

双路注意力引导图卷积网络的关系抽取

李志欣¹, 孙亚茹¹, 唐素勤¹, 张灿龙¹, 马慧芳²

(1. 广西师范大学 广西多源信息挖掘与安全重点实验室, 广西桂林 541004;
2. 西北师范大学计算机科学与工程学院, 甘肃兰州 730070)

摘要: 为了更好地学习节点依赖并利用结构信息, 本文提出一种以完全依赖树作为直接输入的新方法, 利用图卷积网络并结合两个并行的注意力模块, 自主学习如何有选择地关注对关系抽取任务有用的信息. 该方法将样本表示成图上的各节点, 一个模块用于计算节点特征位置之间的影响, 使特征向量可以包含更广范围的语义信息, 另一个用于计算节点依赖的关系特征, 以增强节点间的全局依赖. 两个模块并行相互提升, 可以得到完整的特征表示. 在 TACRED 和 SemEval 数据集上的实验结果表明, 该方法能够更有效地获取对关系抽取任务有益的信息, 在各评价指标上取得了更好的性能.

关键词: 关系抽取; 图卷积网络; 注意力机制; 多跳关系推理

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2021)02-0315-09

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.12263/DZXB.20191105

Dual Attention Guided Graph Convolutional Networks for Relation Extraction

LI Zhi-xin¹, SUN Ya-ru¹, TANG Su-qin¹, ZHANG Can-long¹, MA Hui-fang²

(1. Guangxi Key Laboratory of Multi-source Information Mining and Security, Guangxi Normal University, Guilin, Guangxi 541004, China;
2. College of Computer Science and Engineering, Northwest Normal University, Lanzhou, Gansu 730070, China)

Abstract: To better learn node dependence and make use of structural information, this paper proposes a new method that takes the tree of complete dependence as the direct input. The method uses the graph convolutional network and combines two parallel attention modules to learn how to select the useful information. The method represents the samples as nodes on the graph. One module is used to compute the influence between positions of node features, which allows the feature vector to contain a wider range of semantic information. The other one is used to compute the relational features of node dependence, so as to enhance the global dependence between nodes. The two modules promote each other in parallel to obtain complete feature representation. The experimental results on the TACRED and SemEval datasets show that the method can obtain more useful information for relation extraction, thus achieve better performances on various evaluation metrics.

Key words: relation extraction; graph convolutional network; attention mechanism; multi-hop relational reasoning

1 引言

关系抽取用于检测文本中实体之间的关系, 在自然语言处理中扮演着重要的角色, 是构建知识问答以及知识图谱^[1]的基础, 也是信息抽取的重要支撑技术. 传统的模型主要关注实体识别和规则方法的研究. 大量的研究工作表明, 命名实体识别已经取得令人满意的结果, 但是基于规则方法的模型存在可移植性差的

问题^[2]. 近年来, 深度学习模型在获取实体间的关系上取得了一定的进展, 不需要依赖定义的规则就可以学习到句子中的结构信息.

现有的关系抽取模型可分为基于序列和基于依赖关系两类. 基于序列的模型^[3,4]是在单词序列上运行, 如利用循环神经网络对词进行编码获取句子信息. 基于依赖关系的模型^[5,6]是将句子关系依赖树纳入模型中, 有效利用依赖树结构信息对特征进行表征. 基于依

收稿日期: 2019-09-25; 修回日期: 2020-08-10; 责任编辑: 孙瑶

基金项目: 国家自然科学基金 (No. 61966004, No. 61663004, No. 61967002, No. 61866004, No. 61762078); 广西自然科学基金 (No. 2019GXNSFDA245018, No. 2018GXNSFDA281009, No. 2017GXNSFAA198365)

赖关系的模型可以捕获隐含的非局部句法关系,但是依赖树中所有的信息并不都对实体间的关系信息有利.为了进一步提高系统性能,会采取一些相应的修剪策略来提取依赖信息.文献[7]在关系抽取任务中只考虑树中实体间的最短依赖路径,很大程度上减少了数据的负担.为了更精确地保留有用的依赖信息,文献[8]只考虑实体的最近共同祖先(Lowest Common Ancestor, LCA)的子树.然而,使用这些修剪策略可能会消除树中一些重要的信息.图1给出了对一条文本进行关系抽取的实例.如图1(a)所示,对依赖树采用LCA策略修剪时,当最远依赖路径 K 为1时,实体Gwathmey与

Rosalie之间的隐含信息可能会缺失.为了不丢失重要的信息,并更好的利用树中的隐藏信息,一方面需要将整棵树纳入模型^[9],利用端到端的方式去学习实体间的关联性强度.因此,使模型学习如何在整颗树中保留和排除信息之间保持平衡是任务的关键.另一方面,树的规模可能会影响修剪策略的路径,实体之间会存在多跳关系推理的问题.如图1(b)所示,Gwathmey与Rosalie存在隐含关系,该关系依据Robert Gwathmey存在,若缺失,则隐含关系不成立.因此,多跳关系推理在多元关系抽取中是必不可少的,如何消除多跳对依赖路径的影响也是任务的关键.

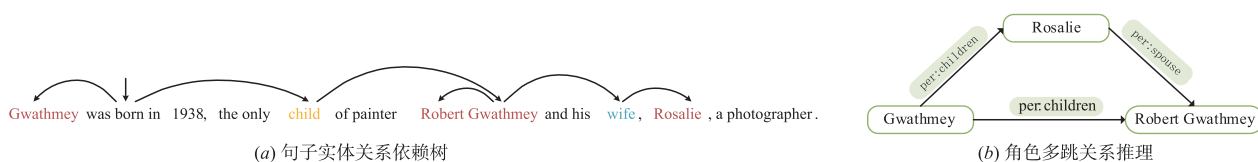


图1 对一条纯文本进行关系抽取

针对以上两个问题,本文提出了双路注意力引导图卷积网络的关系抽取模型(Dual Attention with Graph Convolutional Network, DAGCN).图卷积网络(Graph Convolutional Network, GCN)是一种有效解决多跳关系推理的方法^[10,11],将句子中的词表示成图中的节点,节点的特征依赖于其相邻的节点,在图结构上应用神经网络可以直接获取节点依赖信息,以缓解多跳对依赖路径的影响.利用自注意力机制^[12]不仅可以捕获文本中丰富的语义信息,还可以使模型学习节点之间的关联强度,从而更好地利用依赖树中的信息.因此,本方法利用GCN并结合注意力机制,在GCN尾部添加两个并行的注意力模块:位置注意力模块(Position Attention Module, PA)和关系注意力模块(Relation Attention Module, RA),分别在空间维度和关系维度上对语义依赖信息进行建模.位置注意力模块利用自注意力机制来捕获节点特征任意两个位置之间的空间依赖关系.对于某一位置的特征,通过加权求和将所有位置的特征聚集起来进行更新,其中权值由对应两个位置的特征相似性决定.也就是说,无论在空间维度上相距多远,具有相似特征的任意两个位置都可以相互促进.关系注意力模块利用注意力机制编码节点之间的依赖信息生成关系注意力矩阵,作为节点特征的权重系数,使特征向量的更新都与其相邻节点特征相关联,并用来增强全局依赖.两个注意力模块并行运行相互提升,将其输出的特征进行融合以得到更好的特征表征.实验结果表明,模型在TACRED和SemEval数据集上取得了令人满意的效果.与两个当前先进模型相比,该模型具有更好的性能,对于数据的扰动也更加鲁棒.

本文主要的贡献包括:

(1)提出了一种新的具有自注意力机制的双注意力网络,通过捕获广泛的上下文依赖信息,提高了关系抽取中特征的判别能力.

(2)提出了一个位置注意力模块来学习特征的空间相关性,通过编码更广泛范围的语义信息到局部特征中,以增强特征表征能力.

(3)提出了一个关系注意力模块来捕获节点间的依赖信息,通过编码节点依赖信息到关系特征中,增强节点之间的关系依赖.

(4)本文模型在TACRED和SemEval数据集上取得了良好的性能.

2 相关工作

传统的关系抽取方法过于依赖规则和自然语言处理工具,存在耗时、费力、精确率不高等问题.现有的研究表明,神经网络能够在一定程度上捕捉实体之间的关系,省时省力且精确率更高^[13,14].Zeng等人^[3]提出的模型最早将简单卷积神经网络应用于关系抽取任务,从词法层面提取句子特征,使得模型更有偏重性.随后,Zeng等人^[15]在简单卷积上使用分片最大池化对模型进行了扩展,通过将知识库与非结构化文本对齐来自动构建大量训练数据,增强模型的跨领域适应能力.该方法虽然使得模型对人工标注的依赖减少,但是存在假设过于肯定的缺点,且难免引入大量噪声数据.Nguyen等人^[16]提出一个多窗口版本模型,使用多尺寸卷积核提取文本中更多的卷积特征,并将词法特征用句法特征替代,降低了噪声数据的引入.为了使模型关注对任务有用的信息,Lin等人^[17]将注意力机制引入到关系抽取中,利用图神经网络将句子语义组合表示关

系,并引入注意力机制来尽量减轻样本中错误标签的影响. Le 等人^[18]将关系看作潜在变量,能够在没有任何监管信号的情况下诱导关系,同时以端到端的方式优化实体链接系统.除了信息表征对关系抽取的影响外,关系路径在关系抽取中也有着重要的作用. Zeng 等人^[19]通过中间实体在两个目标实体之间建立推理链,表明解决关系路径是提升关系抽取精确率的关键. Marcheggiani 等人^[20]通过联合训练修改图神经网络的上游表征来产生可以作为关系推理的隐式对象.

以上方法虽从数据层面和模型层面进行了模型优化,但是存在不能充分利用多个实体对之间的多跳依赖及其在句子中的依赖隐含关系信息的问题.为了解决这一结构化任务问题, Zhu 等人^[10]将图神经网络引入到关系抽取任务中,用节点表征实体,将关系与节点间的路径相关联,有效解决了多跳关系推理问题. De Cao 等人^[21]证明了图神经网络可用于构建知识库,通过连接预测因子分解模型丰富信息,使其在关系图的多个推理步骤上积累证据. Santoro 等人^[22]提出联合学习命名实体和关系的端到端模型将文本建模成关系图,利用关系加权来考虑实体间的交互.在对句子中的依赖隐含关系信息提取时,中间向量会存储上一阶段中的所有信息,随着句子长度的增加,该向量可能难以存储足够的信息,这便影响了信息的传递.注意力机制

就是为这个问题而设计的,侧重于传递信息,减少信息的存储.在将整棵树转换为邻接矩阵作为图神经网络的输入后, Guo 等人^[9]利用多头注意力将邻接矩阵学习表示成完全连接的边缘加权图进行关系抽取的方法取得了较好的效果.

3 模型

3.1 概述

DAGCN 模型主要由 3 个模块组成,即 GCN 模块、注意力模块和分类模块,如图 2 所示.

模型的输入为词节点的嵌入及节点的邻接矩阵,通过 GCN 对节点的隐藏信息进行表征,其结果通过一层卷积操作作为注意力模块的输入.卷积运算会导致局部感受域,对关系抽取任务会产生一些不好的影响.一方面,具有相同标签的节点所对应的特征可能会有一些差异.这导致了类内的一致性,影响了识别的精确性.另一方面,节点间的依赖信息可能会丢失.为此,本文通过注意力机制在特征之间建立关联来探索全局上下文信息,能够自适应地聚集远程上下文信息,提高节点特征的表示能力.与单一的注意力引导 GCN 获取上下文信息的方法^[9]不同,本文方法不仅关注节点间的依赖关系,而且关注节点表征的明确性,使节点在包含依赖关系的同时提高节点的识别率.

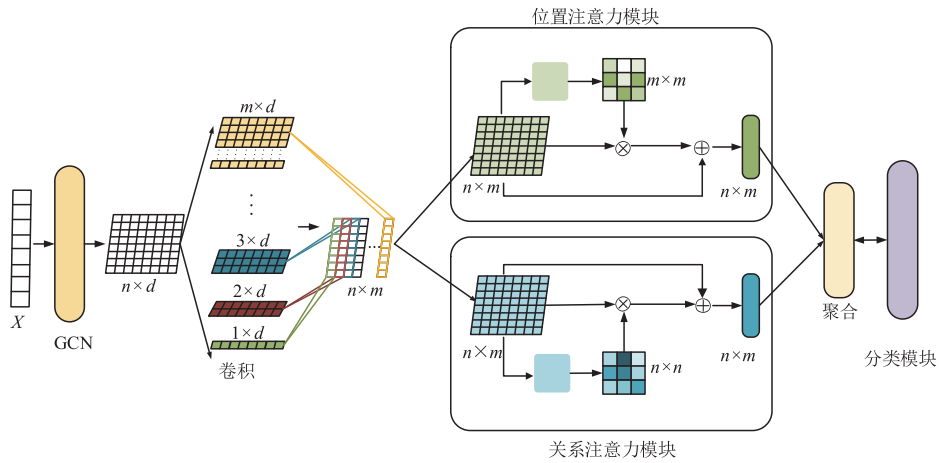


图2 DAGCN模型框架

3.2 GCN 模块

在关系抽取任务中,目的是预测出句子中实体间的关系.以 $\chi = [x_1, \dots, x_n]$ 表示一个句子,其中 x_i 表示第 i 个词. $\chi_e = [x_{e1}, \dots, x_{em}]$ 表示一个句子中的实体.给定 χ 和 χ_e ,则关系抽取的目的便是预测实体间的一个关系 $r \in \mathfrak{R}$ (关系集合).

GCN 通过邻接节点来表征本节点.例如,将句子中的词表示成图中的节点,若图中包含 n 个节点,则可以将该图表示成一个 $n \times n$ 的邻接矩阵 A . a_{ij} 和 a_{ji} 表示节

点 i 和 j 之间存在一条边.在 L 层的 GCN 中,给定输入集合 $\{h_i^0, \dots, h_n^0\}$,目的是输出 $\{h_i^L, \dots, h_n^L\}$.第 l 层的节点 i 的输出向量 h_i^l 由第 $l-1$ 层的节点 i 及其相邻节点表示,即

$$h_i^l = \sigma \left(\sum_{j=1}^n a_{ij} \mathbf{W}^{(l)} h_j^{(l-1)} + \mathbf{b}^{(l)} \right) \quad (1)$$

其中, $\mathbf{W}^{(l)}$ 是一个做线性变换的权重矩阵, $\mathbf{b}^{(l)}$ 是一个偏差向量. h_i^0 是初始输入的 x_i , $x_i \in \mathbb{R}^d$, d 是输入特征的维度.

经过 L 层的 GCN 对每个节点向量的处理,得到节点的隐藏表示,利用这些词表征可以得到一个句子的特征表示,即

$$\mathbf{h}_{\text{sent}} = f(\mathbf{H}^{(L)}) = f(\text{GCN}(\mathbf{H}^{(0)})) \quad (2)$$

其中, $\mathbf{H}^{(L)} \in \mathbb{R}^{n \times d}$ 表示 L 层所有的隐藏表示, L 是超参数, 函数 $f: \mathbb{R}^{n \times d} \rightarrow \mathbb{R}^d$ 将 n 个向量转变成一个句子向量. 同样的, 对第 i 个实体 \mathbf{h}_{ei} 计算如下:

$$\mathbf{h}_{ei} = f(\mathbf{H}_{ei}^{(L)}) \quad (3)$$

3.3 注意力模块

从图 2 可以看到, 本方法对 GCN 的最后一层进行改进, 加入一层卷积, 采用 m 个不同大小的滤波器进行卷积操作生成特征映射. 每个卷积层输出一个大小为 $n \times 1$ 的隐藏向量, 连接这些输出得到一个新的大小为 $n \times m$ 的矩阵, 然后将新的节点向量矩阵输入到两个并行的注意力模块中. 两个注意力模块通过在局部特征上计算全局上下文信息, 以获得更好的节点特征表示.

3.3.1 位置注意力模块

识别词特征的代表是关系抽取任务的关键. 然而, 许多工作表明传统的 GCN 只利用局部的节点信息可能会丢失文本中一些重要的信息, 从而导致对象被错误地分类. 为了在局部特征上建立丰富的上下文关系, 设计了位置注意力模块. 该模块将更广泛的上下文信息编码为局部特征, 从而增强了它们的表示能力. 如图 3 所示, 通过以下三个步骤生成空间上下文信息的新特征. 首先是生成位置注意力矩阵, 该矩阵对特征任意两个位置之间的空间关系进行建模. 接着, 对注意力矩阵和原始特征执行矩阵乘法. 最后, 将相乘后的矩阵与原始特征进行元素求和运算, 以获得全局上下文信息的最终表示.

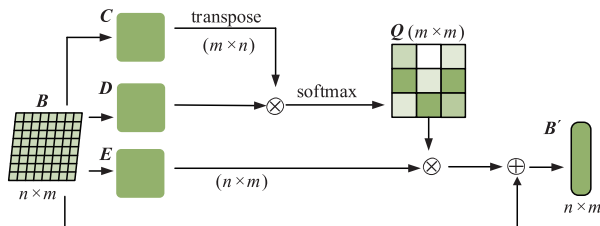


图3 位置注意力模块

自适应聚合空间上下文信息包括以下步骤. 首先利用卷积层获取的节点特征矩阵 $\mathbf{B} \in \mathbb{R}^{n \times m}$ 生成三个新的节点特征矩阵 \mathbf{C} , \mathbf{D} 和 \mathbf{E} , $\{\mathbf{C}, \mathbf{D}, \mathbf{E}\} \in \mathbb{R}^{n \times m}$. 接着, 对 \mathbf{C} 的转置和 \mathbf{D} 进行矩阵乘法操作, 利用一个 softmax 层计算出位置注意力图 $\mathbf{Q} \in \mathbb{R}^{m \times m}$, 即

$$q_{ij} = \frac{\exp(\mathbf{D}_j \cdot \mathbf{C}_i)}{\sum_{j=1}^m \exp(\mathbf{D}_j \cdot \mathbf{C}_i)} \quad (4)$$

其中, q_{ij} 表示第 j 个位置对第 i 个位置的影响. 两个位置

的特征越相似, 相关性就越大. 同时, 在 \mathbf{E} 和 \mathbf{Q} 之间进行矩阵乘法操作. 最后, 将结果乘以一个学习因子 α , 其初始值为 0, 并逐渐学会分配更多权重. 然后与原始特征逐元素累加得到最终的表示 \mathbf{B}' , 即

$$\mathbf{B}'_j = \alpha \sum_{i=1}^m (q_{ij} \mathbf{E}_i) + \mathbf{B}_j \quad (5)$$

由式(5)可知, 每个节点的最终的特征是所有节点的特征与原始特征的加权和. 因此, 它具有全局上下文信息, 并根据位置注意力矩阵有选择地聚合上下文. 相似的语义特征相互促进, 从而提高类内紧凑性和语义一致性.

3.3.2 关系注意力模块

节点之间的依赖信息对关系分类很重要, 是解决多跳推理的关键. 然而, 现有的工作表明利用修剪策略^[7,8]和传统的端到端模型会丢失文本中一些重要的依赖信息, 从而导致关系被错误分类. 为了有效学习节点依赖信息, 设计了关系注意力模块. 该模块强调相互依赖的特征映射, 可以改善节点语义特征的表示. 关系注意力模块将节点依赖信息编码为关系特征, 从而增强节点间的依赖关系. 如图 4 所示, 模块经过三个步骤生成节点关系依赖信息的新特征, 与位置注意力模块不同的是直接计算关系注意力矩阵 $\mathbf{P} \in \mathbb{R}^{n \times n}$. 关系注意力矩阵依据节点之间的依赖关系生成. 例如, 节点 i 与节点 j 的关系用 p_{ij} 或 p_{ji} 表示. 初始时, 两个有关联的节点的值为 1, 若没有关系则为 0. 然后利用自注意力机制生成关系特征值. 接着, 对注意力矩阵和原始特征执行矩阵乘法. 最后, 将相乘后的矩阵与原始特征进行元素求和运算, 以获得节点之间的全局依赖.

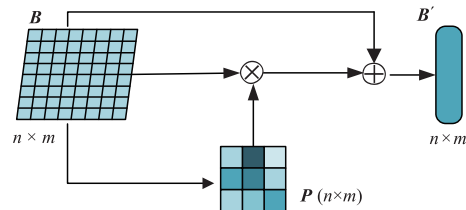


图4 关系注意力模块

具体的, 注意力矩阵 $\mathbf{P} \in \mathbb{R}^{n \times n}$ 通过一个 softmax 层计算得到, 即

$$p_{ij} = \frac{\exp(\mathbf{P}_i \cdot \mathbf{P}_j)}{\sum_{j=1}^n \exp(\mathbf{P}_i \cdot \mathbf{P}_j)} \quad (6)$$

因此 p_{ij} 表示节点 j 对节点 i 的关系影响, 两个节点的关系越紧密, 对这个值的影响越大. 接着, 将注意力矩阵与原始节点特征相乘. 最后, 将结果乘以一个学习因子 β , 并与原始特征逐元素累加得到最终的表示 \mathbf{B}' , 即

$$\mathbf{B}'_i = \beta \sum_{j=1}^n (p_{ij} \mathbf{B}_j) + \mathbf{B}_i \quad (7)$$

其中, β 从 0 逐渐学习, 得到的最终表示便包含了所有节点间的依赖信息, 从而提高了节点关系类别的判断性。

为了利用大范围的上下文信息, 最后将两个注意力模块得到的新特征进行聚合, 通过一层卷积得到最终的节点特征. 维度保持与原节点特征相同, 没有增加太多的参数, 但有效地增强了特征表示。

3.4 分类模块

首先对注意力模块输出的特征通过一层前馈神经网络^[17]输出关系特征表示, 实体 h_{e_i} 与实体 h_{e_j} 的关系 r_{ij} 可表示为

$$r_{ij} = \text{FFNN}(h_{e_i}, h_{e_j}, h_{\text{sent}}) \quad (8)$$

然后, 通过 softmax 函数对输出的关系特征做关系的概率预测, 即

$$P(r_{ij} | h_{e_i}, h_{e_j}, h_s) = \text{softmax}(\text{MLP}(r_{ij})) \quad (9)$$

其中, $\text{MLP}(\cdot)$ 表示一个多层感知机。

分类函数的损失可用交叉熵计算, 即

$$\mathcal{L} = \sum_{s \in \mathcal{S}} \sum_{i \neq j} \log P(r_{ij} | i, j, s) \quad (10)$$

其中, \mathcal{S} 表示句子集合, s 表示集合中的一个句子。

4 实验与分析

4.1 数据集

本工作主要在 TACRED 和 SemEval 两个数据集上对模型进行评估。

TACRED 数据集^[23]包含 TACKBP 挑战赛中提取的超过 106k 的实例, 定义了 41 种关系类型和一个特殊的无关系类. TACRED 中的提到是类型化的, 主题分为个人和组织, 对象分为 16 种细粒度类型 (例如日期和位置). 该数据集中 75050 个样本用于训练, 18660 个样本用于测试. 本文给出了在这个数据集上的 Micro-F1 值。

SemEval 2010 Task 8 数据集^[24]规模较小 (约为 TACRED 的 1/10), 但是在最近的关系抽取工作中得到了广泛的应用. 该数据集定义了 9 种关系类型和一个特殊的其它类. 每个样本都用两个给定实体间的关系进行注释. 该数据集中 8000 个样本用于训练, 2717 个样本用于测试. 本文遵循官方的任务设置, 给出了 Macro-F1 值。

4.2 实验设置

为了在 TACRED 数据集上进行公平的比较, 本文遵循文献[23]中使用的评估方案. 通过从 5 次独立运行中选择具有中间验证 F1 值的模型, 并报告其测试的 F1 值. 在 TACRED 数据集中的 25764 个样本作为验证集, 在 SemEval 2010 Task 8 训练集中选择 800 个样本作为验证集. GloVe 向量^[25]用作单词嵌入的初始化. 通过在验证集上的测试发现, $L=5$ 和 $d=300$ 的设置取得了

最好的效果. 模型是在 NVIDIA GeForce GTX 1050 下采用 CUDA9.0 训练. 训练周期为 100, 每 10 个周期用时约 270s. 在实验结果中给出了在 TACRED 数据集上的 Micro-F1 值, 和在 SemEval 2010 Task 8 数据集上的 Macro-F1 值。

本文模型的复杂性主要在于注意力模块中的位置注意力矩阵和关系注意力矩阵的计算. 在位置注意力模块中, 需要对输入矩阵做 1 次转置、2 次矩阵乘和 1 次加和操作. 复杂度为 $O(nm^2)$, n 为词的个数, m 为卷积核的个数. 在关系注意力模块中, 需要对输入矩阵做 1 次矩阵乘和 1 次加和操作. 复杂度为 $O(mn^2)$. 模型以整个句子作为输入, 因此 n 是固定的. m 相当于并行层数, m 越大, 模型的性能越好. 但是复杂度也会随之增加, 在不大于句子长度的前提下, 当 $m=15$ 时模型性能较好。

对于特征融合, 一般采用两种方式: 一是直接将两个特征进行拼接, 二是组合成复向量. 拼接会增加特征的维度, 而采用复合后的特征与原特征维度不变. 本文在考虑模型复杂度的情况下选择了第二种特征融合操作。

4.3 实验结果

本节的对比模型主要分为基于依赖的模型和基于序列的模型, 大多数模型基于 GCN 或长短期记忆网络 (Long Short-Term Memory, LSTM), 性能对比数据均来自于论文. 基于依赖的模型包括 LR (Logistic Regression)^[23], SDP-LSTM (Shortest Dependency Path LSTM)^[26], Tree-LSTM (Tree-structured LSTM)^[27], GCN 和 C-GCN (Contextualized GCN)^[6], AGGCN (Attention Guided GCN) 和 C-AGGCN (Contextualized AGGCN)^[9]. 对于基于序列的模型, 选择效果最好的 PA-LSTM (Position Aware LSTM)^[23]. 最后, 与 C-GCN 和 C-AGGCN 相似, C-DAGCN (Contextualized DAGCN) 使用双向 LSTM 扩展 DAGCN, 以捕获随后嵌入 DAGCN 层的上下文表示。

各模型在 TACRED 数据集上的评估结果如表 1 所示. 可见 LR 模型获得了最高的精确率. 由于基于特征的方法倾向于高频率的标签, 这种结果可能是数据不平衡的问题. 从表中可看出, 神经网络模型能够更好的平衡 P 值和 R 值. 因为本文模型与 GCN、C-GCN 和 AGGCN、C-AGGCN 相似, 且这两种模型取得了较好的结果, 所以这里主要与它们作对比. 从结果中可以看出, DAGCN 的 F1 值比 GCN 高 2.8%, 比 AGGCN 高 1.7%. C-DAGCN 的 F1 值比 C-GCN 高 4.2%, 比 C-AGGCN 高 1.6%, 主要原因是这两种模型缺乏对文本词序或歧义的信息. 与 GCN 性能差异明显是因为 GCN 基于修剪树. 因此, 在完整树上运行并结合注意力机制可以使模型更好地区分与工作相关和无关的信息。

各模型在 SemEval 数据集上的评估结果如表 2 所示. C-DAGCN 模型仍取得了最优的结果, Macro-F1 值比 C-GCN 高 2.1%, 比 C-AGGCN 高 1.2%.

表 1 各模型在 TACRED 数据集上的结果比较 (%)

模型	P	R	F1
LR ^[23]	73.5	49.9	59.4
SDP-LSTM ^[26]	66.3	52.7	58.7
Tree-LSTM ^[27]	66.0	59.2	62.4
PA-LSTM ^[23]	65.7	64.5	65.1
GCN ^[6]	69.8	59.0	64.0
C-GCN ^[6]	69.9	63.3	66.4
AGGCN ^[9]	69.9	60.9	65.1
C-AGGCN ^[9]	71.8	66.4	69.0
DAGCN	70.1	63.5	66.8
C-DAGCN	72.6	68.7	70.6

下面与没有添加注意力模块的模型作对比, 分别从节点语义关联强度和分类效果对节点进行网络可视化. 图 5 展示了两个例句的可视化结果, 左侧的两个图表示添加了注意力模块的句子节点关联可视化, 右侧

的两个图则表示未添加注意力模块的句子节点关联可视化. 图中的节点代表句子中的单词, 单词在句子中的类别不同颜色会不同, 权重不同大小会不同. 图中标注出的红色的节点表示句子中的实体, 实体在句子中的权重是最大的. 可以看出, 添加了注意力模块的模型使得节点间的语义关联性更强. 右侧图中会有少量节点没有与实体相关联, 而左侧图中的节点几乎全部相关联, 并且从节点间连接的距离也可发现左侧图的关联性更强.

表 2 各模型在 SemEval 数据集上的结果比较 (%)

模型	Macro-F1
SVM ^[28]	82.2
SDP-LSTM ^[26]	83.7
SPTree ^[8]	84.4
PA-LSTM ^[23]	82.7
C-GCN ^[6]	84.8
C-AGGCN ^[9]	85.7
C-DAGCN	86.9

Hwang, architect of the Pyongyang regime's ideology of *juche*, was once secretary of the ruling *workers'* Party and a tutor to current leader.

The government against the disaster and set up makeshift *houses*, ten million quake *survivors* moved into makeshift *houses*.

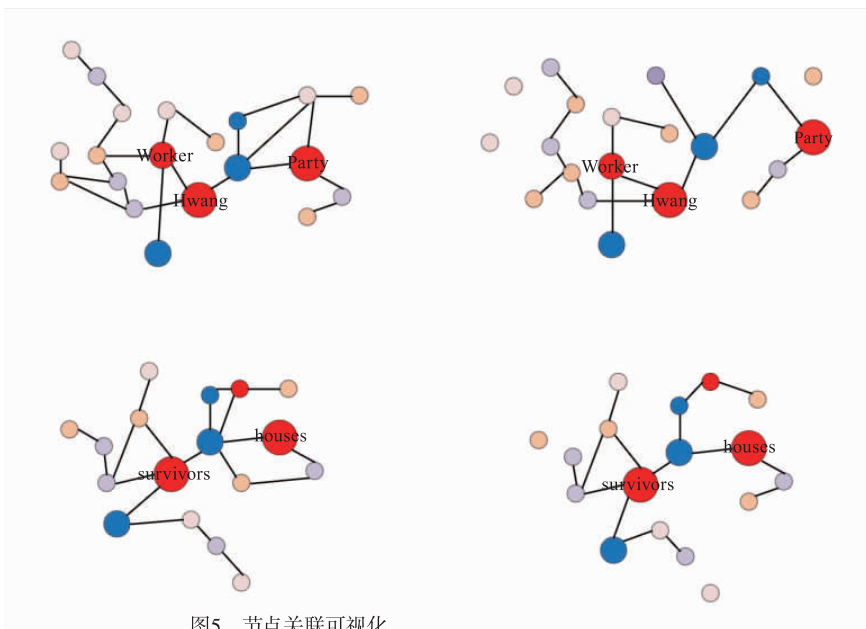


图5 节点关联可视化

对少量样本中的实体关系分类可视化, 每个样本中至少有一对实体对, 不同的颜色表示不同的关系, 如



图6 实体关系分类可视化

图 6 所示. 其中左图是没有添加注意力模块的模型所展示的结果, 从中可以看出实体关系分类存在混乱的现象. 而右图是添加了注意力模块的模型所展示的结果, 其分类边界显然更加清晰.

4.4 分析和讨论

4.4.1 消融研究

本节对模型各模块进行消融研究, 评估结果如表 3 所示. 从结果中发现, 增加注意力模块结果有明显的提高. 添加位置注意力模块提高 3.5%, 添加关系注

注意力模块提高 3.3%, 而两者作用在一起提高了 4.2%。而前馈神经网络对结果的影响不大, 去掉这一部分 F1 下降到 68.1%。去掉双向 LSTM 做词嵌入的部分, 也就是 DAGCN 的结果。当去掉 GCN 部分时, 模型下降的性能与两个并行的注意力模块相同。这说明并行的两个注意力模块扮演着重要的角色, 并且离不开 GCN 对文本信息的整合。可见, 注意力模块有助于 GCN 学习更好的信息聚合, 为图结构生成更好的表示。

表 3 DAGCN 模型的消融实验结果 (%)

模型	P	R	F1
C-DAGCN	72.6	68.7	70.6
— LSTM-embedding 层	70.1	63.5	66.8
— GCN 层	69.8	68.7	66.5
— h_e, h_{sent} , 前馈神经网络	70.5	65.9	68.1
— 位置注意力模块 PA	69.8	64.6	67.1
— 关系注意力模块 RA	70.1	64.7	67.3
— PA, RA	68.7	64.3	66.4

4.4.2 数据集规模对性能的影响

本节对数据集进行统一划分, 设置 5 种训练数据规模分别包含: 20%、40%、60%、80% 和 100% 的训练数据。对 C-GCN、C-AGGCN 及 C-DAGCN 在不同训练设置上进行性能评估, 如图 7 所示。发现随着训练数据规模的增大, 3 种模型性能的差距越明显。尤其在数据集规模达到 80% 时, C-DAGCN 的 F1 值几乎要高于 C-AGGCN 的最高值。以上分析表明, 本文的模型可以更有效的利用训练数据集。

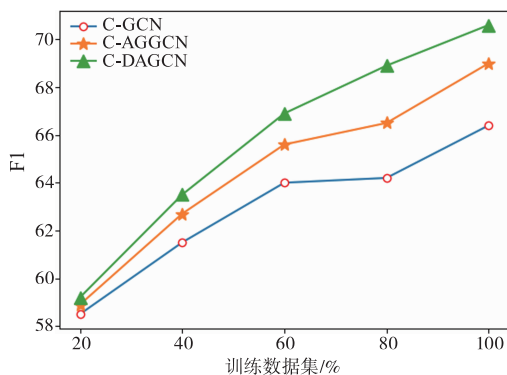


图7 不同训练数据规模下C-GCN、C-AGGCN和C-DAGCN的比较

4.4.3 句子长度对性能的影响

本节设置 5 种不同长度的句子: < 20 、 $[20, 30)$ 、 $[30, 40)$ 、 $[40, 50)$ 、 ≥ 50 , 对 C-GCN、C-AGGCN 及 C-DAGCN 在不同句子长度上进行性能评估, 如图 8 所示。其中所有的模型都基于完整树。发现 C-DAGCN 在不同设置上的性能完全优于另外两个模型。随着句子长度

的增加, 依赖关系图也会随着节点的增加而变大, 信息获取性能会下降, 但是结果表明 C-DAGCN 在更大的依赖关系图的情况下可以更好的获取有益的信息。

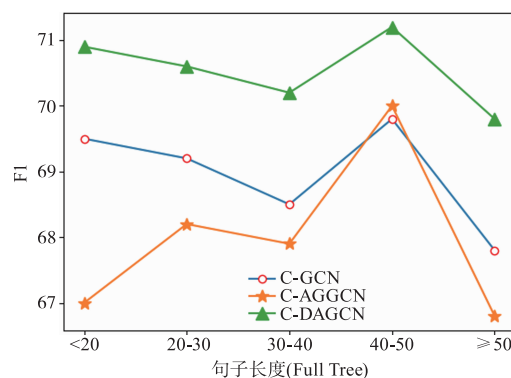


图8 不同句子长度下C-GCN、C-AGGCN和C-DAGCN的比较

4.4.4 不同类别下各模型的统计测试

本节对各模型在不同类别下的性能进行统计测试, 选择 SemEval 数据集的 9 种关系类别进行实验, 统计模型识别出正样本的概率。如表 4 所示, 对于关系清单中相近的关系, 本文模型要比 C-GCN 和 C-AGGCN 识别得更精确一些。例如, Component-Whole 和 Member-Collection 都是部分和整体的特殊情况, GCN 和 C-AGGCN 的表现旗鼓相当, 但本文模型的效果要明显优于这两种模型。对于句子关系所表达的状态是动态的还是静态的也可以很好地区分, 例如 Content-Container、Entity-Origin 和 Entity-Detination 关系。

表 4 SemEval 测试集上各类别精确率统计/%

类别	测试样本数	C-GCN ^[7]	C-AGGCN ^[10]	C-DAGCN
Cause-Effect	328	86.9	87.8	88.7
Instrument-Agency	156	87.2	87.2	90.4
Product-Producer	231	87.4	88.3	89.6
Content-Container	192	79.2	81.3	81.8
Entity-Origin	258	78.7	79.5	82.2
Entity-Detination	292	78.8	80.1	81.2
Component-Whole	312	81.4	80.1	83.6
Member-Collection	233	82.4	82.8	84.1
Message-Topic	261	88.1	87.7	88.5

5 结束语

本文提出了一种双路注意力引导图卷积网络的关系抽取模型, 通过两个并行的注意力模块来聚合特征全局语义信息和增强特征全局依赖。该模型在完整树上运行, 并学习以端到端的方式从树中获取有用的信息。结合图卷积网络可有效缓解多跳关系推理及不能

有效利用依赖树结构信息的难题;并行的两个注意力模块使模型可以学习更有利于关系抽取工作的信息.在 TACRED 和 SemEval 数据集上的实验表明该模型取得了良好的性能.未来的工作将考虑结合无监督学习推进知识抽取的进步.

参考文献

- [1] YU M, YIN W, HASAN K S, et al. Improved neural relation detection for knowledge base question answering[A]. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics [C]. Stroudsburg, USA: ACL, 2017. 571 – 581.
- [2] 冯建周, 宋沙沙, 王元卓, 等. 基于改进注意力机制的实体关系抽取方法[J]. 电子学报, 2019, 47(8): 1692 – 1700.
FENG Jian-zhou, SONG Sha-sha, WANG Yuan-zhuo, et al. Entity relation extraction based on improved attention mechanism[J]. Acta Electronica Sinica, 2019, 47(8): 1692 – 1700. (in Chinese)
- [3] ZENG D, LIU K, LAI S, et al. Relation classification via convolutional deep neural network[A]. Proceedings of the 25th International Conference on Computational Linguistics [C]. Stroudsburg, USA: ACL, 2014. 2335 – 2344.
- [4] WANG L, CAO Z, DE MELO G, et al. Relation classification via multi-level attention CNNs[A]. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics [C]. Stroudsburg, USA: ACL, 2016. 1298 – 1307.
- [5] PENG N, POON H, QUIRK C, et al. Cross-sentence n-ary relation extraction with graph LSTMs[J]. Transactions of the Association for Computational Linguistics, 2017, 5: 101 – 115.
- [6] ZHANG Y, QI P, MANNING C D. Graph convolution over pruned dependency trees improves relation extraction[A]. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing [C]. Stroudsburg, USA: ACL, 2018. 2205 – 2215.
- [7] XU K, FENG Y, HUANG S. Semantic relation classification via convolutional neural networks with simple negative sampling[A]. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing [C]. Stroudsburg, USA: ACL, 2015. 536 – 540.
- [8] MIWA M, BANSAL M. End-to-end relation extraction using LSTMs on sequences and tree structures[A]. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics [C]. Stroudsburg, USA: ACL, 2016. 21 – 29.
- [9] GUO Z, ZHANG Y, LU W. Attention guided graph convolutional networks for relation extraction[A]. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics [C]. Stroudsburg, USA: ACL, 2019. 241 – 251.
- [10] ZHU H, LIN Y, LIU Z, et al. Graph neural networks with generated parameters for relation extraction[A]. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics [C]. Stroudsburg, USA: ACL, 2019. 1331 – 1339.
- [11] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[A]. Proceedings of the 5th International Conference on Learning Representations [C]. La Jolla: ICLR, 2017. 1 – 14.
- [12] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[A]. Advances in Neural Information Processing Systems [C]. Cambridge, UK: MIT Press, 2017. 5998 – 6008.
- [13] 王亚坤, 黄河燕, 冯冲, 等. 基于级联模型的中文情感要素抽取[J]. 电子学报, 2016, 44(10): 2459 – 2465.
WANG Ya-shen, HUANG He-yan, FENG Chong, et al. Chinese evaluation element extraction based on cascaded model[J]. Acta Electronica Sinica, 2016, 44(10): 2459 – 2465. (in Chinese)
- [14] 权宇, 李志欣, 张灿龙, 等. 融合深度扩张网络和轻量化网络的目标检测模型[J]. 电子学报, 2020, 48(2): 390 – 397.
QUAN Yu, LI Zhi-xin, ZHANG Can-long, et al. Fusing deep dilated convolutions network and light-weight network for object detection[J]. Acta Electronica Sinica, 2020, 48(2): 390 – 397. (in Chinese)
- [15] ZENG D, LIU K, CHEN Y, et al. Distant supervision for relation extraction via piecewise convolutional neural networks[A]. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing [C]. Stroudsburg, USA: ACL, 2015. 1753 – 1762.
- [16] NGUYEN T H, GRISHMAN R. Relation extraction: Perspective from convolutional neural networks[A]. Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing [C]. Stroudsburg, USA: ACL, 2015. 39 – 48.
- [17] LIN Y, SHEN S, LIU Z, et al. Neural relation extraction with selective attention over instances[A]. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics [C]. Stroudsburg, USA: ACL, 2016. 2124 – 2133.
- [18] LE P, TITOV I. Improving entity linking by modeling latent relations between mentions[A]. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics [C]. Stroudsburg, USA: ACL, 2018. 1595 – 1604.

- [19] ZENG W, LIN Y, LIU Z, et al. Incorporating relation paths in neural relation extraction [A]. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing [C]. Stroudsburg, USA: ACL, 2017. 1768 – 1777.
- [20] MARCHEGGIANI D, TITOV I. Encoding sentences with graph convolutional networks for semantic role labeling [A]. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing [C]. Stroudsburg, USA: ACL, 2017. 1506 – 1515.
- [21] DE CAO N, AZIZ W, TITOV I. Question answering by reasoning across documents with graph convolutional networks [A]. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies [C]. Stroudsburg, USA: ACL, 2019. 2306 – 2317.
- [22] SANTORO A, RAPOSO D, BARRETT D G, et al. A simple neural network module for relational reasoning [A]. Advances in Neural Information Processing Systems [C]. Cambridge, UK: MIT Press, 2017. 4967 – 4976.
- [23] ZHANG Y, ZHONG V, CHEN D, et al. Position-aware attention and supervised data improve slot filling [A]. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing [C]. Stroudsburg, USA: ACL, 2017. 35 – 45.
- [24] HENDRICKX I, KIM S N, KOZAREVA Z, et al. Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals [A]. Proceedings of the 5th International Workshop on Semantic Evaluation [C]. Stroudsburg, USA: ACL, 2010. 33 – 38.
- [25] PENNINGTON J, SOCHER R, MANNING C D. GloVe: Global vectors for word representation [A]. Proceedings of the Conference on Empirical Methods in Natural Language Processing [C]. Stroudsburg, USA: ACL, 2014. 1532 – 1543.
- [26] XU Y, MOU L, LI G, et al. Classifying relations via long short term memory networks along shortest dependency paths [A]. Proceedings of the Conference on Empirical Methods in Natural Language Processing [C]. Stroudsburg, USA: ACL, 2015. 1785 – 1794.
- [27] TAI K S, SOCHER R, MANNING C D. Improved semantic representations from tree-structured long short-term memory networks [A]. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics [C]. Stroudsburg, USA: ACL, 2015. 1556 – 1566.
- [28] RINK B, HARABAGIU S M. UTD: Classifying semantic relations by combining lexical and semantic resources [A]. Proceedings of the 5th International Workshop on Semantic Evaluation [C]. Stroudsburg, USA: ACL, 2010. 256 – 259.

作者简介



李志欣 (通信作者) 男, 1971 年 10 月出生, 广西桂林人. 现为广西师范大学计算机科学与信息工程学院教授、博士生导师. 研究领域为图像理解、机器学习与跨媒体计算.
E-mail: lizx@gxnu.edu.cn



孙亚茹 女, 1993 年 10 月出生, 山东菏泽人. 广西师范大学计算机科学与信息工程学院硕士研究生. 研究方向为机器学习与自然语言处理.
E-mail: miniecho@163.com

唐素勤 女, 1972 年 12 月出生, 广西河池人. 现为广西师范大学教育学部教授. 研究领域为知识工程与自然语言处理.

张灿龙 男, 1975 年 10 月出生, 湖南娄底人. 现为广西师范大学计算机科学与信息工程学院教授. 研究领域为目标跟踪与模式识别.

马慧芳 女, 1981 年 7 月出生, 甘肃兰州人. 现为西北师范大学计算机科学与工程学院教授. 研究领域为数据挖掘与机器学习.