

融合关系传递信息的双图文档级关系抽取方法

寇嘉颖, 赵卫东, 柳先辉

引用本文

寇嘉颖, 赵卫东, 柳先辉. 融合关系传递信息的双图文档级关系抽取方法[J]. 计算机科学, 2023, 50(12): 229-235.

KOU Jiaying, ZHAO Weidong, LIU Xianhui. [Method of Document Level Relation Extraction Based on Fusion of Relational Transfer Information Using Double Graph](#) [J]. Computer Science, 2023, 50(12): 229-235.

相似文章推荐 (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)

[基于动态负采样的图卷积协同过滤推荐模型](#)

Dynamic Negative Sampling for Graph Convolution Network Based Collaborative Filtering Recommendation Model

计算机科学, 2023, 50(11A): 230200149-7. <https://doi.org/10.11896/jsjcx.230200149>

[结合图注意力机制的知识图谱推荐算法](#)

Knowledge Graph Recommendation Algorithm Combined with Graph Attention Mechanism

计算机科学, 2023, 50(11A): 230100057-7. <https://doi.org/10.11896/jsjcx.230100057>

[基于数据库表的微服务拆分方法](#)

Microservice Splitting Approach Based on Database Table

计算机科学, 2023, 50(11A): 230200102-7. <https://doi.org/10.11896/jsjcx.230200102>

[图像美学信息增强的视觉感知推荐系统](#)

Image Aesthetics-enhanced Visual Perception Recommendation System

计算机科学, 2023, 50(11A): 221100083-8. <https://doi.org/10.11896/jsjcx.221100083>

[基于深度学习的羽毛球知识图谱补全模型构建](#)

Construction of Badminton Knowledge Graph Completion Model Based on Deep Learning

计算机科学, 2023, 50(11A): 220900205-6. <https://doi.org/10.11896/jsjcx.220900205>

融合关系传递信息的双图文档级关系抽取方法

寇嘉颖 赵卫东 柳先辉

同济大学电子与信息工程学院 上海 201804

(jiayingk@tongji.edu.cn)

摘要 文档级关系抽取指在长段落的非结构性文本中抽取实体以及实体之间的关系。相较于传统的句子级关系抽取,文档级关系抽取需要融合多个句子的上下文信息,并且加以逻辑推理,才能抽取关系三元组。针对目前文档级关系抽取方法中存在的文档语义信息建模不够完整且抽取效果具有局限性等问题,提出了一种融合关系传递信息的双图文档级关系抽取方法。利用关系信息传递性将不同句子中提及之间的交互信息引入路径构造中,加以使用同句子中提及的交互信息以及提及之间的共指信息,构建提及节点间的路径集合,提高文档建模的完整性;应用路径集合和提及节点搭建提及层次的图聚合网络,建立文档语义信息模型;经过图卷积网络的信息迭代后,将相同实体的不同提及节点的信息进行融合,形成实体节点,构成实体层次的图推理网络;最终根据实体图节点间的路径信息进行逻辑推理,抽取实体间的关系。在公开数据集 DocRED 上的实验结果表明,相对于基准模型,所提模型的 F1 值有 1.2 的提升,证明了该方法的有效性。

关键词 文档级关系抽取;关系传递;图卷积神经网络;逻辑推理

中图法分类号 TP391

Method of Document Level Relation Extraction Based on Fusion of Relational Transfer Information Using Double Graph

KOU Jiaying, ZHAO Weidong and LIU Xianhui

College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China

Abstract Document-level relation extraction refers to the extraction of entities and their relations from long paragraphs of unstructured text. Compared to traditional sentence-level relation extraction, document-level relation extraction requires the integration of contextual information from multiple sentences and logical reasoning to extract relation triples. In response to the current limitations of document-level relation extraction methods, such as incomplete modeling of document semantic information and limited extraction effects, a double-graph document-level relation extraction method that integrates relational transfer information is proposed. Interactions mentioned between different sentences are introduced into the path construction through the transitivity of relational information, and the interaction information mentioned in the same sentence as well as the coreference information between mentions are used to construct the path set between mention nodes, so as to improve the completeness of document modeling. A mention hierarchy graph aggregation network is constructed using the path set and mention nodes, and a document semantic information model is established. After the information iteration of the graph convolutional network (GCN), the information of different mention nodes of the same entity is fused to form entity node, which constitutes an entity-level graph reasoning network. Finally, logical inference is performed based on the path information between entity graph nodes to extract the relation between entities. The proposed model is experimented on the public dataset DocRED (document-level relation extraction dataset), and the experimental results show a improvement of 1.2 (F1) compared to the baseline model, which proves the effectiveness of the proposed method.

Keywords Document relation extraction, Relational transfer, Graph convolutional network, Logical reasoning

1 引言

关系抽取 (Relation Extraction, RE)^[1] 是信息抽取中的一项重要任务,其目的在于提取文本中的实体对之间的关系。由于

该任务需要模型在构建时很好地理解上下文的语义,因此它也是信息抽取中最关键和最困难的一部分^[2]。早期的关系抽取重点在于提取单个句子中的实体对^[3-4],但据人工标注的数据的统计表明,至少 40% 的实体关系事实必须从多个句子中联合

到稿日期:2023-05-04 返修日期:2023-08-30

基金项目:国家重点研发计划(2020YFB1709303)

This work was supported by the National Key Research and Development Program of China(2020YFB1709303).

通信作者:柳先辉(xianhui_l@163.com)

得出。因此,文档级关系抽取成为目前重点研究的方向^[5]。

文档级关系抽取^[6]任务相对于句子级关系抽取更为复杂^[7-8],图1给出了文档级关系抽取实例,其难点主要体现在以下3个方面:

(1)同一实体可能会在多个句子中出现,且上下文之间会出现实体的自然指代。例如,“The Forbidden City”分别出现在句子[1]和句子[6]中,而句子[2]中的“The palace”其实也指代它。因此,这需要利用更多的上下文信息,实体表示才更加全面。

(2)一种关系的头实体和尾实体可能位于不同的句子中,需要多个句子才能抽取实体之间的关系。例如,“The Imperial Garden”和“China”之间是存在关系,但是需要句子[1]和句子[6]作为支撑论据,才能推理得到。

(3)许多的实体关系事实,往往需要通过复杂的逻辑推理和常识知识推理才能得到。比如“the Forbidden City”本身和“The Dongcheng District”并没有关联,但由于“The palace”与“The Dongcheng District”存在关系且它为“The Forbidden City”的指代,因此需要通过一定的逻辑推理才可以抽取出图中所示的关系三元组。

[1] The Forbidden City is a palace complex in Beijing, China. [2] The palace is situated in the Dongcheng District. ... [6] As Beijing's most popular tourist attraction, the Imperial Garden within the Forbidden City is beloved by visitors.			
Object	Subject	Relation	Evidence
Beijing	China	Country	[1]
the Forbidden City	the Dongcheng District	Location	[1][2]
the Dongcheng District	Beijing	City	[1][2]
the Imperial Garden	China	Country	[1][6]

图1 文档级关系抽取任务实例

Fig. 1 Instance of document-level relation extraction task

当前主流的文档级关系抽取方法可以分为基于序列^[9]的方法和基于图的方法。基于序列的方法是将文档看作一系列存在关系的数据,利用深度学习方法对上下文的信息进行捕捉,从而抽取实体间的关系。该方法模型相对简单、易于应用,但面对较为复杂的文档时,由于上下文信息之间的逻辑推理不足,因此会表现出实体关系抽取不全、不准确等局限性。而基于图的方法是利用实体及其提及构建图结构,通过图传播的方式显式学习实体之间的关联。该方法可以在整个文档范围内建立实体之间的联系,并且增强上下文信息的推理能力^[10],从而更好地实现文档级关系抽取。

因此,本文提出了一种融合关系传递信息的双图文档级关系抽取方法 RTDG (Relational Transfer Information Using Double Graph),通过将不同句子中提及之间的交互信息引入路径构造中,以提高文档建模的完整性;同时利用提及层次图和实体层次图的双图结构,分层次建立文档语义信息模型;最后对实体关系进行推理、分类,完成文章级关系抽取任务。

2 相关工作

目前,已有不少学者在基于文档图方法的研究上取得了进展^[11-12]。Sahu 等^[13]于2019年提出了一种使用图卷积神经网络进行文档级关系抽取的方法 GCNN (Graph Convolu-

tion Neural Network)。他们将文档中的单词作为节点,将句内和句间信息之间存在的依存关系、邻接关系、共指关系等作为连接节点的边,然后使用带标签的 GCN 单独学习不同类型边的图,并借助 bi-affine 模型对文章中出现的所有单词对进行打分,从而预测关系类别。该方法虽然可以捕捉不同类型边的特征,但边的类型过多,且需要对文档中的所有单词对进行打分,网络结构存在较多冗余部分,模型较为复杂,直接影响到关系抽取的效率和效果。

同年,Christopoulou 等认为,对于抽取关系而言,边的表示比节点的表示更为重要,因此提出了 EoG (Edge-oriented Graphs)^[14]模型的概念。该模型使用启发式的方法,将节点分为提及、实体和句子3种类型,并且根据节点种类的不同,对应建立不同的边,包括提及到提及、提及到实体、提及到句子、实体到句子和句子到句子这5种类型。该方法主要训练连接两个实体之间的路径,通过多次迭代加权该路径上的信息,进行逻辑推理,最终生成实体到实体的边,实现关系的抽取。但是,由于 EoG 中提及之间的交互方式不够有效,没有很好地发挥共指信息的作用,且伴随着实体的句间距离增大,模型的性能出现了明显的下降,因此需要进一步改进。

为了解决 EoG 模型节点交互方式不佳且节点全连接会带来大量冗余信息的问题,Nan 等提出了 LSR (Latent Structure Refinement) 模型^[15]。该模型对文档中所有句子进行编码,构造实体、提及和元依赖路径 (Meta Dependency Paths, MDP) 节点,将图结构作为变量输入,通过邻居节点的隐状态自动学习知识,更新节点表示,以端到端的方式进行迭代推理。该方法可以利用 MDP 节点过滤掉更多无关信息,发现更为准确的句间关系。

同样是继承 EoG 模型,Zeng 等提出了一种具有双图特征的图聚合推理网络 (Graph Aggregation-and-Inference Network, GAIN)^[16]。该模型先在提及层次的图上使用 GCN 进行信息迭代,之后将所有提及级别的节点进行加权融合,并对它们之间的连接路径进行压缩,形成实体层次的图,最终由 MLP (Multilayer Perceptron) 网络进行实体关系分类。该方法创新性地将文档建模与关系推理分隔开,提升了对文档级信息的推理能力。但是,该方法在路径构造中仅考虑了同句中提及之间的交互信息和文档内提及的共指信息,不同句子中因关系的传递性所带来的提及之间的交互信息却被忽略,然而这部分信息往往隐含着较为丰富的提及联系。因此导致了文档的显式建模不够完整,从而在实体层次图网络进行逻辑推理后,会出现关系抽取不准确的现象。

针对上述问题,本文探讨了一种路径构造改进方法,通过融合关系传递的信息,将不同句子中提及之间的交互信息引入路径构造中。这样既可以在一定程度上提高文档建模的完整性,又避免了节点全连接带来的大量冗余信息,对于进一步提升文档级关系抽取的效果具有重要意义。

3 融合关系传递信息的双图网络模型

本文提出的 RTDG 模型的架构如图2所示,以文档作为输入,其主要包括编码模块、融合关系传递信息的路径构造模块、提及层次图聚合网络模块、实体层次图推理网络模块和关系分类模块。

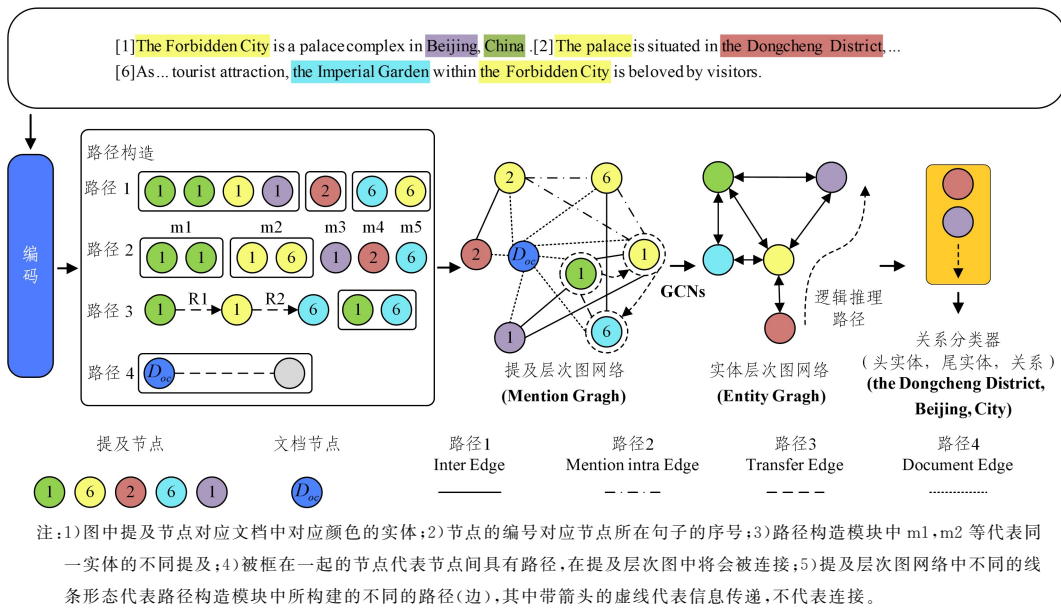


图2 融合关系传递信息的双图文档级关系抽取模型架构示意图(电子版为彩图)

Fig. 2 Architecture diagram of a document-level relation extraction model based on fusion of relational transfer information using double graphs

3.1 编码模块

编码模块将文档文本转化为具有语义信息的向量序列。

在转化之前,对文档级关系抽取问题进行如下定义:

$$A = \{s_i\}_{i=1}^N \quad (1)$$

$$s = \{\tau w_i\}_{i=1}^n \quad (2)$$

$$E = \{e_j\}_{j=1}^P \quad (3)$$

$$e_i = \{m_k\}_{k=1}^Q \quad (4)$$

式(1)表示文档 A 包含 N 个句子 s ,式(2)表示句子 s 中包含 n 个单词 τw ,式(3)表示实体集合 E 包含 P 个实体 e ,式(4)表示某一实体 e_i 在文档中拥有 Q 个具体表达定义(即提及 m)。文档级关系抽取问题就是要在文档 A 中抽取实体关系三元组 $\{(e_i, r_{ij}, e_j) | e_i, e_j \in E, r_{ij} \in R\}$,其中关系 r_{ij} 为预先定义好的关系集合 R 中的关系。

在编码模块中,每个单词的语义信息表示 x_i 由词嵌入编码 $E_w(w_i)$ 、实体类型嵌入编码 $E_t(t_i)$ 和共指关系嵌入编码 $E_c(c_i)$ 组成,如式(5)所示。其中 t_i 和 c_i 分别代表实体类型和实体 ID。对于不属于任何实体的单词,其实体类型为“None”。

$$x_i = [E_w(w_i); E_t(t_i); E_c(c_i)] \quad (5)$$

接下来,将量化的单词语义信息表示送入 BiLSTM(Bi-directional Long Short-Term Memory)^[17] 编码器,即可获得每个单词的上下文语义信息表示 g_i ,如式(6)所示:

$$[g_1, \dots, g_n] = \text{Encoder}([x_1, \dots, x_n]) \quad (6)$$

最终,文档 A 被成功转化为一系列具有语义信息的向量序列 $\{g_i\}_{i=1}^n$ 。

3.2 融合关系传递信息的路径构造模块

在文档级关系抽取任务中,有效的路径构造方法对构建一个能够最大化表示文档语义信息的图网络^[18] 十分重要,它直接影响到关系抽取的准确性^[19]。RTDG 模型在路径构造中,以提及为基本单位,不仅考虑了同句中提及之间的交互信息和文档内提及的共指信息,而且融合了关系传递信息,将不同句子中因关系传递性所带来的提及之间的交互信息也用于

路径构造中,达到最大化表示文档语义信息的效果。如图 3 所示,本文将提及间构造的路径分为“Inter Edge”“Mention Inter Edge”“Transfer Edge”和“Document Edge”4 种类型。

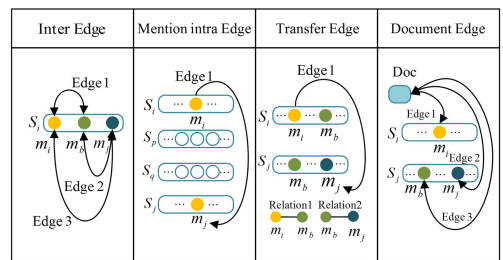


图3 路径构造模块构建的4种路径示意图

Fig. 3 Schematic of four kinds of paths constructed by path construction module

(1) Inter Edge

在关系抽取任务中,距离实体越近的词往往具有越大的关联性,对关系的确定起的作用越大。利用这个信息,在同一句子中的所有提及间都建立一条路径,将其定义为句子内部边,即“Inter Edge”。其在图结构中的显式表现如图 2 中路径 1 所示,由于规定了节点上的数字对应节点所在句子的序号,因此数字相同的节点就代表这些提及出现在同一个句子中,因此需要在它们两两之间建立一条连接。

(2) Mention Inter Edge

同一实体的不同提及可能出现在文档的多个句子之中,虽然所在句距离较远,但它们依然具有极强的关联性。因此,在具有相同共指信息的多个提及间都建立一条路径,即“Mention Inter Edge”,达到加强共指信息联系的效果。其在图结构中的显式表现如图 2 中路径 2 所示,由于规定了节点颜色一致代表提及拥有相同共指信息,因此需要在相同颜色的节点间建立一条连接。

(3) Transfer Edge

对于不同句子中的两个提及节点,当它们之间不存在“Inter Edge”和“Mention Inter Edge”时,在以往的研究中,是

不会在这两个节点中建立连接的。即使它们实际上是存在关系的,该关系事实也只能通过推理才可以得出,这无形中增加了关系抽取不准确的概率。然而,如果这两个提及所在句子中都存在具有相同共指信息的提及,那么根据关系的传递性,可以认为这两个提及有很大概率存在关联,而这个具有相同共指信息的提及可以被定义为“桥梁”。

因此本文提出的 RTDG 模型在路径构造时,在不同句中具有“桥梁”的两个提及之间建立一条路径,将其定义为关系传递路径,即“Transfer Edge”。其在图结构中的显式表现如图 2 中路径 3 所示,即句子[1]中的“The Forbidden City”与“China”存在关系,同时“The Forbidden City”在句子[6]中与“The Imperial Garden”存在关系,因此判断“China”和“The Imperial Garden”存在关系并为其建立一条“Transfer Edge”。

为了避免关系的多次传递带来的信息冗余,限定桥实体只能有一个,即关系只可以传递一次。这里模拟关系的传递情况,即从头部提及开始,遍历所有的桥提及,找到所有尾部提及,在头尾提及间建立一条连接。

(4) Document Edge

由于每个提及都是文档的一部分,因此设置一个文档节点,将其与每个节点都建立一条路径连接,从而使得文档节点或多或少包含任一节点的信息。这样对于两个提及节点而言,即使它们之间不存在“Inter Edge”“Mention Inter Edge”和“Transfer Edge”,也可以在无需在全连接的方式下,通过文档节点实现与另一方的信息交互。这条连接文档节点与提及节点的路径,被定义为“Document Edge”,其在图结构的显式表现如图 2 中路径 4 所示。

3.3 提及层次图聚合网络模块

对文档信息进行路径构造后,得到提及层次的图(Mention Graph, MG),其网络结构如图 4 所示。

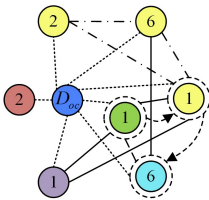


图 4 提及层次图示例

Fig. 4 Mention graph

该模块是在提及层次图上应用 GCN 来聚合来自邻居节点的特征,节点 u 在给定 $l+1$ 层上的语义信息表示 $h_u^{(l+1)}$ 如式(7)所示:

$$h_u^{(l+1)} = \sigma \left(\sum_{k \in K} \sum_{v \in N_k(u)} W_k^{(l)} h_v^{(l)} + b_k^{(l)} \right) \quad (7)$$

其中, K 代表 4 种不同类型的边; $N_k(u)$ 表示通过 k 类型边连接的节点 u 的所有节点集合; $h_v^{(l)}$ 为 $N_k(u)$ 中的某个节点 v 在 l 层上的语义信息表示; $W_k^{(l)}$ 和 $b_k^{(l)}$ 是可训练的参数; σ 是激活函数,如 ReLU (linear rectification function)。

由于 GCN 具有多层结构,其中每一层都包含迭代到该层的提及图的特征,因此为了覆盖所有层的特征,将每层的隐藏状态串联起来,节点 u 的最终表示 m_u 的表达式如式(8)所示:

$$m_u = [h_u^{(0)}; h_u^{(1)}; \dots; h_u^{(N)}] \quad (8)$$

其中, $h_u^{(0)}$ 是节点 u 的初始表示,采用全局定位的方式对其进行

确定。例如这个提及在全篇文档中的位置是第 s 个单词到第 t 个单词,那么根据与之对应的编码信息 $g_j, h_u^{(0)}$ 可以表示为:

$$h_u^{(0)} = \frac{1}{t-s+1} \sum_{j=s}^t g_j \quad (9)$$

另外,文档节点语义信息使用文档经过编码层输出的向量进行初始化。

3.4 实体层次图推理网络模块

实体层次的图(Entity Graph, EG)实质上是将迭代后的提及层次的图按照相同共指信息进行融合,即将表示同一实体的多个提及及节点融合为一个实体节点,并将所有连接这些节点的边进行合并,其网络结构如图 5 所示。

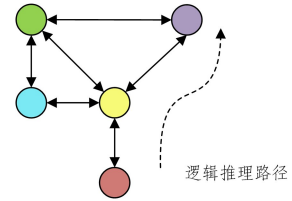


图 5 实体层次图示例

Fig. 5 Entity graph

节点的融合是采用均值的方式,将所有提及节点的信息综合到实体节点中。例如,第 i 个实体节点 e_i 拥有 n 个提及节点 m_i ,则实体节点 e_i 可以表示为:

$$e_i = \frac{1}{N} \sum_n m_n \quad (10)$$

在对边进行合并之后,实体层次图的构建就完成了,通过融合操作后的实体间连接蕴含了潜在的逻辑推理信息。EG 中实体节点 e_i 到实体节点 e_j 的有向边 e_{ij} 表示如式(11)所示:

$$e_{ij} = \sigma(W_q[e_i; e_j] + b_q) \quad (11)$$

其中, W_q 和 b_q 是可训练参数, σ 是激活函数。

如果头实体 e_i 到尾实体 e_j 的第 k 条路径经过实体 e_b ,则该路径 $p_{i,j}^k$ 的表示如式(12)所示。这里的路径仅考虑了实体节点间两跳的路径,同样也可以扩展到多跳的路径中。

$$p_{i,j}^k = [e_{ib}; e_{bj}; e_{ji}] \quad (12)$$

由于同一个实体节点到另一个实体节点的路径可能有多条,因此采用注意力机制^[20-21],使用实体对进行查询,融合不同路径的信息。其中 α_k 为第 k 条路径的归一化注意权值。

$$s_k = \sigma([e_i; e_j] \cdot W_l \cdot p_{i,j}^k) \quad (13)$$

$$\alpha_k = \frac{e^{s_k}}{\sum_h e^{s_h}} \quad (14)$$

$$p_{i,j} = \sum_k \alpha_k p_{i,j}^k \quad (15)$$

通过该方式,将更加关注有用路径的逻辑推理信息,最终使用 $p_{i,j}$ 对实体对 (e_i, e_j) 之间的逻辑推理信息进行表达。

经过实体层次图推理网络模块的处理,每个实体节点都融合了该实体所有提及及节点的信息,并且相关联实体节点间的路径信息也得到了表示。同时采用注意力机制,融合实体对之间的所有路径信息,并对它们之间的逻辑推理信息进行表达。

3.5 关系分类模块

由于整个关系抽取任务被定义为多标签分类任务,关系分类模块就是要将实体对对应到预先设定好的关系中去,预测实体对可能为每一种关系的概率,从而判断实体间的关系。

对于每个实体对 (e_i, e_j) ,首先定义其分类权重值 $I_{i,j}$,如式(16)所示:

$$I_{i,j} = [e_i; e_j; d(e_i, e_j); e_i \odot e_j; m_{\text{doc}}; p_{i,j}] \quad (16)$$

其中, $d(e_i, e_j)$ 表示头尾实体节点的欧氏距离, 用于描述实体向量间的差异, 其计算式如式(17)所示; $e_i \odot e_j$ 表示头尾实体节点的向量点乘, 用于更好地融合头尾实体的信息; 另外, 文档节点 m_{doc} 代表综合的文档信息, 该节点聚合了交叉句子的信息, 也提供了文档的显式表示。

$$d(e_i, e_j) = \sqrt{\sum_{k=1}^n (e_{ik} - e_{jk})^2} \quad (17)$$

接下来就是计算实体对为每一种关系的概率, 计算式如式(18)所示。

$$P(r | e_i, e_j) = \text{sigmoid}(W_r \sigma(W_a I_{i,j} + b_a) + b_b) \quad (18)$$

其中, W_a, W_b, b_a, b_b 是可训练参数, σ 是激活函数。

使用二元交叉熵作为分类损失, 以端到端的方式训练模型。

$$\mathcal{L} = - \sum_{D \in S} \sum_{i \neq j} \sum_{r_k \in R} \mathbb{I}(r_k = 1) \log P(r_k | e_i, e_j) + \mathbb{I}(r_k = 0) \log(1 - P(r_k | e_i, e_j)) \quad (19)$$

其中, S 为整个语料库, \mathbb{I} 为指示函数。

4 实验

4.1 数据集

为验证模型的有效性, 本文在公开数据集 DocRED(Document-Level Relation Extraction Dataset)^[22] 上进行了实验。该数据集内的原始文本数据来源于维基百科和维基数据, 它对 5053 个文档进行人工标注, 得到了 132 375 个实体和 56 354 个关系事实, 是目前最庞大且数据质量较高的人工标注文档级关系抽取数据集。并且, 该数据集中的关系类型多达 96 种, 使得数据具有高覆盖性和多样性。同时, DocRED 中至少有 40.7% 的关系事实只能从多个句子中抽取得到, 这要求模型必须具备全面综合文档信息的能力和有效的逻辑推理能力^[23]。正是因为相比其他关系抽取数据集, DocRED 具有以上这些优势, 所以它能够更好地模拟真实场景中的文档级关系抽取任务。

实验遵循数据集的标准分割, 将 3 053 篇文档用于训练, 1 000 篇用于验证, 另外 1 000 篇用于测试。

4.2 实验设置

在模型的代码实现中, 使用了 3 层 GCN, dropout rate 设置为 0.5, learningrate 设置为 0.001。实验使用 Pytorch^[24] 框架, 并采用 AdamW^[25] 优化器, 使用 DGL 图神经网络的工具包进行模型训练。实验使用了两类预训练模型设置: 一类采用 GloVe(100d)^[26] 和 BiLSTM(256d) 作为嵌入层编码器, 另一类使用 BERT^[27] 编码器。

4.3 实验设计与实验指标

4.3.1 实验设计

本文设置了 3 组对比实验, 用于验证本文提出的 RTDG 模型的性能。

(1) 与传统的关系抽取模型进行对比实验

将 RTDG 模型与 CNN^[28], LSTM^[29], BiLSTM, BERT, CoreBERT^[30] 这 5 种传统关系抽取模型进行对比实验。其中 RTDG 模型分别使用实验设置的两类预训练模型进行编码, 传统模型则是将文档编码为隐藏状态向量序列的模型 $\{h_i\}_{i=1}^n$ 。

(2) 与使用 GCN 的模型进行对比实验

为了进一步验证本文提出方法的有效性, 相较于传统的卷积神经网络和循环神经网络(RNN)^[31], GCN 能够更好地在文档层面上建立实体之间的联系, 从多个角度挖掘文档中的实体表达和语义关联。将 RTDG 与使用 GCN 的经典关系抽取模型进行对比实验, 如 GCNN、EoG、AGGCN^[32]、LSR、模型分层推理网络(Hierarchical Inference Network, HIN)^[33] 和 GAIN 模型。

(3) 应用不同预训练模型的对比实验

分别使用 3 种不同预训练模型, 即 GloVe, BERT-base 和 BERT-large, 与同样使用双图网络的 GAIN 模型进行比较, 进一步验证本文提出方法的鲁棒性与泛化能力。

为了保持一致性, 在 3 组对比实验中均使用相同的数据集与训练参数, 并且进行多轮实验以保证结果的稳定性。

4.3.2 实验指标

实验使用被广泛采用的 F1 和 AUC 文本分类指标, 用于评估在测试集上的模型性能。此外, 还将 Ign F1 和 Ign AUC 指标引入实验中, 通过排除训练集、验证集和测试集中的共同关系事实来计算 F1 和 AUC, 避免了模型在测试集上遇到已经学习过的内容, 从而更加真实地反映模型在新数据上的性能, 使结果更有说服力。

4.4 实验结果

与传统的关系抽取模型进行对比实验的结果如表 1 所列。在 DocRED 数据集上, 结合 GloVe 预训练模型的 RTDG, 以 56.15 的 F1 分数明显高于 CNN, LSTM, BiLSTM, BERT 这 4 种基准模型, 说明其性能较好。而使用 BERT 预训练模型的 RTDG 在测试集中以 62.04 的 F1 分数位居榜首, 说明其性能最佳。

表 1 与传统方法的对比实验结果

Table 1 Comparison experimental results with traditional methods

Model	Dev			Test		
	Ign F1	Ign Auc	F1	Ign F1	Ign Auc	F1
CNN	41.58	36.85	43.45	39.39	40.33	42.26
LSTM	48.44	46.62	50.68	49.48	47.71	50.07
BiLSTM	48.87	47.61	50.93	50.25	48.78	51.04
RTDG-Glove	53.94	52.92	56.31	56.98	53.94	56.15
BERT(base)-RE	—	—	54.16	—	—	53.20
CoreBERT(base)-RE	55.32	—	57.51	—	54.54	56.96
RTDG-BERT(base)	60.03	58.87	61.97	62.04	60.23	62.04

注: RTDG 为本文路径选择的双图模型; 连接符“—”后代表使用的编码模型。

与使用 GCN 的模型进行对比实验的结果如表 2 所列, 在与目前主流的使用图神经网络进行文档关系抽取的模型的对比中, RTDG 模型在 F1 分数上表现出明显优势, 相较于其他模型有了 1.07~4.7 的提升。说明融合关系传递信息的双图文档关系抽取方法可以更好地利用 GCN 的网络结构, 从而在提取文档内关系信息方面达到更好的效果。

应用不同预训练模型的对比实验结果如表 3 所列, 相较于同样使用双图网络的 GAIN 模型, 本文分别使用了 GloVe, BERT-base 和 BERT-large 这 3 个预训练模型与两个模型进行对比。结果发现, 在使用 GloVe 的实验中, RTDG 模型的 F1 指标在验证集和测试集上分别提高了 1.02 和 1.07; 在使用 BERT 的实验中, RTDG-BERT(large) 的 F1 指标在验证集

和测试集上也提高了 1.4 和 1.2,证明了融入关系传递信息操作的有效性。

表 2 与图神经网络相关模型的对比实验结果

Table 2 Comparison experimental results with graph neural network related models

Model	Dev			Test		
	<i>Ign F1</i>	<i>Ign Auc</i>	F1	<i>Ign F1</i>	<i>Ign Auc</i>	F1
GCNN	46.22	—	51.52	—	49.59	51.62
EoG	45.94	—	52.15	—	49.48	51.82
AGCNN	46.29	—	52.47	—	48.89	51.45
LSR-GloVe	48.82	—	55.17	—	52.12	54.18
HIN-GloVe	51.06	—	52.95	—	51.15	53.30
GAIN-Glove	53.05	52.57	55.29	55.44	52.66	55.08
RTDG-Glove	53.94	52.92	56.31	56.98	53.94	56.15

表 3 与双图模型的对比实验结果

Table 3 Comparison experimental results with dual-graph model

Model	Dev			Test		
	<i>Ign F1</i>	<i>Ign Auc</i>	F1	<i>Ign F1</i>	<i>Ign Auc</i>	F1
GAIN-Glove	53.05	52.57	55.29	55.44	52.66	55.08
GAIN-BERT(base)	59.14	57.76	61.22	60.96	59.00	61.24
GAIN-BERT(large)	60.87	61.79	63.09	64.75	60.31	62.76
RTDG-Glove	53.94	52.92	56.31	56.98	53.94	56.15
RTDG-BERT(base)	60.03	58.87	61.97	62.04	60.23	62.04
RTDG-BERT(large)	61.45	62.98	64.49	66.01	61.13	63.96

4.5 消融实验

为进一步证明模型方法的有效性,本文进行了以下两组消融实验,实验结果如表 4 所列。

表 4 消融实验结果

Table 4 Ablation experimental results

Model	Dev			Test		
	<i>Ign F1</i>	<i>Ign Auc</i>	F1	<i>Ign F1</i>	<i>Ign Auc</i>	F1
RTDG-Glove	53.94	52.92	56.31	56.98	53.94	56.15
-Transfer Edge	52.98	51.96	55.12	56.03	53.01	55.03
-Mention Graph	51.77	49.65	53.95	52.59	51.61	53.91
RTDG-BERT(base)	60.03	58.87	61.97	62.04	60.23	62.04
-Transfer Edge	58.97	57.63	60.78	61.15	59.12	60.97
-Mention Graph	57.57	51.99	59.65	55.14	57.80	60.03

第一组实验分别在 Glove 和 BERT(base)这两种预训练模型的设置下在 RTDG 模型里去掉了提及层次图中的“Transfer Edge”,实验结果的 F1 值相较于去掉之前分别下降了 1.12 和 1.07。这表明“Transfer Edge”的构建可以有效帮助 RTDG 获取由关系传递性带来的文档提及节点间的交互信息,促进了相关联提及之间的信息交换,使得文档建模的语义信息更加全面。

第二组实验同样也分别在 Glove 和 BERT(base)这两种预训练模型的设置下在 RTDG 模型里去掉了提及层次图部分,将式(10)中的 m_i 替换为 $h_i^{(0)}$ 后,使用该公式对实体层次图中的实体节点进行初始化,随后对实体层次图进行 GCN 迭代,将不同层的特征连接起来得到 e_i ,再经后续模块进行关系抽取。实验结果表明,在没有提及层次图的情况下,RTDG-Glove 和 RTDG-BERT(base)的验证集上的 F1 分数急剧下降了 2.36 和 2.32。这表明,使用提及层次图不仅有助于捕获相同和不同实体的提及之间的信息,而且在发现文档潜在语义逻辑信息方面也起着至关重要的作用。

4.6 分析和讨论

通过以上 3 组对比实验的结果可以发现,相对于传统的实验方法,RTDG 的 F1 指标始终优于传统模型,在 DocRED 数据集上展现出了广阔的应用前景,证明了图神经网络在文档级关系抽取任务中的有效性。

同时,在与目前存在的使用图神经网络的文档关系抽取方法进行对比时,RTDG 在 F1 分数上也表现出了明显优势,这表明使用提及层次图和实体层次图的双图结构将文档建模与关系推理分隔开,对提升文档级信息的推理能力有很大帮助。而且融合关系传递信息进行路径构造,使得模型能够更完整地文档语义信息进行建模,从而更有效地进行关系抽取。

在与同样使用双图网络的 GAIN 模型进行对比实验时,RTDG 在不同预训练模型下都取得了较好的效果,这更加表明了融合关系传递信息路径构造模块的有效性,也印证了 RTDG 为文档级关系抽取任务提供了一种新的文档建模思路。同样从表 3 中可以观察到,与使用 GloVe 模型相比,使用 BERT 模型的实验结果在 F1 和 Auc 指标方面的表现都更好,尤其是使用 BERT-large 模型时,其 F1 和 Auc 指标都有较大的提升。这说明 BERT-large 模型的预训练效果最好,也更适用于本文提出的 RTDG 模型。同时,这样的结果也进一步验证了 RTDG 可以更好地利用预训练模型来实现更好的关系抽取效果。

结束语 本文提出了一种融合关系传递信息的双图文档级关系抽取方法模型 RTDG,通过将不同句子中提及之间的交互信息引入路径构造中,来提高文档建模的完整性;同时利用 GCN 对复杂信息的聚合推理能力,采用提及层次图和实体层次图的双图结构,分层次建立文档语义信息模型;最后对实体关系进行推理、分类,完成文档级关系抽取任务。

在文档级关系抽取任务中最权威的数据集 DocRED 上进行实验,实验结果表明,相较于没有融合关系传递信息的 GAIN 网络,所提方法的 F1 值有了 1.2 的提升。这证明融合关系传递信息可以更加完整地文档语义信息建模,对文档级关系抽取任务更为有效。此外,本文还探究了不同的编码器对实验结果的影响。实验结果表明,BERT 预训练模型更加契合于本模型,但目前仍有很多预训练模型还未用于实验,未来可能会发现更有效的模型来提高整体的关系抽取效果。

本文提出的文档信息建模思路不仅可以应用于文档级关系抽取,还可以应用于其他信息抽取领域。例如,该方法可以用于事件抽取任务中,融合事件之间相关联的信息,帮助研究人员更好地理解和分析事件的发生和演变过程。未来工作将进一步探索更多的预训练模型,并不断优化方法,以提高模型的性能和可扩展性。

参考文献

- [1] HAN X,GAO T,LIN Y,et al. More data, more relations, more context and more openness: A review and outlook for relation extraction[J]. arXiv,2004. 03186,2020.
- [2] NAYAK T,MAJUMDER N,GOYAL P,et al. Deep neural approaches to relation triplets extraction: A comprehensive survey [J]. Cognitive Computation,2021,13:1215-1232.

- [3] MIWA M, BANSAL M. End-to-end relation extraction using lstms on sequences and tree structures[J]. arXiv:1601.00770, 2016.
- [4] YAN Y, SUN H, LIU J. A Review and Outlook for Relation Extraction[C] // Proceedings of the 5th International Conference on Computer Science and Application Engineering, 2021.
- [5] WANG X, WANG Z, SUN W, et al. Enhancing Document-Level Relation Extraction by Entity Knowledge Injection[C] // 21st International Semantic Web Conference, Virtual Event. Cham: Springer International Publishing, 2022:39-56.
- [6] ZHANG N, CHEN X, XIE X, et al. Document-level relation extraction as semantic segmentation[J]. arXiv:2106.03618, 2021.
- [7] XU B, WANG Q, LYU Y, et al. Entity structure within and throughout: Modeling mention dependencies for document-level relation extraction[C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2021:14149-14157.
- [8] TAN Q, HE R, BING L, et al. Document-level relation extraction with adaptive focal loss and knowledge distillation[J]. arXiv:2203.10900, 2022.
- [9] GIORGI J, BADER G D, WANG B. A sequence-to-sequence approach for document-level relation extraction[J]. arXiv:2204.01098, 2022.
- [10] PENG X, ZHANG C, XU K. Document-level Relation Extraction via Subgraph Reasoning[C] // Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence(IJ-CAD). 2022:4331-4337.
- [11] WU T, KONG F. Document-Level Relation Extraction Based on Graph Attention Convolutional Neural Network[J]. Journal of Chinese Information Processing, 2021, 35(10):73-80.
- [12] CHEN H, HONG P, HAN W, et al. Dialogue relation extraction with document-level heterogeneous graph attention networks[J]. Cognitive Computation, 2023, 15(2):793-802.
- [13] SAHU S K, CHRISTOPOULOU F, MIWA M, et al. Inter-sentence relation extraction with document-level graph convolutional neural network[J]. arXiv:1906.04684, 2019.
- [14] CHRISTOPOULOU F, MIWA M, ANANIADOU S. Connecting the dots: Document-level neural relation extraction with edge-oriented graphs[J]. arXiv:1909.00228, 2019.
- [15] NAN G, GUO Z, SEKULIĆ I, et al. Reasoning with latent structure refinement for document-level relation extraction[J]. arXiv:2005.06312, 2020.
- [16] ZENG S, XU R, CHANG B, et al. Double graph based reasoning for document-level relation extraction[J]. arXiv:2009.13752, 2020.
- [17] LAMPLE G, BALLESTEROS M, SUBRAMANIAN S, et al. Neural architectures for named entity recognition[J]. arXiv:1603.01360, 2016.
- [18] ZHOU J, CUI G, HU S, et al. Graph neural networks: A review of methods and applications[J]. AI Open, 2020, 1:57-81.
- [19] XU W, CHEN K, ZHAO T. Discriminative reasoning for document-level relation extraction[J]. arXiv:2106.01562, 2021.
- [20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30:5998-6008.
- [21] VERGA P, STRUBELL E, MCCALLUM A. Simultaneously self-attending to all mentions for full-abstract biological relation extraction[J]. arXiv:1802.10569, 2018.
- [22] YAO Y, YE D, LI P, et al. DocRED: A large-scale document-level relation extraction dataset[J]. arXiv:1906.06127, 2019.
- [23] ZHU H, LIN Y, LIU Z, et al. Graph neural networks with generated parameters for relation extraction[J]. arXiv:1902.00756, 2019.
- [24] IMAMBI S, PRAKASH K B, KANAGACHIDAMBARESAN G. PyTorch[M] // Programming with TensorFlow. Springer, 2021:87-104.
- [25] LOSHCHEV I, HUTTER F. Decoupled weight decay regularization[J]. arXiv:1711.05101, 2017.
- [26] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation[C] // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing(EMNLP). 2014:1532-1543.
- [27] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv:1810.04805, 2018.
- [28] WANG L, CAO Z, DE MELO G, et al. Relation classification via multi-level attention cnns[C] // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics(Volume 1: Long Papers). 2016:1298-1307.
- [29] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8):1735-1780.
- [30] YE D, LIN Y, DU J, et al. Coreferential reasoning learning for language representation[J]. arXiv:2004.06870, 2020.
- [31] YU Y, SI X, HU C, et al. A review of recurrent neural networks: LSTM cells and network architectures[J]. Neural computation, 2019, 31(7):1235-1270.
- [32] GUO Z, ZHANG Y, LU W. Attention guided graph convolutional networks for relation extraction[J]. arXiv:1906.07510, 2019.
- [33] TANG H, CAO Y, ZHANG Z, et al. HIN: Hierarchical inference network for document-level relation extraction[C] // Advances in Knowledge Discovery and Data Mining: 24th Pacific-Asia Conference (PAKDD 2020). Singapore. Springer International Publishing, 2020:197-209.



KOU Jiaying, born in 1999, postgraduate. Her main research interests include relation extraction and information extraction.



LIU Xianhui, born in 1979, Ph.D, associate professor. His main research interests include machine learning, data mining and big data, networked manufacturing.