



图推理嵌入动态自注意力网络的文档级关系抽取

李云洁, 王丹阳, 刘海涛, 汪华东, 汪培庄

引用本文:

李云洁, 王丹阳, 刘海涛, 等. 图推理嵌入动态自注意力网络的文档级关系抽取[J]. 智能系统学报, 2025, 20(1): 52–63.

LI Yunjie, WANG Danyang, LIU Haitao, et al. Document-level relation extraction of a graph reasoning embedded dynamic self-attention network[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(1): 52–63.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202311021>

您可能感兴趣的其他文章

基于反馈注意力机制和上下文融合的非模式实例分割

Feedback attention mechanism and context fusion based amodal instance segmentation
智能系统学报. 2021, 16(4): 801–810 <https://dx.doi.org/10.11992/tis.202007042>

结合卷积特征提取和路径语义的知识推理

Knowledge-based inference on convolutional feature extraction and path semantics
智能系统学报. 2021, 16(4): 729–738 <https://dx.doi.org/10.11992/tis.202008007>

用于关系抽取的注意力图长短时记忆神经网络

Attention graph long short-term memory neural network for relation extraction
智能系统学报. 2021, 16(3): 518–527 <https://dx.doi.org/10.11992/tis.202008036>

混合神经网络和条件随机场相结合的文本情感分析

Text sentiment analysis combining hybrid neural network and conditional random field
智能系统学报. 2021, 16(2): 202–209 <https://dx.doi.org/10.11992/tis.201907041>

基于注意力融合的图像描述生成方法

An image caption generation method based on attention fusion
智能系统学报. 2020, 15(4): 740–749 <https://dx.doi.org/10.11992/tis.201910039>

图神经网络推荐研究进展

Research advances in graph neural network recommendation
智能系统学报. 2020, 15(1): 14–24 <https://dx.doi.org/10.11992/tis.201908034>

DOI: 10.11992/tis.202311021

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20240923.1515.002>

图推理嵌入动态自注意力网络的文档级关系抽取

李云洁¹, 王丹阳², 刘海涛^{1,3}, 汪华东⁴, 汪培庄³

(1. 辽宁工程技术大学 理学院, 辽宁 阜新 123000; 2. 中国热带农业科学院科技信息研究所, 海南 海口 571000; 3. 辽宁工程技术大学 智能工程与数学研究院, 辽宁 阜新 123000; 4. 清华大学 计算机系, 北京 100084)

摘要: 文档级关系抽取是指从文档中抽取所有具有语义关系的实体对并判断其关系类别, 与句子级关系抽取不同, 这里实体关系的确定需要根据文档中多个句子推理得到。现有方法主要采用自注意力进行文档级关系抽取, 但是运用自注意力进行文档级关系抽取需要面临两个技术挑战: 即长文本语义编码存在的高计算复杂度和关系预测需要的复杂推理建模, 故提出一种图推理嵌入动态自注意力网络 (graph reasoning embedded dynamic self-attention network, GSAN) 模型。该模型借助门限词选择机制动态选择重要词计算自注意力实现对长文本语义依赖的高效建模, 同时考虑以选择词为全局语义背景与实体候选、文档节点一起构建文档图, 将文档图的图推理聚合信息嵌入到动态自注意力模块中, 实现模型对复杂推理建模的能力。在公开的文档级关系数据集 CDR 和 DocRED 上的实验结果表明, 文中提出的模型较其他基线模型有显著提升。

关键词: 文档级关系抽取; 图推理; 动态自注意力网络; 自注意力机制; 门限词选择机制; 文档图; 图注意力网络; 关键词

中图分类号: TP391 文献标志码: A 文章编号: 1673-4785(2025)01-0052-12

中文引用格式: 李云洁, 王丹阳, 刘海涛, 等. 图推理嵌入动态自注意力网络的文档级关系抽取 [J]. 智能系统学报, 2025, 20(1): 52-63.

英文引用格式: LI Yunjie, WANG Danyang, LIU Haitao, et al. Document-level relation extraction of a graph reasoning embedded dynamic self-attention network[J]. CAAI transactions on intelligent systems, 2025, 20(1): 52-63.

Document-level relation extraction of a graph reasoning embedded dynamic self-attention network

LI Yunjie¹, WANG Danyang², LIU Haitao^{1,3}, WANG Huadong⁴, WANG Peizhuang³

(1. Institute of Mathematics and Systems Science, Liaoning Technical University, Fuxin 123000, China; 2. Institute of Scientific and Technical Information, Chinese Academy of Tropical Agricultural Sciences, Haikou 571000, China; 3. Institute of Intelligence Engineering and Mathematics, Liaoning Technical University, Fuxin 123000, China; 4. Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract: Document-level relation extraction refers to the extraction of all entity pairs with semantic relationships from documents and judging their relationship categories. It is different from sentence-level relation extraction, where the determination of entity relationships needs to be inferred from multiple sentences in the document. The existing methods mainly use self-attention for document-level relation extraction, but the use of self-attention for document-level relation extraction needs to address two technical challenges: the high computational complexity of long text semantic encoding and the complex reasoning modeling required for relationship prediction. Therefore, a graph reasoning embedded dynamic self-attention network model (GSAN) is proposed. With the aid of gated word selection mechanism, GSAN dynamically selects important words to calculate self attention, achieving high-efficiency modeling for semantic dependency of long text sequences. At the same time, it is considered to construct a document graph with the word selection as the global semantic background, entity candidates and document nodes. Then, the graph reasoning aggregation information of the document graph being embedded into the dynamic self-attention module enables the model to model complex reasoning. The experimental results demonstrate that the proposed model is a significant improvement over other baseline models on the public document-level relation dataset CDR and DocRED.

Keywords: document-level relation extraction; graph reasoning; dynamic self-attention network; self-attention mechanism; gated token selection mechanism; document graph; graph attention network; key word

收稿日期: 2023-11-17. 网络出版日期: 2024-09-24.

基金项目: 国家自然科学基金项目 (61350003); 辽宁省教育厅高等学校基本科研项目重点攻关项目 (LJKZZ-20220047); 中央级公益性科研院所基本科研业务费专项 (1630072023005).

通信作者: 刘海涛. E-mail: haitao641@163.com.

关系抽取目的是识别文本中所含实体之间的语义关系, 其在各种基于知识的应用中扮演着重要角色, 例如知识库问答^[1]、知识图谱构建^[2]等。现有的方法^[3-7]主要集中在句子级关系抽取, 即识

别单句子中实体之间的关系。但句子级实体关系抽取方法无法很好识别跨句子实体之间的关系,而在实际文档中大量实体关系蕴含于复杂的长文本中。Yao等^[8]对Wikipedia文档的人工标注数据统计发现,至少有40.7%的关系事实需要建立在多句子学习基础上才能获取。因此,研究文档级实体关系抽取对知识获取和文档知识理解非常有必要。

文档级关系抽取主要存在如下几个挑战。首先,由于文档中实体关系的实体可能位于不同的句子中,传统句子级别关系抽取方法无法很好处理。其次,文档级文本序列较长,传统文本编码主要限制在句子级编码,若直接将其应用在文本序列上会导致高计算复杂度或高内存消耗,如现在流行的自注意力模型,需要存储注意力矩阵,其内存复杂度将关于序列长度呈二次方指数增长^[9]。第三,大量的关系识别需要在文档的多个句子中进行推理,要求模型能够在获取长期背景依赖基础上具有一定的推理能力。

近年来,研究者开始关注跨句子文档级关系抽取工作^[8,10-13]。早期研究^[14-15]主要限制在短文本文档关系抽取,通常仅包含几个句子。部分研究^[14,16]尝试研究长文档的关系抽取,处理文档中一个二元实体关系或 N 元实体关系。但其忽略了文档中多个实体关系之间的推理和信息交互,所以无法处理一般的文档级关系抽取。最近研究者提出构建文档级图模型,以利用长文本中多个实体的语义交互信息。例如,Sahu等^[11]将词作为节点、利用句法依存关系和词序列信息构建边,但是其依赖于依存句法解析的结果。Christopoulou等^[17]利用实体候选、实体、句子作为节点构建文档图,提出了边导向的图神经网络,但其不能构建词与词之间的交互关系。

随着Transformer的提出,自注意力机制因其高可并行性和长期依赖建模能力,使得其在自然语言处理各种任务中被广泛采用。Zhuang等^[18]在用自注意力建模多篇章阅读理解问题时,提出了动态自注意力网络模型。该模型核心是门限词选择机制选择部分重要的词计算注意力,其不仅可以实现对词与词之间长期依赖的建模,同时可以显著降低自注意力内存消耗。受该方法启发,本文将动态自注意力网络模型应用于文档级关系建模,考虑文档级实体关系推理建模的需求,提出了图推理嵌入的动态自注意力网络模型。本文的主要贡献包括以下3点:

1)提出了针对文档级关系抽取的图推理嵌入

的动态自注意力网络模型,实现对长文本依赖和实体关系推理的高效建模,同时提出了基于文档节点表示的门限选择机制进行关键词的动态选择;

2)在构建文档图网络时,考虑在文档图中引入动态关键词作为图节点,动态关键词为基于文档图的图神经网络提供了全局语义背景信息;

3)对比实验结果表明提出的方法取得了目前最好效果,也验证了提出方法的有效性和优越性。

1 相关工作

文档级关系抽取是近年来的研究热点,其不仅局限于抽取单个句子间实体关系。Yao等^[8]构建了一个大规模人工标注的文档级关系抽取数据集DocRED,其中大约有40.7%的实体关系需要在多个句子推理基础上得到。因此,如何抽取跨句子关系和推理关系建模,是目前文档级关系抽取的关键问题。部分研究工作^[14-16,19]利用依存图来获取文档层面的特征信息,但是忽略了对文档中丰富的推断建模。之后针对该问题提出许多方法。Tang等^[20]从实体级、句子级、文档级的不同层次信息提出了层次推断网络。Christopoulou等^[17]提出了边导向的图神经网络,将文档中实体候选(entity mention)、实体、句子作为图网络节点。Liu等^[21]利用实体候选、句子构建文档图和实体图,提出了图信息聚合和跨句子推理网络。Nan等^[22]采用依存树来获取文档语义信息,并将实体候选、实体加入图网络中,利用图神经网络实现推理建模。Zeng等^[23]则从实体候选级和实体级两个层次分别构建图,提出了基于双图推理的神经网络模型。Verga等^[10]提出了基于Transformer的文档级关系抽取网络,利用自注意力来实现对摘要级长文本序列的长期依赖建模。以上这些方法要么无法建模词与词之间的交互关系,在图推理过程中忽略了文档的背景语义信息,要么在语义编码阶段都采用BERT(bidirectional encoder representation of Transformer)、LSTM(long short-term memory)、自注意力等神经网络进行编码,忽视了长文本序列计算复杂度和内存消耗问题,导致这些方法的扩展性受到很大局限。

本文主要关注自注意力网络^[9,18,24-25],其作为Transformer^[24]、BERT^[26]等自然语言处理模型的核心模块,得到了广泛研究。Vaswani等^[24]探索利

用自注意力来替换循环神经网络对背景依赖建模的可能性。一些研究提出了自注意力机制的各种变型^[9,27-29]。此外, Shen 等^[9]探索了降低自注意力计算复杂度问题。目前有大量研究探索如何设计高效的自注意力网络^[30], 如 Zhuang 等^[18]在研究多篇章阅读理解问题时, 提出了动态自注意力方法实现对词水平下的长期语义依赖建模。本文将在该方法的基础上研究基于动态自注意力的文档级关系抽取, 考虑之前方法无法对实体关系进行推理建模, 因此, 提出了图推理嵌入的动态自注意力方法。

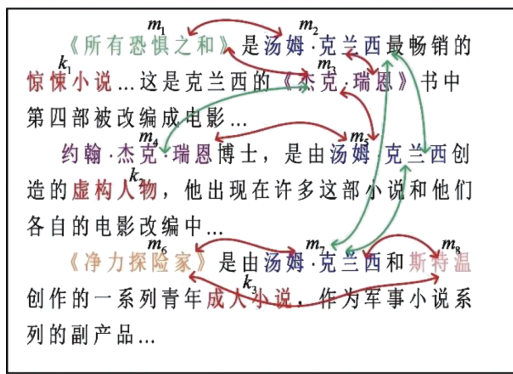
2 图推理嵌入动态自注意力网络的文档级关系抽取模型设计

2.1 任务定义

文档级关系抽取。给定一个含有 N 个句子文档 $D = \{S_i\}_{i=1}^N$ 和其含有的实体集合 $E = \{e_i\}_{i=1}^p$, 其中 $S_i = \{w_{ij}\}_{j=1}^{M_i}$ 表示第 i 个句子, 其包含 M_i 个字, $e_i = \{m_{ij}\}_{j=1}^{Q_i}$ 表示该实体包含 Q_i 个实体候选, 每个实体候选 m_i 为一个连续的词序列片段。文档级关系抽取的目的是预测实体集合 E 中任意实体对之间存在的实体关系, 即

$$\{(e_i, r_{ij}, e_j) : e_i, e_j \in E, i \neq j, r_{ij} \in R\}$$

其中 R 表示给定的实体关系类别集合。



输入文档

文档节点 d
文档图

图 1 融入动态关键词的文档图

Fig. 1 Document graph incorporating dynamic keywords

2.2.1 文档图构建

文档图是一个建立在文档基础上的图关系网络, 文档图节点包含 3 种类型, 实体候选节点、文档节点和动态关键词节点, 并包含以下 5 种类型边:

实体共指关系: 共指同一个实体的所有实体候选之间相互连接, 有助于实体与实体候选之间进行信息交互。

2.2 图推理模块

基于文档图的图推理模块可以对文档关系抽取所需的丰富且复杂的推理进行建模, 本文采用类似 Zeng 等^[23]的图神经网络作图推理模块。图推理模块包含文档图构建和图神经网络两部分。首先构建一种实体候选级包含 3 种类型节点的文档图, 除 Zeng 等^[23]提到的实体候选节点和文档节点外, 还引入动态关键词节点(2.3 节将给出如何构建动态关键词)。通常一个实体在文档中包含多个共指的实体候选, 每个实体候选都对应一个实体。文档节点用来获取文档级的全局信息, Zeng 等^[23]指出文档节点可以作为中枢帮助长距离实体候选之间进行信息交互。动态关键词是指文档里在图推理过程中具有重要作用的词。在文档图的图神经网络推理过程中, 引入动态关键词到图网络中进行推理, 其是建立在如下直觉上: 人在判断文档中实体关系时, 需要反复观察文档并快速得到一些关键的信息进行逻辑归纳, 这些信息粒度一般较小, 主要是文档中的一些重要的词。目前建立的图神经网络的关系抽取模型, 在图推理过程中忽略了这一点。图 1 给出了文档图构建的过程, 其中 m_i 表示实体候选, 相同颜色的实体候选共指同一个实体; k_1 、 k_2 、 k_3 表示 3 个动态关键词, 与其相连虚线表示关键词与其他节点的连接关系; 黑色表示文档节点 d 。

实体共现关系: 如果两个不同的实体候选同时出现在同一个句子中, 在这两个实体之间增加边。通过该连接边可以增强句子内的实体之间的交互。

实体候选与文档关系: 在图中增加所有实体候选到文档节点的边。利用文档节点可以增强长距离实体之间的交互, 同时文档节点表示反映的

主题信息也有助于实体关系的预测。

实体与动态关键词关系: 在实体关系图推理过程中需要反复观察其所在文档中的重要背景信息。因此, 在图中增加动态关键词节点与实体的连接边, 这里仅连接与关键词出现在相同段落(或句子)的实体候选。通过关键词可以增强图推理中获取文档的全局背景语义信息。这些关键词通过动态自注意力中的门限选择机制获得, 因此称之为动态关键词。

文档与动态关键词关系: 在图中考虑动态关键词与文档节点之间的关系, 增加该关系与关键词没有直接连接的实体候选也能够通过文档节点关注到文档中的全局背景语义信息。

2.2.2 图神经网络

根据 2.2.1 节中所述节点和关系构建文档级图, 并将文档图输入多层图注意力网络(graph attention networks, GAT)^[31]进行特征表示和图推理学习。给定第 l 层的节点 u , 根据图注意力操作计算节点 u 在第 $l+1$ 层的表示:

$$\mathbf{h}_u^{(l+1)} = \sigma \left(\sum_{r \in R} \sum_{v \in N_r(u)} \alpha_{uv}^r \mathbf{W}_v \mathbf{h}_v^{(l)} \right) \quad (1)$$

式中: $\mathbf{W}_v \in \mathbb{R}^d$ 表示 Value 的线性变换参数, $N_r(u)$ 表示节点关于边关系 r 的邻居节点集合, R 表示给定的实体关系类别集合, α_{uv}^r 表示 u 关于其邻居节点 $v \in N_r(u)$ 的注意力得分, 即

$$\begin{aligned} b_{uv}^r &= \mathbf{W}_2^r \sigma(\mathbf{W}_1^r [\mathbf{h}_u^{(l)}; \mathbf{h}_v^{(l)}; \mathbf{h}_r^{(l)}]), \\ \mathbf{W}_1^r &\in \mathbb{R}^{d \times 3d}, \mathbf{W}_2^r \in \mathbb{R}^{1 \times d} \quad (2) \\ \alpha_{uv}^r &= \text{softmax}_{uv}(b_{uv}^r) = \frac{\exp(b_{uv}^r)}{\sum_{r \in R} \sum_{v \in N_r(u)} \exp(b_{uv}^r)} \end{aligned}$$

式中: b_{uv}^r 反映的是关系 (u, r, v) 的重要性, 由两层前馈网络得到; σ 表示非线性变换(如 LeakyReLU 激活函数); $[\mathbf{h}_u^{(l)}; \mathbf{h}_v^{(l)}; \mathbf{h}_r^{(l)}]$ 表示对第 l 层节点 u 、 v 和关系 r 的表示向量的拼接。

图网络采用多头注意力^[24]方式获得第 l 层的节点 u 表示。对以上过程重复 H 次, 得到多头图注意力的节点表示:

$$\mathbf{h}'_u = \parallel_{k=1}^H \mathbf{h}_u^k \quad (3)$$

式中: \mathbf{h}_u^k 表示节点 u 第 k 个头注意力的输出, 通过式(2)计算得到; \parallel 表示拼接操作。利用多头注意力, 能够使每个节点从不同视角去获取网络节点的信息。然后, 对拼接后的表示向量采用分位置线性变换和残差连接进行信息聚合, 得到节点 u 在第 $l+1$ 层的节点输出:

$$\mathbf{h}_u^{(l+1)} = \text{LayerNorm}(\mathbf{F}^Y(\mathbf{h}'_u) + \mathbf{h}_u^{(l)}) \quad (4)$$

式中 \mathbf{F}^Y 表示分位置线性变换。

为获得图注意力网络所有层的表示, 把节点 u 在所有层的表示向量进行拼接作为该节点的最终表示向量:

$$\mathbf{m}_u = [\mathbf{h}_u^{(0)} \quad \mathbf{h}_u^{(1)} \quad \dots \quad \mathbf{h}_u^{(L)}]$$

式中: L 表示图注意力网络的层数, $\mathbf{h}_u^{(0)}$ 表示节点 u 的初始输入向量。假设节点 u 表示的实体候选是由 l 个词组成的序列片段, 假设其对应的词向量序

列为 $[\mathbf{w}_i \quad \mathbf{w}_{i+1} \quad \dots \quad \mathbf{w}_{i+l}]$, 那么 $\mathbf{h}_u^{(0)} = \sum_{j=1}^l \mathbf{w}_{i+j}/l$, $\mathbf{h}_r^{(0)} = (\mathbf{h}_u^{(0)} + \mathbf{h}_v^{(0)})/2$ 。文档节点的初始向量可以根据编码器的输出获得, 采用 BERT 作为编码器时, 可以采用 [CLS] 对应的输出向量, 若采用 Bi-LSTM, 可以采用序列首尾词分别在后向 LSTM 和前向 LSTM 中的隐向量表示的拼接。对于动态关键词节点, 直接采用关键词对应的词向量表示, 其表示来自动态自注意力网络。

2.3 图推理嵌入动态自注意力网络

为了使得动态自注意力模块具有图推理能力, 融入图结构知识信息, 提出一种图推理嵌入的动态自注意力网络(graph reasoning embedded dynamic self-attention network, GSAN), 图 2 给出了该网络模型的整体框架。主要思想是: 首先, 根据门限选择机制动态选择最重要的 Top K 词; 其次, 根据以下两步操作得到选择词的两种表示向量: 1) 基于图推理网络的特征表示, 即将选择的 K 个词作为动态关键词, 输入到图推理模块构建文档图并经过一层或多层 GAT 推理学习得到各节点特征表示, 取出动态关键词表示(本文采用 1 层 GAT); 2) 基于自注意力的特征表示, 即将选择的词输入到自注意力模块, 并得到注意力特征表示输出。最后, 实现多头自注意力, 对前两步重复 H 次, 得到 H 组选择词的两种特征表示, 然后与输入特征输入到特征融合模块中, 得到融合后特征表示。这里采用 H 组目的是使该模型能够在不同的特征子空间获取重要信息。

2.3.1 门限词选择机制

门限值可以动态选择最重要的 Top K 词, 不同于 Zhuang 等^[18]的研究, 该机制采用图网络中文档节点表示 d 与动态自注意力中背景表示计算每个词的重要性权重:

$$\mathbf{G} = \text{sigmoid}(\langle \mathbf{F}^D(d), \mathbf{F}^U(U) \rangle) \quad (5)$$

式中: $\langle \cdot, \cdot \rangle$ 表示内积运算, \mathbf{F}^D 和 \mathbf{F}^U 为仿射变换。式(5)中 \mathbf{G} 包含 H 个头, 其中第 h 行 $\mathbf{g}_h \in \mathbb{R}^L$ 表示门限关于第 h 注意力头的输出。

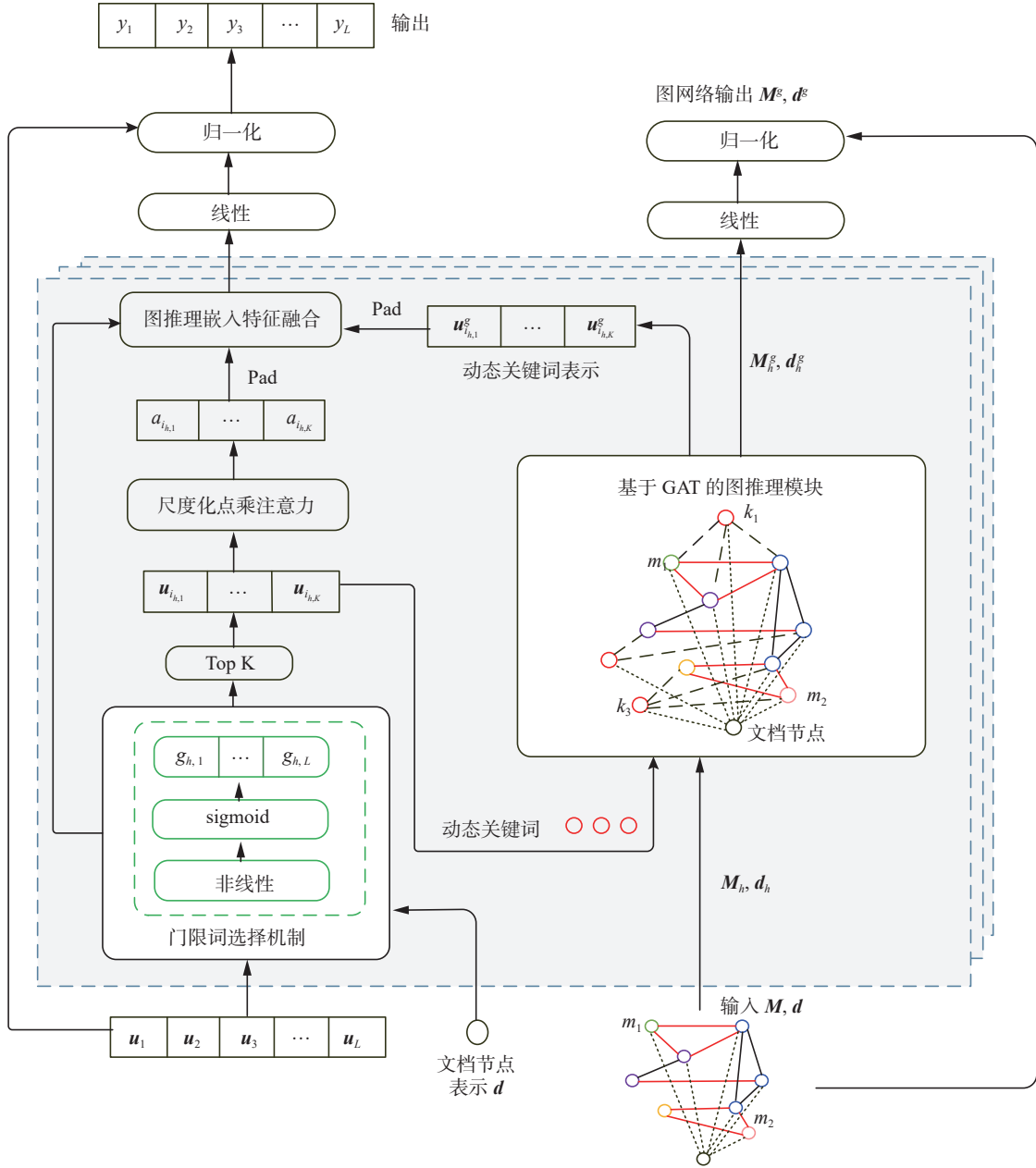


图 2 图推理嵌入动态自注意力网络模型

Fig. 2 Graph reasoning embedded dynamic self-attention network model

2.3.2 选择词的自注意力表示

在选择词背景表示上计算自注意力：

$$A_h = \text{self-attention}(Q_h, K_h, V_h)$$

式中： $Q_h \in R^{H \times K}$ 、 $K_h \in R^{H \times K}$ 、 $V_h \in R^{H \times K}$ 分别为第 h 头的检索 Query 和键值 Key 矩阵，其由 U_h 分别线性变换得到； $A_h \in R^{H \times K}$ 为第 h 头的动态自注意力输出。在选择词上计算自注意力不仅可以对长文本序列的长期依赖进行编码，同时可以有效降低模型对长文本背景表示学习的内存复杂度。此时，自注意力的空间复杂度为 $O(K^2)$ ，因 K 远小于词序列长度 L ， $O(K^2)$ 远小于标准的动态注意力复杂度 $O(L^2)$ 。

2.3.3 选择词的图推理特征表示

记选择词对应的背景表示为 $U = \{u_{h1}, u_{h2}, \dots, u_{hK}\}$ 。记文档图中实体候选节点关于第 h 头的特征表示为 $U_h = [u_{i_{h,1}} \dots u_{i_{h,j}} \dots u_{i_{h,K}}] \in R^{D \times K}$ ，其中表示选择词的位置，文档节点关于 h 头的特征表示 d_h 。将 $U_h = [u_{i_{h,1}} \dots u_{i_{h,j}} \dots u_{i_{h,K}}] \in R^{D \times K}$ 作为动态关键词代入图推理模块中，根据式 (3) 得到该组的图推理特征表示：

$$U_h^g, M_h^g, d_h^g = \text{GAT}_G(U_h, M_h, d_h) \quad (6)$$

式中： GAT 表示图注意力网络， G 表示图， U_h^g 、 M_h^g 、 d_h^g 分别表示动态关键词节点、实体候选节点、文档节点的对应的输出。

2.3.4 图推理嵌入特征融合

对选择词的图推理特征表示和自注意力表示进行融合。首先, 两种特征表示分别填充零向量得到整个句子上的输出 $A_h^* = \text{Pad}(A_h) \in R^{\frac{D}{H} \times L}$ 和 $U_h^* = \text{Pad}(U_h^g) \in R^{\frac{D}{H} \times L}$, 则第 h 头的输出表示为

$$Z_h = (F_h + A_h^* + U_h^*) \cdot \frac{g_h}{\max(g_h)}$$

式中 $F_h \in R^{\frac{D}{H} \times L}$ 是为了阻止梯度消失。在特征融合模块中, 融合了选择词在图推理模块中的特征表示。最后, 对所有头注意力输出进行融合得到动态自注意力模块的输出 $Y \in R^{D \times L}$:

$$Y = F^Y([Z_1; Z_2; \dots; Z_H]) + U \quad (7)$$

式中: F^Y 表示线性变换; $[\cdot; \cdot]$ 表示对所有头张量按最后 1 维进行拼接操作, 保证门限激活值能够在选择重要的词时具有稀疏性, 对 G 增加一个 L_1 正则。

2.4 实体对特征表示模块

本模块主要是根据图推理嵌入动态自注意力网络和图推理模块的输出得到文档中任意实体对之间特征表示, 并利用该表示进行实体对关系类别预测。目前实体对特征表示存在多种方法, 如 Zeng 等^[23] 给出了实体级图推理模块, 首先将对应相同实体的实体候选表示聚合得到对应的实体表示, 即 $e_i = \frac{1}{Q_i} \sum_{j=1}^{Q_i} m_{ij}$, Q_i 表示对应第 i 实体 e_i 实体候选个数 (见小节 2.1 中任务定义), m_{ij} 为实体候选 $m_{ij} \in e_i$ 的特征表示; 并根据实体对之间是否存在实体候选连接边构建实体级图的连接边。构建完实体图之后, 定义每个连接实体对之间有向表示为

$$e_{ij} = \sigma(W[e_i; e_j] + b)$$

式中: $W \in R^{d \times 2d}$ 和 $b \in R$ 为可训练参数。然后, 基于向量边表示考虑了实体对之间路径信息, 定义了任意实体对 (e_h, e_t) 之间二阶路径表示 $p_{h,t} =$

$$\sum_i \alpha_i p_{h,t}^i, \text{ 其中 } p_{h,t}^i = [e_{ho}; e_{ot}; e_{to}; e_{oh}] \text{ 表示头实体 } e_h \text{ 和尾实体 } e_t \text{ 之间通过实体 } e_o \text{ 的第 } i \text{ 条路径, } \alpha_i = \frac{e^{s_i}}{\sum_j e^{s_j}}$$

表示第 i 条路径的注意力权重, $s_i = \sigma([e_h; e_t] \cdot W_l \cdot p_{h,t}^i)$ 。路径信息的引入有助于在关系预测时引入潜在的逻辑推理链信息。最终, 给定任意待预测实体对 (e_h, e_t) , 则其特征表示可以定义为

$$E_{ht} = [e_h; e_t; d^g; p_{h,t}] \quad (8)$$

式中: d^g 为图推理嵌入动态自注意力网络的文档节点表示。

2.5 关系类别预测模块

给定实体对特征表示之后, 实体对关系类别预测可以看成多分类问题^[32], 也可以看成多标签分类问题^[23]。本文采用多标签分类的方式, 得到实体对 (e_h, e_t) 具有关系类别 r 的概率:

$$P_{ht} = \text{sigmoid}(F_2(E_{ht}))$$

式中: F_2 表示具有 $|R|$ 个输出节点的两层前馈神经网络。最终, 模型在单个文档样本上的优化损失定义为

$$\mathcal{L} = - \sum_{e_h, e_t \in E, h \neq t} \sum_{r \in R} [I(r=1) \log P_{ht}(r) + I(r=0) \log (1 - P_{ht}(r))] \quad (9)$$

式中: $I(\cdot)$ 表示示性函数, $P_{ht}(r)$ 表示实体对具有关系类别 r 的概率。

3 实验仿真与结果分析

本文将在化学-疾病关系数据集 (CDR) 和文档级关系抽取数据集 (DocRED) 两个文档级关系抽取数据集上对提出的方法有效性进行验证, 并与现有方法进行对比实验。表 1 给出了两个数据集的统计信息。

表 1 两个数据集 CDR 和 DocRED 的统计信息

Table 1 Statistics information for two datasets CDR and DocRED

数据集	数据划分	文档	关系类别	实体关系	非实体关系
CDR ^[33]	训练集	500	1	1 038	4 280
	验证集	500	1	1 012	4 136
	测试集	500	1	1 066	4 270
DocRED ^[8]	训练集	3 053	96	38 269	1 163 035
	验证集	1 000	96	12 332	385 263
	测试集	1 000	96	12 842	379 316

3.1 数据集

化学-疾病关系数据集 (CDR) 是由 Li 等^[33] 创建, 其包含了 1 500 PubMed 的摘要, 并均等地

分为训练集、验证集和测试集。该数据集是由人工标注的包含一种实体关系类型, 即化学药品和疾病之间是否存在诱导关系。

文档级关系抽取数据集 (DocRED) 由清华大学自然语言处理实验室构建^[8], 文档数据来源于维基百科和 Wikidata, 覆盖了科学、艺术、个人生活等领域, 包含 96 种实体关系类别。该数据集提供了人工标注和远程监督的两个部分数据。在本试验中, 仅采用人工标注的数据集, 包括近 5 万份文档, 数据集按照 3:1:1 划分成训练集、验证集和测试集。

3.2 实验设置和评价标准

提出的模型在公开代码 GAIN^[23] 基础上实现, 采用 GPU V100 作为所有实验的计算平台。实验文本序列背景编码采用 BERT 预训练语言模型。模型优化器为 AdamW^[34], 其权重退化率为 1×10^{-4} , 学习率为 1×10^{-5} 。除模型 BERT 之外其他参数采用学习率为 1×10^{-3} 。模型采用具有图推理嵌入层的动态自注意力模块, 其中图推理层数为 1, 自注意模块和图注意力模块均采用 2 个注意力头 ($H=2$)、256 维隐向量维度。其他参数设置与 GAIN^[23] 保持一致。对于数据集 DocRED, 预训练模型为 BERT-base, 采用 AUC (area under curve) 值和 F_1 值作为评价指标, 此外其测试集和训练集中实体关系存在重合, 这里还给出排除与训练集实体关系重合部分的 AUC 值和 F_1 值, 分别记为 Ing AUC 值和 Ing F_1 值。对于 CDR 数据集, 采用精确率 (P)、召回率 (R) 和 F_1 值作为评价指标, 采用与 Wang 等^[32] 相同的预处理方式, 并转成与 DocRED 相同的数据格式, 与文献 [32] 相同, 这里 BERT 模型同样选择 BioBERT-Base v1.1^[35], 其在生物医学数据集上重新训练得到。其他参数设置两个模型保持一致。

3.3 基线模型

1) 针对数据集 CDR, 采用以下基准模型。

GCN (graph convolutional network)^[36]: 一种利用剪枝的依存树作为图输入图卷积网络模型, 属于句子级关系抽取方法。

Ensemble^[37]: Zhou 等^[37] 针对 CDR 数据集探索了利用句法和语义特征信息关系抽取方法, 该方法是基于特征抽取、树核、神经网络等方法的整合模型。

GCN-MEM (maximum entropy model)^[38]: 利用卷积网络和最大熵结合方法。

Graph-Kernel^[39]: Panyam 等^[39] 探索了一种图核的关系抽取方法。

Bi-SAN (self-attending network)^[10]: 一种基于自注意力网络的摘要级关系抽取方法, 采用自注意力对长文本词序列的长期依赖进行建模, 给出了

利用双线性网络实现对命名实体和实体关系联合抽取, 这里列出的是其实体给定下关系抽取的实验结果。

Stack-BiLSTM^[40]: 一种基于多个 BiLSTM 的层次网络。

Umls-ScispaCy^[41]: Li 等^[41] 针对 CDR 数据集探索了利用实体链接方法, 该方法用于度量实体提及能够正确链接到该实体。

有向边图模型 (edge-oriented graphs, EoG)^[17]: 在生物领域的文档级关系抽取任务中表现出色, 首先利用启发式方法构建图, 然后利用有向边模型进行推理; 注意力引导的图卷积网络 (attention-driven GCN, AGCN)^[42-43] 是目前句子级关系抽取表现最优的模型, 并通过自注意力^[44] 构建潜在图结构。

Seq2rel^[45]: 对关系提取使用 Seq2seq 学习, 是基于将关系提取的信息线性化为字符串, 因为该线性化具有足够的表达能力。

BERT-GLRE (global-to-local neural network for relation extraction)^[32]: 一种图神经网络模型, 该方法从实体的全局和局部表示、背景关系表示^[46-49] 等多个角度综合对文档信息进行编码。其实验结果是本文在公开代码和相同设置基础上重新运行的实验结果, 与文献 [32] 公开的结果有些差异。

Two-Phase-BERT^[50]: 针对 DocRED 给出了基于 BERT 的两阶段预测方法, 首先预测实体之间是否存在关系, 然后预测实体对存在的实体类别。

GCN、Two-Phase-BERT、EoG 这 3 种方法在 CDR 的实验结果来自 Wang 等^[32]。

2) 针对数据集 DocRED, 除了 Two-Phase-BERT、EoG、BERT-GLRE 外还采用如下的基准模型。

BiLSTM: Yao 等^[8] 在提出 DocRED 数据集的同时给出了几种基准模型的实验结果, 分别采用 CNN、BiLSTM 作为文本序列编码器, 并得到每个实体的表示向量, 最后对每个实体对进行关系类别预测, 其中 BiLSTM 表现最优。此外, 该工作也提供了背景敏感的关系抽取 (Context-Aware) 模型的实验结果, 除采用 LSTM 作为编码器之外, 利用注意力机制来获取背景关系信息来改善句子内的关系抽取性能。

基于 BERT 关系抽取模型: 直接采用各种预训练语言模型作为文档序列的编码器进行关系抽取, 这类方法有 RoBERTa^[50-51]、BERT^[52] 和 CorefBERT^[53]。

BERT-LSR (latent structure refinement)^[22]: 通过自动生成的文档级图进行关系推理学习。同时,

提供了其他 2 种图方法的实验结果, 分别是: GAT^[31], 其根据局部注意力机制学习加权的图结构; 图卷积神经网络 (graph convolutional neural network, GCNN)^[11], 其通过共指链接构建文档级图, 然后应用关系图卷积网络进行推理。

BERT-HIN (hierarchical inference network)^[20]: 采用 BERT 输出作为词向量, 构建层次推断网络从 3 个层面 (实体级、句子级和文档级) 汇聚信息进行关系抽取。

RSMAN^[54]: 一种关系特定提及注意网络, 通过对候选关系的不同实体提及进行选择关注。

GRACR (graph information aggregation and cross-sentence reasoning network)^[21]: 基于 BERT 对文档进行编码, 通过构建文档级图汇聚文档中所有实体候选和句子的语义信息, 并构建实体级图对长距离跨句实体对之间的关系进行抽取。

BERT-GAIN (graph aggregation-and-inference network)^[23]: 利用 BERT 分别对文档进行编码, 然后从实体候选级构建图网络进行推理学习, 之后在此基础上整合实体表示构建实体级图网络。

以上基于 BERT 的模型, 均采用 BERT-base 大小的预训练模型, 即 12 层、8 个头、768 的隐层维度。

3.4 实验结果与分析

表 2 和表 3 给出了提出模型 GSAN 分别在 CDR 和 DocRED 两个数据集上实验结果。从表中可以观察到如下结果。

表 2 在文档数据集 CDR 的实验结果

Table 2 Experimental results on the document dataset CDR %

模型	P	R	F_1
GCN ^[36]	52.3	72.0	60.6
Ensemble ^[37]	64.9	49.3	56.0
GCN-MEM ^[38]	55.7	68.1	61.3
Graph-Kernel ^[39]	55.6	68.4	61.3
Bi-SAN ^[10]	55.6	70.8	62.1
Stack-BiLSTM ^[40]	45.2	68.1	54.3
ScispaCy ^[41]	53.7	63.3	58.1
EoG ^[17]	62.7	66.3	64.5
Two-Phase-BERT ^[52]	61.9	68.7	65.1
Seq2rel ^[48]	68.2	66.2	67.2
BERT-GLRE ^[32]	65.1	72.2	68.5
BERT-GSAN	80.7	74.2	77.3

注: 加粗表示该列最优结果。

表 3 在 DocRED 数据集上的实验结果

Table 3 Experimental results on the DocRED dataset %

模型	验证集				测试集	
	Ign F_1	Ign AUC	F_1	AUC	Ign F_1	F_1
BiLSTM ^[8]	48.47	47.61	50.94	50.26	48.78	51.06
Context-Aware ^[8]	48.94	47.22	51.09	50.17	48.40	50.70
GAT ^[31]	45.17	—	51.44	—	47.36	49.51
GCNN ^[11]	46.22	—	51.52	—	49.59	51.62
EoG ^[17]	45.94	—	52.15	—	49.48	51.82
AGGCN ^[42]	46.29	—	52.47	—	48.89	51.45
BERT ^[52]	—	—	54.16	—	—	53.20
RoBERTa ^[52]	53.85	48.27	56.05	51.35	53.52	55.77
Two-Phase-BERT ^[52]	—	—	54.42	—	—	53.92
BERT-HIN ^[20]	54.29	—	56.31	—	53.70	55.60
CorefBERT ^[53]	55.32	—	57.51	—	54.54	56.96
BERT-GLRE ^[32]	—	—	—	—	55.40	57.40
BERT-LSR ^[22]	52.43	—	59.00	—	56.97	59.05
RSMAN ^[54]	57.22	—	59.25	—	57.02	59.29
GRACR ^[21]	57.85	—	59.73	—	56.47	58.54
BERT-GAIN ^[23]	57.56	55.60	59.78	58.74	57.00	59.46
BERT-GSAN	58.37	54.56	60.58	57.79	58.11	60.64

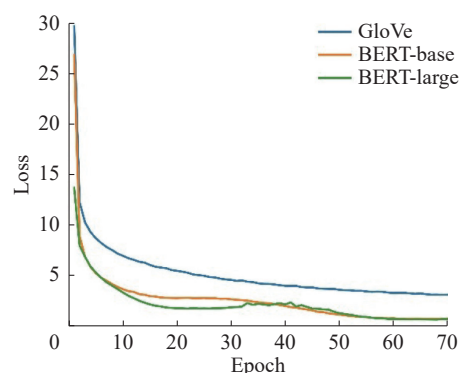
注: 加粗表示该列最优结果。

1)从表中可以观察到 BERT-GASN 在两个数据集上相对于基准模型均得到了最优结果。因为本文模型与现有最优模型^[32]差异在于将图推理嵌入动态自注意力模块代替了该方法中的实体候选级图聚合模块,因此该结果说明图推理网络中引入动态关键词有助于提升图推理的性能。从表 2 可以观察到, BERT-GSAN 模型相较于 BERT-GLRE^[32] F_1 的分数提升了 8.8%, 这表明使用实体候选节点、文档节点和动态关键词节点构建的文档图,可以有效提升文档中的实体关系聚合推理的能力,同时也可以有效地提升实体关系分类的效果。表 3 中通过去除训练集、验证集以及测试集中的共同关系事实计算出的 Ign F_1 分数较其他基准模型有更优的效果,避开 BERT-GSAN 模型已经学习过的知识,从而更加准确反映模型性能。

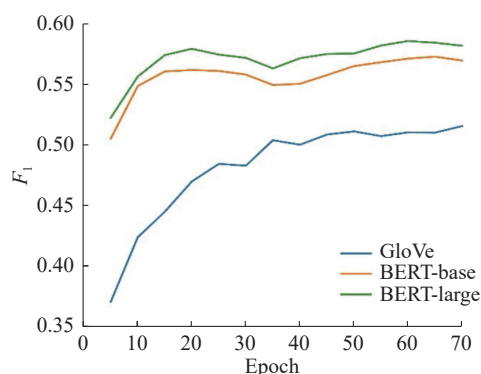
2)文档级关系抽取方法的性能总体上要优于句子级关系抽取方法,这说明了直接应用句子级模型到文档级关系抽取任务存在的局限性,原因在于句子级方法不能够利用全局背景语义信息进行关系预测,同时也不能够预测跨句子的实体关系。注意到,相比于 BERT-GAIN,文中所提 BERT-GSAN 构建带有动态关键词节点的文档图,可实现文档中复杂推理建模的能力,例如, BERT-GSAN 中提出的实体与动态关键词的关系,该关键词在关系图推理过程中可快速获取文档中全局背景语义信息;文档与动态关键词的关系,增加该关系与关键词没有直接连接的实体候选也可以通过文档节点关注到文档中的全局背景语义信息。

3.5 应用不同预训练模型的对比实验

为评估不同词向量编码器对模型性能的影响,基于 DocRED 数据集,分别用 GloVe、BERT-base 和 BERT-large 的输出作为词向量表示输入到 GASN 模块,固定词向量表示模块,在训练中观察不同周期(Epoch)下模型在训练集上的损失(Loss)和验证集上文档关系预测性能 F_1 分数。实验结果如图 3 所示,由图 3(a)可以看出, BERT 预训练模型较 GloVe 预训练模型损失收敛速度更快,表明采用预训练语言模型 BERT 作为文本编码器的模型都要显著优于基于传统的词向量模型,体现了 BERT 在文档级关系抽取任务中的强大预测能力。图 3(b)中显示 BERT-large 模型的 F_1 分数远高于 GloVe 和 BERT-base 模型。结果表明基于图神经网络的模型其训练的收敛速度对于词向量编码器具有很强依赖性,优质的词向量表征可以加速模型的收敛,同时提升图神经网络的预测能力。



(a) 不同预训练模型损失曲线



(b) 不同预训练模型 F_1 曲线

图 3 不同预训练模型的对比实验

Fig. 3 Comparative experiments with different pre-trained models

4 结束语

本文提出了图推理嵌入的动态自注意力方法,其通过动态注意力机制实现对文档级长文本序列实现高效背景语义表示学习,同时在模块中嵌入图推理模块以获得文档级图结构和推理信息。构建了基于实体候选、文档节点、动态关键词等 3 种类型节点的图网络,并采用多头图注意力网络获取图中蕴涵的潜在关系推理信息。在文档级关系抽取数据集 DocRED 和 CDR 上实验结果表明,在图推理模型中引入动态关键词作为全局背景语义信息,有助于提升图推理的性能。此外,本文对未来的研究工作提出了展望。1)噪声的文档级关系抽取:本文研究的模型是建立在人工标注的文档级关系抽取数据集上,认为文档标注的实体和关系是无误的,未来可以考虑借助知识图谱的远程监督的方式标注数据的关系抽取方法,研究解决远程监督数据存在噪声情况下抗噪的关系抽取方法。2)研究基于 Transformer-Decoder 架构的预训练语言模型的文档级关系抽取方法:目前基于 Transformer-Decoder 的自回归模型架构已经成为大语言模型的主流架构,本文目

前研究的是基于 Transformer-Encoder 架构 BERT 模型, 后续可以研究在自回归架构下文档级关系抽取, 结合 Prompt 工程、指令学习等方式。

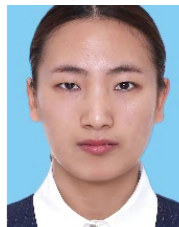
参考文献:

- [1] YU Mo, YIN Wenpeng, HASAN K S, et al. Improved neural relation detection for knowledge base question answering[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver: Association for Computational Linguistics, 2017: 571–581.
- [2] DAS R, MUNKHDALAI T, YUAN Xingdi, et al. Building dynamic knowledge graphs from text using machine reading comprehension[EB/OL]. (2018–10–12) [2023–10–01]. <http://arxiv.org/abs/1810.05682v1>.
- [3] QIN Pengda, XU Weiran, WANG W Y. Robust distant supervision relation extraction via deep reinforcement learning[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne: Association for Computational Linguistics, 2018: 2137–2147.
- [4] ZENG Daojian, LIU Kang, LAI Siwei, et al. Relation classification via convolutional deep neural network[C]//Proceedings of COLING 2014–25th International Conference on Computational Linguistics. Dublin: ACL, 2014: 2335–2344.
- [5] ZENG Xiangrong, HE Shizhu, LIU Kang, et al. Large scaled relation extraction with reinforcement learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2018: 5658–5665.
- [6] ZHOU Wenxuan, CHEN Muhao. An improved baseline for sentence-level relation extraction[C]//Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing. Stroudsburg: ACL, 2022: 161–168.
- [7] HAN Xu, LIU Zhiyuan, SUN Maosong. Neural knowledge acquisition via mutual attention between knowledge graph and text[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2018: 4832–4839.
- [8] YAO Yuan, YE Deming, LI Peng, et al. DocRED: a large-scale document-level relation extraction dataset[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019: 764–777.
- [9] SHEN Tao, ZHOU Tianyi, LONG Guodong, et al. DiSAN: directional self-attention network for RNN/CNN-free language understanding[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2018: 5446–5455.
- [10] VERGA P, STRUBELL E, MCCALLUM A. Simultaneously self-attending to all mentions for full-abstract biological relation extraction[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. New Orleans: Association for Computational Linguistics, 2018: 872–884.
- [11] SAHU S K, CHRISTOPOULOU F, MIWA M, et al. Inter-sentence relation extraction with document-level graph convolutional neural network[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019: 4309–4316.
- [12] XU Wang, CHEN Kehai, MOU Lili, et al. Document-level relation extraction with sentences importance estimation and focusing[C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: Association for Computational Linguistics, 2022: 2920–2929.
- [13] 吴婷, 孔芳. 基于图注意力卷积神经网络的文档级关系抽取[J]. 中文信息学报, 2021, 35(10): 73–80.
- [14] WU Ting, KONG Fang. Document-level relation extraction based on graph attention convolutional neural network[J]. Journal of Chinese information processing, 2021, 35(10): 73–80.
- [15] PENG Nanyun, 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.
- [16] QUIRK C, POON H. Distant supervision for relation extraction beyond the sentence boundary[C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. Stroudsburg: Association for Computational Linguistics, 2017: 1171–1182.
- [17] 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. Palo Alto: AAAI, 2019: 6513–6520.
- [18] CHRISTOPOULOU F, MIWA M, ANANIADOU S. Connecting the dots: document-level neural relation extraction with edge-oriented graphs[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2019: 4925–4936.
- [19] ZHUANG Yimeng, WANG Huadong. Token-level dynamic self-attention network for multi-passage reading comprehension[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2019: 2252–2262.
- [20] SONG Linfeng, WANG Zhiguo, YU Mo, et al. Explor-

- ing graph-structured passage representation for multi-hop reading comprehension with graph neural networks [EB/OL]. (2018-9-06)[2023-10-01]. <http://arxiv.org/abs/1809.02040v1>.
- [20] TANG Hengzhu, CAO Yanan, ZHANG Zhenyu, et al. HIN: hierarchical inference network for document-level relation extraction[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020: 197-209.
- [21] LIU Hongfei, KANG Zhao, ZHANG Lizong, et al. Document-level relation extraction with cross-sentence reasoning graph[M]//Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2023: 316-328.
- [22] NAN Guoshun, GUO Zhijiang, SEKULIC I, et al. Reasoning with latent structure refinement for document-level relation extraction[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2020: 1546-1557.
- [23] ZENG Shuang, XU Runxin, CHANG Baobao, et al. Double graph based reasoning for document-level relation extraction[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2020: 1630-1640.
- [24] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2017: 6000-6010.
- [25] 李祥宇, 隋璘, 熊伟丽. 基于自注意力机制与卷积 ONLSTM 网络的软测量算法[J]. 智能系统学报, 2023, 18(5): 957-965.
- LI Xiangyu, SUI Lin, XIONG Weili. Soft sensor algorithm based on self-attention mechanism and convolutional ONLSTM network[J]. CAAI transactions on intelligent systems, 2023, 18(5): 957-965.
- [26] DEVLIN J, CHANG Mingwei, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of NAACL- HLT. Stroudsburg: ACL, 2019: 4171-4186.
- [27] HU Minghao, PENG Yuxing, HUANG Zhen, et al. Reinforced mnemonic reader for machine reading comprehension[C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. Stockholm: International Joint Conferences on Artificial Intelligence Organization, 2018: 4099-4106.
- [28] SHAW P, USZKOREIT J, VASWANI A. Self-attention with relative position representations[EB/OL]. (2018-03-06)[2023-12-01]. <http://arxiv.org/abs/1803.02155v2>.
- [29] YANG Baosong, WANG Longyue, WONG D, et al. Convolutional self-attention network[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: ACL, 2019: 4040-4045.
- [30] TAY Y, DEGHANI M, BAHRI D, et al. Efficient transformers: a survey[J]. ACM computing surveys, 2023, 55(6): 1-28.
- [31] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. (2017-10-30)[2023-10-01]. <http://arxiv.org/abs/1710.10903v3>.
- [32] WANG Difeng, HU Wei, CAO Ermei, et al. Global-to-local neural networks for document-level relation extraction[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2020: 3711-3721.
- [33] LI Jiao, SUN Yueping, JOHNSON R J, et al. BioCreative V CDR task corpus: a resource for chemical disease relation extraction[J]. Database, 2016, 2016: baw068.
- [34] LOSHCILOV I, HUTTER F. Decoupled weight decay regularization[EB/OL]. (2017-11-14)[2023-10-01]. <http://arxiv.org/abs/1711.05101v3>.
- [35] LEE J, YOON W, KIM S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining[J]. Bioinformatics, 2020, 36(4): 1234-1240.
- [36] ZHANG Yuhao, QI Peng, MANNING C D. Graph convolution over pruned dependency trees improves relation extraction[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2018: 2205-2215.
- [37] ZHOU Huiwei, DENG Huijie, CHEN Long, et al. Exploiting syntactic and semantics information for chemical-disease relation extraction[J]. Database, 2016, 2016: baw048.
- [38] GU Jinghang, SUN Fuqing, QIAN Longhua, et al. Chemical-induced disease relation extraction via convolutional neural network[J]. Database, 2017, 2017: bax024.
- [39] PANYAM N C, VERSPOOR K, COHN T, et al. Exploiting graph kernels for high performance biomedical relation extraction[J]. Journal of biomedical semantics, 2018, 9(1): 7.
- [40] ZHENG Wei, LIN Hongfei, LI Zhiheng, et al. An effective neural model extracting document level chemical-induced disease relations from biomedical literature[J]. Journal of biomedical informatics, 2018, 83: 1-9.
- [41] LI Jing, WANG Yequan, ZHANG Shuai, et al. Rethinking document-level relation extraction: a reality check[C]//Findings of the Association for Computational Linguistics: ACL 2023. Stroudsburg: Association for Computational Linguistics, 2023: 5715-5730.
- [42] GUO Zhijiang, ZHANG Yan, LU Wei. Attention guided graph convolutional networks for relation extraction[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2023: 5715-5730.

- ation for Computational Linguistics, 2019: 241–251.
- [43] 陈容珊, 高淑萍, 齐小刚. 注意力机制和图卷积神经网络引导的谱聚类方法[J]. 智能系统学报, 2023, 18(5): 936–944.
- CHEN Rongshan, GAO Shuping, QI Xiaogang. A spectral clustering based on GCNs and attention mechanism[J]. CAAI transactions on intelligent systems, 2023, 18(5): 936–944.
- [44] 朱金霞, 孟祥福, 邢长征, 等. 融合图卷积注意力机制的协同过滤推荐方法[J]. 智能系统学报, 2023, 18(6): 1295–1304.
- ZHU Jinxia, MENG Xiangfu, XING Changzheng, et al. Collaborative filter recommendation approach fused with graph convolutional attention mechanism[J]. CAAI transactions on intelligent systems, 2023, 18(6): 1295–1304.
- [45] 赵鹏武, 李志义, 林小琦. 基于注意力机制和卷积神经网络的中文人物关系抽取与识别[J]. 数据分析与应用, 2022, 6(8): 41–51.
- ZHAO Pengwu, LI Zhiyi, LIN Xiaoqi. Chinese character relation extraction and recognition based on attention mechanism and convolution neural network[J]. Data Analysis and knowledge discovery, 2022, 6(8): 41–51.
- [46] 张鲁, 段友祥, 刘娟, 等. 基于 RoBERTa 和加权图卷积网络的中文地质实体关系抽取[J]. 计算机科学, 2024, 51(8): 297–303.
- ZHANG Lu, DUAN Youxiang, LIU Juan, et al. Chinese geological entity relation extraction based on RoBERTa and weighted GCNs[J]. Computer science, 2024, 51(8): 297–303.
- [47] 袁泉, 陈昌平, 陈泽, 等. 基于 BERT 的两次注意力机制远程监督关系抽取[J]. 计算机应用, 2024, 44(4): 1080–1085.
- YUAN Quan, CHEN Changping, CHEN Ze, et al. Twice attention mechanism distantly supervised relation extraction based on BERT[J]. Journal of computer applications, 2024, 44(4): 1080–1085.
- [48] GIORGI J, BADER G D, Wang Bo. A sequence-to-sequence approach for document-level relation extraction [EB/OL]. (2022–04–03) [2023–10–01]. <http://arxiv.org/abs/2204.01098>.
- [49] 文坤建, 陈艳平, 黄瑞章, 等. 基于提示学习的生物医学关系抽取方法[J]. 计算机科学, 2023, 50(10): 223–229.
- WEN Jiankun, CHEN Yanping, HUANG Ruizhang, et al. Biomedical relationship extraction method based on prompt learning[J]. Computer science, 2023, 50(10): 223–229.
- [50] 赵晋斌, 王琦, 马黎雨, 等. 基于知识图谱的远程监督关系抽取降噪方法[J]. 火力与指挥控制, 2023, 48(10): 160–169.
- ZHAO Jinbin, WANG Qi, MA Liyu, et al. A noise reduction method for distant supervision relation extraction based on knowledge graph[J]. Fire control and command control, 2023, 48(10): 160–169.
- [51] 曾碧卿, 李砚龙, 蔡剑. 基于外部知识增强的远程监督关系抽取模型[J]. 计算机应用系统, 2023, 32(5): 253–261.
- ZENG Biqing, LI Yanlong, CAI Jian. Distantly-supervised relation extraction model via external knowledge enhancement[J]. Computer systems and applications, 2023, 32(5): 253–261.
- [52] WANG Hong, FOCKE C, SYLVESTER R, et al. Fine-tune bert for DocRED with two-step process[EB/OL]. (2019–09–26)[2023–10–01]. <http://arxiv.org/abs/1909.11898v1>.
- [53] YE Deming, LIN Yankai, DU Jiaju, et al. Coreferential reasoning learning for language representation[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2020: 7170–7186.
- [54] YU Jiaxin, YANG Deqing, TIAN Shuyu. Relation-specific attentions over entity mentions for enhanced document-level relation extraction[EB/OL]. (2022–05–28)[2023–10–01]. <http://arxiv.org/abs/2205.14393v1>.

作者简介:



李云洁, 硕士研究生, 主要研究方向为自然语言处理、机器学习。E-mail: 2621991259@qq.com。



王丹阳, 研究实习生, 主要研究方向为智能信息处理。E-mail: danyang.wang@catas.cn。



刘海涛, 副教授, 博士, 主要研究方向为自然语言处理、机器学习、因素空间理论。中国运筹学会模糊信息与工程分会理事。曾获市科技进步一等奖 1 项, 市级自然科学学术成果奖特等奖、一等奖各 1 项。以第一作者身份发表学术论文 30 余篇, 出版专著《因素空间与人工智能》获国家出版基金资助, 并被评为“十三五”国家重点图书出版规划重大出版工程项目。E-mail: haitao641@163.com。