

基于句间信息的图注意力卷积网络的文档级关系抽取

段建勇, 杨潇, 王昊, 何丽, 李欣

引用本文

段建勇, 杨潇, 王昊, 何丽, 李欣 [基于句间信息的图注意力卷积网络的文档级关系抽取](#) [J]. 计算机科学, 2023, 50(6A): 220800189-6.

DUAN Jianyong, YANG Xiao, WANG Hao, HE Li, LI Xin. [Document-level Relation Extraction of Graph Attention Convolutional Network Based on Inter-sentence Information](#) [J]. Computer Science, 2023, 50(6A): 220800189-6.

相似文献推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于动态时空神经网络的城市交通流量预测方法](#)

City Traffic Flow Prediction Method Based on Dynamic Spatio-Temporal Neural Network
计算机科学, 2023, 50(6A): 220600266-7. <https://doi.org/10.11896/jsjcx.220600266>

[面向交通流量预测的时空Graph-CoordAttention网络](#)

Spatial-Temporal Graph-CoordAttention Network for Traffic Forecasting
计算机科学, 2023, 50(6A): 220200042-7. <https://doi.org/10.11896/jsjcx.220200042>

[基于多模态特征融合的时间序列异常检测](#)

Anomaly Detection of Time-series Based on Multi-modal Feature Fusion
计算机科学, 2023, 50(6A): 220700094-7. <https://doi.org/10.11896/jsjcx.220700094>

[联合人体姿态估计和多目标跟踪的跨数据集学习](#)

Cross-dataset Learning Combining Multi-object Tracking and Human Pose Estimation
计算机科学, 2023, 50(6A): 220400199-7. <https://doi.org/10.11896/jsjcx.220400199>

[基于改进Yolov4-tiny的轻量级目标检测算法](#)

Lightweight Target Detection Algorithm Based on Improved Yolov4-tiny
计算机科学, 2023, 50(6A): 220700006-7. <https://doi.org/10.11896/jsjcx.220700006>

基于句间信息的图注意力卷积网络的文档级关系抽取

段建勇^{1,2} 杨潇^{1,2} 王昊^{1,2,3} 何丽^{1,2} 李欣^{1,2}

1 北方工业大学信息学院 北京 100144

2 CNONIX 国家标准应用与推广实验室 北京 100144

3 北京城市治理研究基地 北京 100144

摘要 为解决现有模型对文档的结构信息挖掘不足的问题,提出一种基于句间信息的图注意力卷积网络模型。该模型改进了一种文档级编码器,该编码器使用了一种新的注意力机制——句间注意力机制,使得句子的最终表示更加关注前一个句子和之前文档中的重要信息,更有利于挖掘文档的结构信息。实验结果表明,所提模型在 DocRED 数据集上的 F_1 评价指标达到 56.3%,性能优于基线模型。在融入句间注意力机制时,由于模型需要对每一句话分别进行句间注意力操作,因此训练模型时需要消耗更多的内存和时间。基于句间信息的图注意力卷积网络模型可以有效地对文档中的相关信息进行聚合,并且增强对文档的结构信息的挖掘能力,从而使得模型在文档级关系抽取任务中效果得到提升。

关键词: 文档级关系抽取;注意力机制;文档级编码器;图卷积网络

中图法分类号 TP391

Document-level Relation Extraction of Graph Attention Convolutional Network Based on Inter-sentence Information

DUAN Jianyong^{1,2}, YANG Xiao^{1,2}, WANG Hao^{1,2,3}, HE Li^{1,2} and LI Xin^{1,2}

1 School of Information, North China University of Technology, Beijing 100144, China

2 CNONIX National Standard Application and Promotion Lab, Beijing 100144, China

3 Beijing Urban Governance Research Center, Beijing 100144, China

Abstract In order to solve the problem of insufficient structural information mining for documents in existing models, a graph attention convolution network model based on inter-sentence information is proposed. In this model, a document level encoder is improved, which uses a new attention mechanism, inter-sentence attention mechanism, so that the final representation of a sentence pays more attention to the important information in the previous sentence and the previous document, which is more conducive to mine the structural information of the document. Experiments show that the F_1 evaluation index of this model on DocRED data set reaches 56.3%, and its performance is better than that of the baseline model. When integrating the inter sentence attention mechanism, the model needs to perform inter-sentence attention operations for each sentence, so it needs to consume more memory and time when training the model. The graph attention convolution network model based on inter-sentence information can effectively aggregate the relevant information in the document, and enhance the mining ability of the document structure information, so that the model can improve the effect of the document level relationship extraction task.

Keywords Document-level relation extraction, Attention mechanism, Document-level encoder, Graph convolutional network

1 引言

随着信息时代的高速发展,互联网在给人们带来便利的同时也产生了数以万计的数据。为了应对信息爆炸带来的挑战,迫切需要一种自动内容抽取工具帮助人们快速从海量数据中挖掘出感兴趣的信息。在这种背景下,信息抽取应运而生^[1]。文档级关系抽取是信息抽取任务中的一个分支,主要目标是从自然语言文本中识别并判定实体对之间存在的特定关系。目前实体关系抽取的相关研究多集中在句子级别,

即只关注句内两个实体之间的关系,对跨句子的情况关注相对较少^[2]。而现实生活中,大多数的数据都是以文档(篇章)的形式存在的。随着远程监督在实体关系抽取任务中的有效应用,Quirk 等^[3]于 2017 年借助远程监督生成了生物领域的跨句子关系抽取数据集。文档级关系抽取任务得到了更多的关注。

文档级关系抽取需要整合文档多个句子内部和多个句子之间的信息,并捕获句子间实体之间的复杂关系。在图卷积网络 GCN^[4]流行之后,出现了许多基于依赖图的文档级关系

基金项目:国家自然科学基金(61972003);教育部人文社科基金项目(21YJA740052);北京哲社基地建设专项经费(110051360019XN142);北京市教育委员会科学研究计划项目(KM202210009002)

This work was supported by the National Natural Science Foundation of China(61972003), Humanities and Social Sciences Fund of the Ministry of Education(21YJA740052), Special Funds for the Construction of the Beijing Zheshe Base(110051360019XN142) and Beijing Municipal Commission of Education Scientific Research Program Project(KM202210009002).

通信作者:段建勇(duanjy@ncut.edu.cn)

抽取模型。然而,文档中相关信息的有效聚合和对文档结构的深入挖掘仍然是一个具有挑战性的研究问题^[5]。为了解决以上问题,本文提出了一种基于句间信息的图注意力卷积网络模型。

2 相关工作

关于文档级关系抽取的模型目前有两个方向:基于图卷积网络的文档级关系抽取模型和基于端到端的文档级关系抽取模型。在基于图卷积网络的文档级关系抽取模型的发展中,Gupta等^[6]于2019年提出iDepNN模型,分别基于最短依存路径(SDP)和子树的增广依存路径(ADP)两类特征进行建模。Nan等^[7]于2020年提出了LSR模型,相较于之前一些构建同质图然后做特征传播的模型,该模型构建了异质图并使用了GCN进行特征传播,设计了一种关系推理,构建了一个实体图用于关系推理。Zhou等^[8]于2021年提出的ATLOP模型重点关注对决定实体对的关系有用的上下文信息,在基于端到端的文档级关系抽取模型中效果显著。数据集方面,Yao等^[9]于2019年发布了DocRED数据集,同时给出了关系抽取的几个常用模型在该数据集上的表现。

现有的在基于图卷积网络的文档级关系抽取模型中的文档级编码器只是简单地将文档级的数据转化为句子级别,将句子级别的数据输入到对应的编码器中进行编码,之后再行拼接,以这样的方式很难进行跨句子之间的信息关联从而丢失一部分关键信息。因此,本文提出了一种基于句间信息的图注意力卷积网络模型。本文的主要贡献有:

- (1)提出了一个新的注意力机制,增强了句子之间的信息推理能力,提高了模型在整体任务上的性能。
- (2)提出了一种新的文档级编码器,通过调整文档编码的方式来增强句子之间的实体的关联。
- (3)使用图卷积网络进一步获取跨句子之间的信息。

3 基于句间信息的图注意力卷积网络模型

本节将介绍基于句间信息的图注意力卷积网络模型。模型由3部分组成:文档级编码器、加强推理层和分类层。其中文档级编码器又分为初始编码层和句间注意力机制。初始编码层首先对输入文档的每个句子进行编码,得到初始的上下文表示,之后再通过句间注意力机制得到最终表示。在加强推理层,我们使用图卷积网络,将对应于传感器中提及和实体提取设置为节点,并对其进行迭代细化。最后通过分类层分别计算实体对之间存在某种关系的概率。具体模型结构如图1所示。

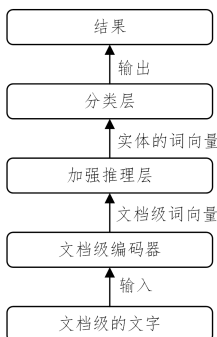


图1 文档级编码器结构图

Fig. 1 Structure diagram of document-level encoder

3.1 文档级编码器

文档级编码器的作用是将原本的文字信息转化为编码信息。输入是一篇文章,输出是对应的编码信息。模型中设计的文档级编码器的结构类似于句子级别的RNN模型^[10],当前句子的最终表示是由前一个句子和之前所有的文章内容共同决定时,考虑到文档中首次出现的实体在后文中存在共指的问题,因此该模型只采用单向的结构,来避免后文中的共指对于前文中实体表达的影响。文档级编码器整体结构如图2所示,其中的核心是句间注意力机制。

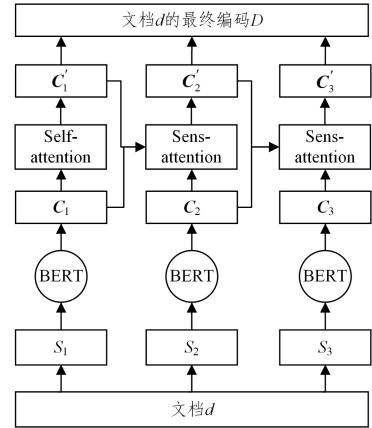


图2 文档级编码器结构图

Fig. 2 Structure diagram of document-level encoder

给定一个文档 d ,其中的每个句子 S_i 都被输入到初始编码器,得到每个句子的初始编码表示 C_i , i 表示第 i 个句子。初始编码器可以是RNN,LSTM^[11],BERT^[12]等传统的端到端序列模型。以BERT为例,将文本中的每一个句子转化为初始的词向量表示。得到每个句子的初始表示之后,通过句间注意力机制得到每个句子的最终表示。由于每篇文章的第一个句子 C_1 是没有上文信息的,因此句间注意力机制的第一个句子选择自注意力机制^[13],让第一个句子的最终表示 C_1' 更加关注自身重要的信息。 $C_i \in R^{a \times e}$, $C_i' \in R^{a \times e}$,其中 a 为第一个句子的长度。文档级编码器的编码过程如式(1)–式(3)所示:

$$d = \{S_1, S_2, \dots, S_n\} \quad (1)$$

$$C_n = BERT(S_n) \quad (2)$$

$$C_n' = \begin{cases} Self\text{-attention}(C_n), & n=1 \\ Sens\text{-attention}(C_n, C_{n-1}, C_{n-1}'), & n \neq 1 \end{cases} \quad (3)$$

本文提出了一种新的注意力机制——句间注意力机制,其结构如图3所示,句间注意力机制是让当前句子的表示关注到前一个句子以及之前所有句子的表示,这里句间注意力机制的输入有3个,分别是当前句子的初始表示 C_n 、前一个句子的初始表示 C_{n-1} 和前一个句子的最终表示 C_{n-1}' 。当前句子的初始表示 C_n 分别关注前一个句子的初始表示 C_{n-1} 和前一个句子的最终表示 C_{n-1}' ,得到两个分别关注前一个句子和之前所有文档的编码表示 A_{n5} 和 B_{n5} 。 A_{n5} 和 B_{n5} 会更加关注前文的重要信息,其效果如图4所示。在得到 A_{n5} 和 B_{n5} 之后,再与当前句子的初始表示 C_n 进行拼接,得到当前句子的最终表示 C_n' 。

下面进行具体说明:得到关注前一个句子编码表示 A_{n5} 的方法如式(4)–式(8)所示。

当前句子的初始表示 $C_n \in R^{b \times e}$,前一个句子的初始表示 $C_{n-1} \in R^{a \times e}$, a 和 b 为当前句子和前一个句子的长度, e 为

词嵌入的维度。将 C_{n-1} 转置并与 C_n 进行矩阵相乘。得到的 $A_{n1} \in R^{b \times a}$ 。矩阵相乘之后,再进行计算得到初始注意力得分 A_{n2} 。生成一个与 A_{n1} 同维度的 $mask$ 矩阵,并将原本的注意力得分矩阵进行遮盖,遮盖后得到 A_{n3} 。之前注意力得分矩阵 A_{n2} 是按照句子的最大维度进行存储的 $A_{n2} \in R^{r \times r}$, r 为最大句子长度,但实际句子长度肯定会比这个最大维度低,所以需要将原本的注意力得分矩阵 A_{n2} 利用 $mask$ 矩阵进行遮盖,得到实际的注意力得分矩阵 $A_{n3} \in R^{b \times b}$ 。最后将 A_{n3} 进行归一化得到最终的注意力得分矩阵 A_{n4} 。再将前一个句子的初始表示 C_{n-1} 与最终的注意力得分矩阵 A_{n4} 进行矩阵相乘得到关注前一个句子的当前句子的表示 A_{n5} 。

$$A_{n1} = \text{MatMul}(C_n, \text{Transpose}(C_{n-1})) \quad (4)$$

$$A_{n2} = \frac{A_{n1}}{\sqrt{e}} \quad (5)$$

$$A_{n3} = \text{Mask}(A_{n2}) \quad (6)$$

$$A_{n4} = \text{SoftMax}(A_{n3}) \quad (7)$$

$$A_{n5} = \text{MatMul}(A_{n4}, C_{n-1}) \quad (8)$$

由于前一个句子的初始表示 C_{n-1} 和最终表示 C'_{n-1} 是同一维度的,同理可得关注之前所有文章重点的当前句子的表示 B_{n5} 。最终将3个矩阵进行加权相加得到当前句子的最终表示 C'_n ,拼接方法如式(9)所示:

$$C'_n = C_n + \alpha * A_{n5} + \beta * B_{n5} \quad (9)$$

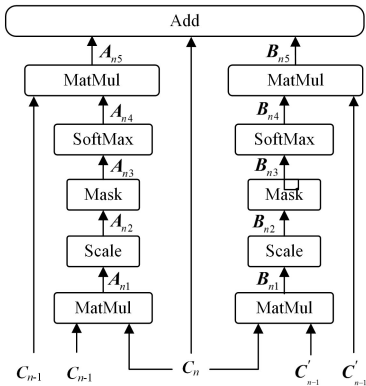


图3 句间注意力机制结构图

Fig. 3 Structure diagram of inter-sentence attention

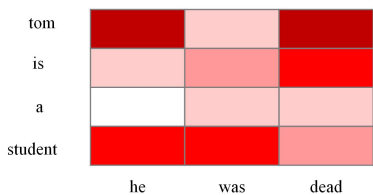


图4 句间注意力机制效果图

Fig. 4 Display of inter-sentence attention

3.2 加强推理层

加强推理层是使用图卷积网络(GCN)进一步进行跨句子之间的信息关联与推理。使用图卷积网络可以得到文档级别的图结构作为文档内在的结构信息。加强推理层通过构建迭代蕴含文档结构信息的图,最终得到可以计算出实体与实体之间关系的概率。加强推理层分为两个部分:构建层与迭代层。

3.2.1 构建层

构建层又分为两个部分:节点的构造层和边的生成层。

(1)节点的构造层

我们在表示文档结构的图中选取了3种不同的节点:提及及节点 $N_{mention}$ 、实体节点 N_{entity} 、句子节点 $N_{sentence}$ 。每个节点的编码表示都是通过计算不同元素编码表示的平均值。

1)提及节点

提及节点 $N_{mention}$ 对应的是输入文档中的所有实体的不同提及。提及节点的具体表示为提及所包含的单词 W 的平均值。定义如式(10)所示,其中 i 表示整篇文章一共有 i 个提及, j 表示在第 i 个提及中共包含有 j 个单词。

$$N_{mention}^{(i)} = \text{Average}(W_j) \quad (10)$$

2)实体节点

实体节点 N_{entity} 代表了文章中所有独特的实体。实体节点的具体表示为与实体关联的提及节点 $N_{mention}$ 表示的平均值。定义如式(11)所示,其中 i 表示整篇文章中实体的个数, j 表示一共有 j 个提及对应第 i 个实体。

$$N_{entity}^{(i)} = \text{Average}(N_{mention}^{(j)}) \quad (11)$$

3)句子节点

句子节点 $N_{sentence}$ 对应文档中的所有句子。句子节点的具体表示为这个句子中的所有单词 W 的平均值。定义如式(12)所示,其中 i 表示整篇文章一共有 i 个句子, j 表示第 i 个句子一共有 j 个单词。

$$N_{sentence}^{(i)} = \text{Average}(W_{ij}) \quad (12)$$

(2)边的生成层

本模型最终选择在图节点之间构造无向边,是否链接两个节点取决于文档元素之间的自然关联,即提及、实体和句子。节点之间的连接基于预定义的数据集或是预处理后的文档级关联数据。由于无法事先知道两个实体是否相关,因此本模型不会直接连接实体与实体的节点,而是通过构造和迭代后的图中各个节点之间边的关系,进行实体与实体之间的关系预测。下文将展示本模型预定义的5种边的类型,若两个节点之间不存在联系,则两个节点之间的边的数值默认为0。

1)提及节点 $N_{mention}$ 与提及节点 $N_{mention}$ 之间的边(以下简称提及提及边)

提及与提及之间也有可能具有一定联系,特别是两者位置比较接近时,因此预定义了提及提及边来保留这样的联系。判断是否生成提及提及边的办法定义如下:如果两个提及 i 和 j 在同一个句子 n 中出现,并且任意提及 i 或 j 出现在了本句子的上句或下句时,在这两个节点之间生成提及提及边。提及提及边的具体数值为同一个句子中两个节点的距离,如果存在多个距离选取最小的距离,其定义如式(13)所示。这样定义的好处是可以增强跨句子之间的实体与实体的联系。

$$\text{if } mention_i \in S_n \& mention_j \in S_n \& (mention_i \in S_{n\pm 1} \parallel mention_j \in S_{n\pm 1}), \text{ then } E_{mn} = \text{Min}(D_{ij}) \quad (13)$$

2)句子节点 $N_{sentence}$ 与提及节点 $N_{mention}$ 之间的边(以下简称句子提及边)

定义1 当句子 j 中出现某一提及 i 时,该提及与相应句子存在联系,此时在两个节点之间生成一条句子提及边。

句子提及边的具体数值为 1, 其定义如式(14)所示:

$$\text{if } mention_j \in S_j, \text{ then } E_{sm} = 1 \quad (14)$$

3) 实体节点 N_{entity} 与提及节点 $N_{mention}$ 之间的边(以下简称实体提及边)

定义 2 如果实体 i 可以与提及 j 相关联, 那么就在这两个节点之间生成一条实体提及边, 实体提及边的具体数值为 1, 其定义如式(15)所示:

$$\text{if } mention_j \in entity_{y_i}, \text{ then } E_{em} = 1 \quad (15)$$

4) 实体节点 N_{entity} 与句子节点 $N_{sentence}$ 之间的边(以下简称实体句子边)

定义 3 如果实体 i 中存在某一提及 j 出现在句子 k 中, 该实体与对应句子存在联系, 那么就在这两个节点之间生成一条实体句子边。实体句子边的具体数值为 1, 其定义如式(16)所示:

$$\text{if } mention_j \in entity_{y_i} \& mention_j \in S_k, \text{ then } E_{sn} = 1 \quad (16)$$

5) 句子节点 $N_{sentence}$ 与句子节点 $N_{sentence}$ 之间的边(以下简称句子句子边)

句子句子边是在两个句子节点之间直接生成一条句子句子边, 其具体数值为两个句子 i 和 j 之间的距离, 其定义如式(17)所示:

$$E_{ss} = D_{ij} \quad (17)$$

3.2.2 迭代层

给定 n 个节点构成的图, 生成对应邻接矩阵 $M_{ij} \in R^{k \times n \times n}$ 。 k 为生成边种类的数量。邻接矩阵 M_{ij} 是由上述生成的节点之间的边构成的矩阵。具体的迭代公式如式(18)所示:

$$l^i = \text{ReLU}(\sum_{j=1}^n M_{ij} w^i l^{i-1} + b^i) \quad (18)$$

其中, l^i 是第 i 层时 n 个节点的节点表示, l^0 表示 n 个节点的初始化节点表示; w^i 为权重矩阵, b^i 为偏移量。在每一层的迭代过程中都是将邻接矩阵 M_{ij} 分成 K 个矩阵分别迭代。

3.3 分类层

在经历了 n 次迭代之后, 我们得到了每个节点的最终表示。接下来我们计算每一组实体对在所有关系下的可能性, 如式(19)所示:

$$P(n | entity_{y_i}; entity_{y_j}) = \text{Sigmoid}(N_{entity}^{(i)} W N_{entity}^{(j)} + b) \quad (19)$$

其中, $N_{entity}^{(i)}$ 和 $N_{entity}^{(j)}$ 是经过迭代后的实体 i 和 j 的最终表示, W 与 b 是经过训练后的权重矩阵以及偏移量。

4 实验

4.1 数据集

实验使用的数据集为 CDR^[14] 和 2019 年由清华大学发布的 DocRED, 其中以 DocRED 为主数据集, 数据集的训练数据规模如表 1 所列。CDR 是生物领域的一个人工标注的一个数据集, 其任务是预测化学和疾病概念之间的二元相互作用。相比于 CDR, DocRED 涉及的领域更广, 同时有至少 40% 的相关事实需要对多个句子进行阅读和推理。DocRED 包含 3053 个训练数据、1000 个开发数据和 1000 个测试数据, 总共包含 96 种关系、132 375 个实体和 56 354 个相关事实。

表 1 实验所用数据集的数据规模

Table 1 Size of data set used in the experiment

	训练集	开发集	测试集
CDR	500	500	500
DocRED	3053	1000	1000

4.2 评价标准

本文采用 F_1 值作为最终的评价指标, 定义如式(20)一式(22)所示:

$$F_1 = \frac{2 * P * R}{P + R} \quad (20)$$

$$P = \frac{\text{预测正确的关系数}}{\text{预测的关系总数}} \quad (21)$$

$$R = \frac{\text{预测正确的关系数}}{\text{正确的关系总数}} \quad (22)$$

4.3 实验设置

实验基于 Python3.6 和深度学习框架 PyTorch1.4 实现, 在 GPU 上训练、验证并测试。实验分别使用 LSTM, GRU, BERT 作为初始编码器进行实验, LSTM 和 GRU 使用 GloVe 词嵌入。BERT 使用自带的词嵌入, 词嵌入的维度为 768 维。句间注意力机制中通过多次实验手工调节得到 α 取值为 1, β 取值为 0.15。针对不同的初始编码器有不同的参数, 具体参数如表 2 所列。

表 2 实验最佳效果参数

Table 2 Optimum experimental effect parameters

参数	值-LSTM	值-GRU	值-BERT
BATCH_SIZE	10	10	20
学习率	0.001	0.001	0.001
EPOCH	150	100	100
DECAY_RATE	0.96	0.955	0.97

4.4 实验基线

为验证模型在文档级关系抽取任务上的有效性, 将其与以下基线模型进行对比:

(1) 被广泛应用于自然语言处理任务的 3 个模型: LSTM, GRU^[15] 和 BERT。

(2) AGGCN^[16]: 一种以全依赖树作为输入的注意力引导图卷积网络模型, 使用多头注意力机制, 对 GCN 的邻接矩阵使用软剪枝策略。

(3) EoG^[17]: 在传统的 GCN 的模型中, 通过构造一种新的节点——边界点来完成整篇文章结构的更新。

(4) Multi-GCN^[18]: 针对文档级关系抽取, 提出的一种基于图注意力卷积卷积网络模型。该模型的结构也适用于 LSTM 和 BERT 两种模型, 同时该模型与本文模型的结构大致相似, 都使用了图神经网络和注意力机制。

4.5 实验结果与分析

4.5.1 模型主实验

模型主要在 DocRED 数据集上进行实验同时, 在 CDR 数据集上进行小规模实验, 实验结果如表 3 所列。

通过主实验, 可以得到以下结论:

(1) 通过对比序号 1, 2, 3, 8, 9, 10 的实验可以看出, DocRED 数据集上本文提出的文档级编码器在 LSTM, GRU 和 BERT 上分别提升了 2.16%, 1.74% 和 2.27%。在一定程度上验证了该文档级编码器具有兼容性, 可以很好地提升经典

模型在文档级关系抽取任务上的效果。

(2)通过对比序号 4,5,10 的实验,在一定程度上可以验证在文档级关系抽取任务上本模型相比其他基于图卷积网络的模型具有一定优势。

(3)通过对比序号 6,7,8,10 的实验可以看出,在以 LSTM 和 BERT 为初始编码器的模型中,该模型的效果比 Multi-GCN 的效果分别高出了 0.35% 和 0.87%,验证了相比于自注意力机制,句间注意力机制更能挖掘出跨句子之间的信息。

表 3 模型主实验结果

Table 3 Main experimental results of model
(单位:%)

序号	模型	DocRED	CDR
1	LSTM	50.57	58.43
2	GRU	51.93	59.02
3	BERT	54.03	60.57
4	AGGCN	52.47	—
5	EoG	52.15	63.60
6	LSTM+Multi-GCN	52.38	—
7	BERT+Multi-GCN	55.43	—
8	LSTM+Ours	52.73	60.05
9	GRU+Ours	53.67	61.24
10	BERT+Ours	56.30	64.15

4.5.2 消融实验

(1)文档级编码器中的自注意力机制的影响

在文档级编码器中,每篇文章的第一个句子 C_1 没有前一个句子的信息,因此第一个句子的最终表示 C_1 是否需要变化,成为了实验需要验证的一个问题,实验结果如表 4 所列。

表 4 消融实验 a

Table 4 Ablation experiment a

模型	DocRED
BERT +Ours(without Self)	56.06
BERT +Ours	56.30

通过实验,可以验证文档级编码器中的自注意力机制对于让第一个句子的最终表示 C_1' 更加关注自身的重要信息,同时对后面的句子的最终表示 C_n' 有一定影响。

(2)句间注意力机制中关注前一个句子重点的当前句子的表示 A_{n5} 和关注之前所有文章重点的当前句子的表示 B_{n5} 的影响

通过研究数据集,可以发现某些实体在上下句中存在着一定的关联,这类关联也出现在跨度比较大的句子之间。因此 A_{n5} 和 B_{n5} 权重分配问题也成为了实验需要验证的另一个问题,实验结果如表 5 所列。

表 5 消融实验 b

Table 5 Ablation experiment b

模型	DocRED
BERT +Ours(without A_{n5})	55.25
BERT +Ours(without B_{n5})	55.94
BERT +Ours	56.30

通过实验可以得出: A_{n5} 和 B_{n5} 在文档级关系抽取任务中都有一定的作用。与 B_{n5} 相比, A_{n5} 对于整体任务的性能影响更大。

(3)加强推理层的影响

本实验主要验证加强推理层对于整体模型的影响。

表 6 消融实验 c

Table 6 Ablation experiment c

模型	DocRED
BERT +Ours(without improve)	55.06
BERT +Ours	56.30

通过实验可以得出,加强推理层对于整体模型有一定的作用。

结束语 为了使文档中的相关信息更有效聚合,进一步对文档结构进行挖掘,本文提出了一种基于句间信息的图注意力卷积网络模型。在基于图卷积神经网络的关系抽取模型中,使用了一种新的注意力机制——句间注意力机制,用于更好地捕获上下文的关键信息。改进了一种文档级编码器,使得新的编码更有利于文档结构信息的挖掘。实验表明,所提出的句间信息的图注意力卷积网络模型性能优于基线模型。不仅是文档的表达对于基于图卷积神经网络的关系抽取模型有一定的作用,图的构建对其也有着一定影响。因此在未来的工作中,也可以通过改变图的结构来提升模型性能。

参考文献

- [1] CHE W X, LIU T, LI S. Automatic extraction of entity relationship [J]. Chinese Journal of Information, 2005, 19(2): 2-7.
- [2] HUANG Q Z, ZHU S Q, FENG Y S, et al. Three sentences are all you need: Local path enhanced document relation extraction [C]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). 2021.
- [3] QUIRK C, POON H. Distant supervision for relation extraction beyond the sentence boundary [C] // Proceedings of EACL. 2017.
- [4] KIPF T N, WELING M. Semisupervised classification with graph convolutional networks [C]// Proceedings of ICLR. 2017.
- [5] HUANG K, QI P, WANG G T, et al. Entity and evidence guided document-level relation extraction [C]// Proceedings of the 6th Workshop on Representation Learning for NLP. 2021: 140-149.
- [6] GUPTA P, RAJARAM S, SCHÜTZE H, et al. Neural relation extraction within and across sentence boundaries [C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2019: 6513-6520.
- [7] NAN G, GUO Z, SEKULIC I, et al. Reasoning with Latent Structure Refinement for Document-Level Relation Extraction [C]// ACL. 2020: 1073-1087.
- [8] ZHOU W X, HUANG K, MA T Y, et al. Document-level relation extraction with adaptive thresholding and localized context pooling [C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2021: 1631-1640.
- [9] YAO Y, YE D M, LI P, et al. DocRED: A large-scale document-level relation extraction dataset [C] // Proceedings of ACL. 2019: 189-198.
- [10] TODERICI G, O'MALLEY S M, HWANG S J, et al. Recurrent

- Neural Network Regularization[J]. arXiv:1409.2329,2015.
- [11] ZHOU P, SHI W, TIAN J, et al. Attention-based bidirectional long short-term memory networks for relation classification [C]//Proceedings of ACL. 2016:1412-1421.
- [12] DEVLIN J, WEI M, CHANG K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [J]. arXiv:1810.04805v2,2019.
- [13] VASWANI A, SHAZEER N, PARMAR N. Attention Is All You Need[J]. arXiv:1706.03762v5,2017.
- [14] GU J H, SUN F Q, QIAN L H, et al. Database: The Journal of Biological Databases and Curation[J]. The Journal of Biological Databases and Curation,2017:2049-2058.
- [15] CHUNG J Y, GULCEHRE C, CHO K H. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling[J]. arXiv:1412.3555,2014.
- [16] GUO Z J, ZHANG Y, LU W. Attention Guided Graph Convolutional Networks for Relation Extraction [C] // Proceedings of ACL. 2019:241-251.
- [17] CHRISTOPOULOU F, MIWA M, ANANIADOU S. Connecting the dots: Document-level neural relation extraction with edge-oriented graphs [C] // Proceedings of EMNLP. 2019:1532-1543.
- [18] WU T, KONG F. Document level relation extraction based on graph attention convolution neural network [J]. Chinese Journal of information, 2021, 35(10):73-80.



DUAN Jianyong, born in 1978, Ph. D, professor, is a member of China Computer Federation. His main research interests include natural language processing and information retrieval.