 中的 $d$ 是输入特征维数。

### 1.4 线性连接层

在本文的模型中包含了一个线性连接层,当依存树通过多头注意力层后,可以得到 $n$ 个不同的注意力权重矩阵 $\mathbf{A}(t)$ ,相当于在图卷积网络的输入邻接矩阵输入模型中进行特征提取,可以得到 $N$ 个不同的输出结构,本文利用线性连接层集成 $n$ 个不同的结果,整合成最终的输出特征 $h_{comb}$ ,其输出的数学定义如式(5)所示:

$$h_{comb} = \mathbf{W}_{comb} h_{out} + \mathbf{b}_{comb} \quad (5)$$

其中, $h_{out}$ 表示注意力引导层的输出结果, $h_{out} = [h^{(1)}, h^{(2)}, \dots, h^{(n)}]$ ;  $\mathbf{W}_{comb}$ 为权重矩阵; $\mathbf{b}_{comb}$ 为线性

偏置向量。

## 1.5 关系抽取

应用本文模型的数据,获得了所有标记的隐藏表示,关系抽取的任务就是预测两实体之间的关系。连接句子表示和实体表示,得到关系分类的最终表示<sup>[10]</sup>。首先,需要获得句子表示  $h_{sent}$ , 计算公式(6)如下:

$$h_{sent} = f(h_{mask}) = f(ASP - GCN(x)) \quad (6)$$

其中,  $h_{mask}$  表示被 mask 所有隐藏表示,这里 mask 的意思是只选择在句子中不是实体标记的标记表示,  $f: R^{d \times n} \rightarrow R^{d \times 1}$  是一个最大池化函数,将其  $n$  个输出向量映射到 1 个句子向量。

类似的,也可以获得实体表示。对于第  $i$  个实体的隐藏表示,其表示为式(7):

$$h_{e_i} = f(h_{e_i}) \quad (7)$$

实体表示与句子表示连接起来形成新的表示,新的表示应用于前馈神经网络(Feed Forward Neural Network, FFNN),  $h_{final}$  作为 logistic 回归分类器的输入进行关系的预测,式(8):

$$h_{final} = FFNN([h_{sent}; h_{e_1}; \dots; h_{e_i}]) \quad (8)$$

## 2 实验数据及分析

### 2.1 数据集和参数设置

本文使用 TACRED 数据集,该数据集包含 41 种关系类型和一种特殊的关系类型,即“无关系”类型,其中训练集 106 264 条,验证集 22 631 条,测试集 15 509 条。

自然语言处理任务中准确率 (*precise*), 召回率 (*recall*),  $F_1$  值是模型的评价指标。训练过程中,为了使模型的效果更佳,需要对模型的参数进行调节,本文的模型中对于超参数的设置见表 1。

表 1 模型超参数设置

Tab. 1 Model hyperparameter setting

超参数名称	超参数
Word embedding	300
Dropout	0.5
Heads	3
GCN layer	2
Epoch	100
Bi-LSTM(h)	300
Bi-LSTM hidden layer	1
Lr-decay	0.9
Max_grad_norm	5

### 2.2 基线模型及评估方案

将本文模型与以下几种模型进行对比:逻辑回归分类器模型(LR),最短路径模型(SDP-LSTM),树结构神经网络模型(Tree-LSTM),均使用修剪树

过后的 GCN 模型和 C-GCN 模型,最先进的位置感知模型(PA-LSTM),注意力导向图卷积神经网络模型(AGGCN),实验结果见表 2。

表 2 各模型在 TACRED 数据集上的实验结果比较

Tab. 2 Comparison of experimental results of each model on TACRED data set

模型	准确率	召回率	$F_1$
LR	73.5	49.9	59.4
SDP-LSTM	66.3	52.7	58.7
Tree-LSTM	66.0	59.2	62.4
GCN	69.8	59.0	64.0
C-GCN	69.9	63.3	66.4
PA-LSTM	65.7	64.5	65.1
AGGCN	69.9	60.9	65.1
ASP-GCN(BERT)	73.4	64.9	68.9

由表 2 可知,回归分类器(LR)的准确率在所有模型中最高,但是召回率相对较低;与神经网络模型(SDP-LSTM),(Tree-LSTM)和(PA-LSTM)的准确率相比,ASP-GCN 模型的优势较为明显;以模型 SDP-LSTM 为例,ASP-GCN 模型的准确率得到了 10.2% 的提高,这主要得益于 GCN 对于图的特征提取能力;对于综合指标  $F_1$  值,ASP-GCN 模型明显优于同样使用 GCN 网络的 C-GCN 模型和 AGGCN 模型,说明本文模型中结合 BERT 预训练, Bi-LSTM 网络,最短路径为中心的修剪方法,多头注意力机制对模型的抽取性能是有帮助的。由此可见,本文提出的模型也是关系抽取任务中一个具备一定竞争力的模型。

### 2.3 消融实验

为了检验模型中各因素对关系抽取的影响,采用控制变量法进行实验。实验步骤如下:

- (1) 检验 BERT 预训练模型对本模型抽取性能的影响,主要与 glove 词嵌入模型进行比较;
- (2) 检验 Bi-LSTM 对本模型抽取性能的影响;
- (3) 检验最短路径修剪树方法对本模型抽取性能的影响;
- (4) 检验多头注意力机制对本模型抽取性能的影响。

4 种模型的抽取结果见表 3。由表 3 可知:

(1) 将 ASP-GCN 模型中 BERT 预训练模型换成 glove 模型后  $F_1$  值降低了 1.1%,这说明 BERT 强大的语义表征能力对于模型后续任务有很大的帮助;

(2) 删除 Bi-LSTM 层后,ASP-GCN 模型的  $F_1$  降低了 2.7%,这说明 Bi-LSTM 对文本特征进行初次提取有益于模型后续的处理;

(3) 删除最短路径为中心的剪枝方式后,ASP-

GCN模型的 $F_1$ 值降低了1.1%,说明筛选出文本中的无用信息,不仅可以提高模型的运行速率,还能提高模型的抽取性能;

(4)删除多头注意力时,模型 $F_1$ 值降低了2%,说明通过关注不同子空间的特征对模型的抽取任务有很大的作用。

表3 ASP-GCN模型的消融实验结果

模型	$P$	$R$	$F_1$
ASP-GCN(ours)	73.4	64.9	68.9
—BERT	72.3	63.9	67.8
—Bi-LSTM	66.7	65.7	66.2
—SDG	72.6	63.6	67.8
—Multi-Attention	71.9	62.5	66.9

## 2.4 注意力多头数分析

为了验证模型中多头注意力机制对模型结果的影响,实验分别对比了注意力头数 $N$ 的不同取值的不同模型效果,实验结果见表4所示,由表4可以看出,不同头数的注意力机制会对模型产生不同的影响,模型的最优值属于一定的范围,取值较小时,模型的特征提取能力较弱,取值较大时,可能会带来信息冗余,当 $N=3$ 时,模型效果相对较好。

表4 不同多头取值模型效果

$N$	1	2	3	4	5	6
Precision	72.3	72.5	73.4	71.3	70.5	70.3
Recall	64.7	64.7	64.9	63.9	63.4	63.9
$F_1$	68.3	68.4	68.9	67.4	66.7	66.9

## 2.5 以路径为中心的剪枝效果

为证明本文提出的剪枝方法的有效性,比较了GCN模型,Tree-LSTM模型和本文模型的剪枝距离 $L$ 变化对模型的影响。在TACRED数据集上分别采用 $L=\{0,1,2,3\}$ 时得到的结果如图6所示。

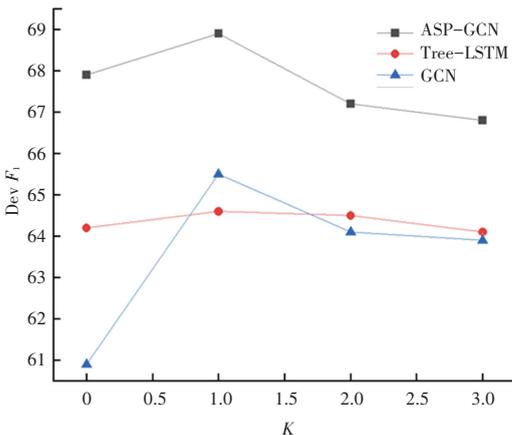


图6 以路径为中心的修剪距离( $K$ )

Fig. 6 Pruning distance centered on path ( $K$ )

当 $K=1$ 时,3个模型的性能都达到了峰值,优于各自模型的其他依赖路径模型。表明通过 $K=1$ 的路径为中心的修剪方式可以使模型得到改善。

## 3 结束语

本文提出了一种基于BERT和注意力引导图卷积网络的关系抽取模型,通过实验证实了在使用图卷积进行关系抽取的任务中,通过BERT预训练模型与Bi-LSTM结合对文本进行表征,加入最短路径修剪树,以及利用多头注意力机制等方式对于整个图卷积模型进行关系抽取任务来说是有益的,得到的结果均优于本文提及的其它基于深度学习的基线模型,从而证明了本文模型的合理性。

## 参考文献

- [1] 鄂海红, 张文静, 肖思琪, 等. 深度学习实体关系抽取研究综述[J]. 软件学报, 2019, 30(6): 1793-1818.
- [2] PENG N, POON H, QUIRKC, et al. Cross-sentence n-ary relation extraction with graph LSTMs [J]. Transactions of the Association for Computational Linguistics, 2017, 5: 101-115.
- [3] ZHANG Y, QI P, MANNING C D. Graph convolution over pruned dependency trees improves relation extraction [J]. arXiv preprint arXiv:1809.10185, 2018.
- [4] Zhang Y, Guo Z, Lu W. Attention Guided Graph Convolutional Networks for Relation Extraction [J]. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019:241-251.
- [5] TIAN Y, CHEN G, SONG Y, et al. Dependency-driven Relation Extraction with Attentive Graph Convolutional Networks [C]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2021:2331-2341.
- [6] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [J]. arXiv preprint arXiv:1810.04805, 2018:4171-4186.
- [7] MIKOLOV T, M KARAFIÁT, BURGET L, et al. Recurrent neural network based language model [C]// Interspeech, Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September. DBLP, 2015: 1388-1426.
- [8] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in neural information processing systems, 2017, 30:5998-6008.
- [9] MARCHEGGIANI D, TITOV I. Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling [J]. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017:1506-1515.
- [10] ZHANG Y, QI P, MANNING C D. Graph convolution over pruned dependency trees improves relation extraction [J]. arXiv preprint arXiv:1809.10185, 2018.