

基于注意力机制语义增强的文档级关系抽取

柳先辉¹, 吴文达¹, 赵卫东^{1,2}, 侯文龙¹

(1. 同济大学 电子与信息工程学院, 上海 201804; 2. 上海视觉感知与智能计算工程技术研究中心, 上海 200092)

摘要: 文档级关系抽取旨在从文档中抽取多个实体对之间的关系, 具有较高的复杂性。针对文档级关系抽取中的多实体、关系相关性、关系分布不平衡等问题, 提出了一种基于注意力机制(Attention)语义增强的文档级关系抽取方法, 能够实现实体对之间关系的推理。具体来说, 首先在数据编码模块改进编码策略, 引入更多实体信息, 通过编码网络捕获文档的语义特征, 获得实体对矩阵; 然后, 设计了一个基于 Attention 门控机制的 U-Net 网络, 对实体对矩阵进行局部信息捕获和全局信息汇总, 实现语义增强; 最后, 使用自适应焦点损失函数缓解关系分布不平衡的问题。在 4 个公开的文档级关系抽取数据集 (DocRED、CDR、GDA 和 DWIE) 上评估了 Att-DocuNet 模型并取得了良好的实验结果。

关键词: 文档级关系抽取; 注意力机制; 语义增强; 焦点损失
中图分类号: TP391 **文献标志码:** A

Document-Level Relation Extraction Method Based on Attention Semantic Enhancement

LIU Xianhui¹, WU Wenda¹, ZHAO Weidong^{1,2}, HOU Wenlong¹

(1. College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China; 2. Shanghai Visual Perception and Intelligent Computing Engineering Technology Research Center, Shanghai 200092, China)

Abstract: Document-level relation extraction aims to extract the relations between multiple entity pairs from a document, a task characterized by high complexity. This paper proposes a method for document-level relation extraction based on attention semantic enhancement to address challenges such as handling multiple entities, capturing relationship correlations, and dealing with imbalanced relationship distributions within documents. The method proposed facilitates the inference of relationships between entity pairs. Specifically, the data

encoding module enhances the encoding strategy by incorporating additional entity information, capturing semantic features of the document through the encoding network, and generating an entity pair matrix. Subsequently, a U-Net network employing an attention gating mechanism is devised to capture local information and aggregate global information from entity pair matrices, thereby achieving semantic enhancement. Finally, this paper introduces an adaptive focal loss function to mitigate imbalanced relationship distributions. The Att-DocuNet model proposed is evaluated on four publicly available document-level relation extraction datasets (DocRED, CDR, GDA, and DWIE), yielding promising experimental results.

Keywords: document-level relation extraction; attention mechanism; semantic enhancement; focal loss

关系抽取任务是信息抽取领域中一项重要的子任务, 它对于知识库的构建、文本的理解等都有着重要的作用, 也广泛应用于一些自然语言处理中的课题, 如知识问答^[1]、知识图谱构建等^[2]。关系抽取现有的大部分工作都是针对单个句子, 旨在抽取单个句子中多个实体的对应关系, 也已经取得了很好的效果。然而, 有研究表明, 现实生活中的很多关系都是隐藏在多个句子中的^[3]。因此, 文档级关系抽取问题越来越受到研究人员的关注。

相较于单句的关系抽取, 文档级关系抽取面临着更多的问题。首先是多实体问题, 文档中存在多个实体且每个实体可能多次出现, 模型需要聚合相同实体的多维度提及, 还要识别并关注每个实体的相关上下文。第二是关系类别分布不均衡问题, 大多数实体对之间没有关系, 即使存在关系也是几种常见的关系类别。第三是实体关系的相关性问题,

收稿日期: 2022-11-29

基金项目: 国家重点研发计划(2020YFB1709303)

第一作者: 柳先辉, 副教授, 工学博士, 主要研究方向为自然语言学习。E-mail: lxh@tongji.edu.cn

通信作者: 吴文达, 硕士生, 主要研究方向为知识图谱。E-mail: 2130766@tongji.edu.cn



论文
拓展
介绍

实体对之间的关系可能存在联系,部分实体关系三元组需要通过逻辑推理得到。举个例子,已知“IBM实验室,位于,南美洲”和“IBM实验室,位于,圣保罗”2个三元组,根据这2个关系三元组的相关性可以推理出“圣保罗,位于,南美洲”。注意到关系的相关性问题并加以利用,可以帮助模型实现更好的关系抽取。

注意到多个关系之间存在相关性,DocuNet^[4]将文档级关系抽取描述为一个语义分割问题,并在矩阵特征图上使用U-Net^[5]模块来捕捉三元组之间的全局相互依赖关系。应该注意的是,关系相关性问题仅涉及到小部分实体,它使得这些实体间的关系存在依赖而变得更加难以抽取。因此,为了更好地解决关系相关性问题,本文提出Attention-Unet语义增强模块,在U-Net模块中进一步添加注意力门控机制,帮助捕捉局部特征,使得模型能够更好地捕获全局和局部信息,提升模型针对局部的实体进行复杂关系抽取的能力。

同时,上述的方法都没有针对多实体问题进行改进。Zhou和Chen^[6]已经验证了引入实体类型信息可以提升句子级关系抽取的效果。因此,本文引入带有实体类型和实体id信息的实体标记方法,帮助模型更好地匹配同一实体的不同提及,利用实体类型丰富实体信息,从而获得更完整的上下文语义向量。

文档级关系抽取中还存在关系类别不平衡问题,大部分实体对之间不包含关系,包含关系时也大多是几种常见的关系类别。本文引入了一个自适应焦点损失函数,能够平衡关系类别的权重,缓解关系分布不平衡的影响。

1 相关工作

以往的关系抽取方法主要集中于识别句子中2个实体之间的关系。许多方法^[7-9]能够有效处理句子层面的关系抽取任务。然而,句子级的关系抽取面临着不可避免的问题,那就是许多现实世界的关系只能通过阅读多个句子来抽取,因此,文档级关系抽取任务成为了许多研究人员的研究重点。

文档级关系抽取的各种方法主要包括基于图的方法和基于Transformer的方法。基于图的方法因为其在关系推理中的有效性,在文档级关系抽取中被广泛采用。Jia等^[10]提出了一个结合了不同文本跨度的表示的模型,包括文档级和句子级的表示。

Christopoulou等^[11]提出了一个基于边的图神经网络模型(EoG),利用多种类型的节点和边聚合全局信息,通过基于边的推理抽取出实体间的关系。Nan等^[12]基于EoG做了一些改进,引入元依赖路径节点,并且自动地更新学习文档图的结构,实现关系推理。Li等^[13]用图形增强的双重注意网络(GEDA)实现关系实例和句子的信息交互,利用图卷积网络进行信息增强。Zhang等^[14]提出了一种具有双层异构图(DHG)的模型,利用结构建模层对文档进行编码,关系推理层实现多跳推理。Zhou等^[15]提出了一种全局上下文增强图卷积网络(GCGCN),以实体作为节点,实体对的上下文作为节点之间的边,应用多头注意力生成边的加权图。Wang等^[16]提出了一种模型(GLRE),该模型学习全局和局部实体表示,结合上下文关系表示进行关系分类。Zeng等^[17]提出了具有双图的图聚合和推理网络(GAIN),分别构造提及级别和实体级别的图。Xu等^[18]提出了一种编码器—分类器—重建器模型(HeterGSAN),该模型利用重建器从图形结构中建模路径依赖关系。图形结构对文档中存在的依赖关系进行多跳路径的显式建模,提升了可解释性,减轻长距离依赖问题,可以获得较好的模型性能。

相比之下,Transformer架构利用大量的自注意力机制隐式地建模全局依赖关系。一些研究人员尝试了放弃图形结构,直接利用预训练的语言模型进行文档级关系抽取。Wang等^[19]使用BERT(Bidirectional Encoder Representation from Transformers)进行文档数据编码,并提出一个2步训练方法,注意到关系分布不平衡的问题,并利用模型更强的推理能力实现关系识别和分类。Tang等人^[20]提出了一种分层推理网络(HIN)分别对实体级、句子级、文档级的向量表示进行编码,实现不同粒度的信息聚合与推理。Zhou等^[21]针对BERT提出增强策略,针对关系标签多分类问题和多实体问题,使用自适应阈值和局部上下文池实现了较好的关系抽取效果(ATLOP)。然而,以往的研究大多集中在局部的实体表示上,更多地关注局部句子信息而忽略了三元组之间的全局连接和信息交互,没有很好地解决多个关系之间的相关性问题。

2 问题定义

给定包含一组实体 $\{e_i\}_{i=1}^n$ 的文档 D ,文档级关

系抽取任务是预测实体对 $(e_s, e_o)_{s,o \in \{1 \dots n\}, s \neq o}$ 之间的关系类型,其中 e_s 和 e_o 分别表示头实体和尾实体。关系集定义为 $R \cup \{NR\}$,其中NR表示无关系。一个实体可能在文档中出现多次,因此对于每个实体 e_i ,可以被多次提及 $\{m_j\}_{j=1}^{N_i}$ 。如果实体对 (e_s, e_o) 之间不存在关系,则将其标记为NR。在测试期间,将会预测所有实体对 $(e_s, e_o)_{s,o \in \{1 \dots n\}, s \neq o}$ 的分类标签。本质上,这是一个多标签分类问题,因为 e_s 和 e_o 之间可能存在多种关系。

3 模型

如图1所示,本文提出的Att-DocuNet模型主要分为3层结构:向量编码层、语义增强层、关系分类层。向量编码层对输入数据进行上下文编码,获得实体对矩阵;语义增强层对矩阵进行特征提取,增强模型推理能力;最后,关系分类层针对实体对向量获得关系分类结果。

3.1 向量编码层

使用预训练模型作为编码器。将文档 D 视为包含 l 个单词的序列, x_t 表示第 t 个单词。在实体的编码时添加实体类型信息可以筛选出不可能的关系,例如,2个类型为PER的实体不可能存在“位于”的关系。因此,本文改进实体标记模块,引入实体类型信息。在每个实体提及的开始和结束位置分别插入 $\langle e_i \rangle$ 和 $\langle e_{id} \rangle$,其中 e_i 表示实体类型, e_{id} 表示实体id。

这样一方面添加实体类型信息可以补充相关信息,提高关系抽取性能;另一方面,在文档中一个实体可能多次提及,利用实体id信息可以辅助对齐同一实体中不同提及的信息。

然后,将标记处理后的文档输入预训练语言模型,获得文档中每一个单词 x_i 的上下文向量表示 h_i 。

$$H = [h_1, \dots, h_l] = B([x_1, \dots, x_l]) \quad (1)$$

式中: B 表示使用BERT模型进行编码; H 表示文档中向量表示的集合。

在这之后,使用 $\langle e_i \rangle$ 和 $\langle e_{id} \rangle$ 的向量的平均值来表示提及的向量,记为 h_{m_j} 。

针对实体 e_i ,假设其有 N_i 个提及,第 j 个提及表示为 m_j 。参考Zhou等^[21]研究成果,通过logsumexp运算获得其向量表示 h_{e_i} 。

$$h_{e_i} = \lg \sum_{j=1}^{N_i} e^{h_{m_j}} \quad (2)$$

更进一步地,针对每一个实体对 (e_s, e_o) 构造一个实体级关系矩阵 $F(e_s, e_o)$ 。先计算一个基于实体对的上下文表示 $c_{s,o}$,它表示实体 e_s 和 e_o 共同关注的文档中的上下文。在实体对上下文表示 $c_{s,o}$ 的基础上根据仿射变换得到 D 维特征向量 $F(e_s, e_o)$,其计算式为

$$a_{s,o} = f_{\text{softmax}}(A_s * A_o) \quad (3)$$

$$c_{s,o} = H a_{s,o} \quad (4)$$

$$F(e_s, e_o) = W_F c_{s,o} \quad (5)$$

式中: f_{softmax} 表示softmax函数计算; $A_s(A_o)$ 是指实体

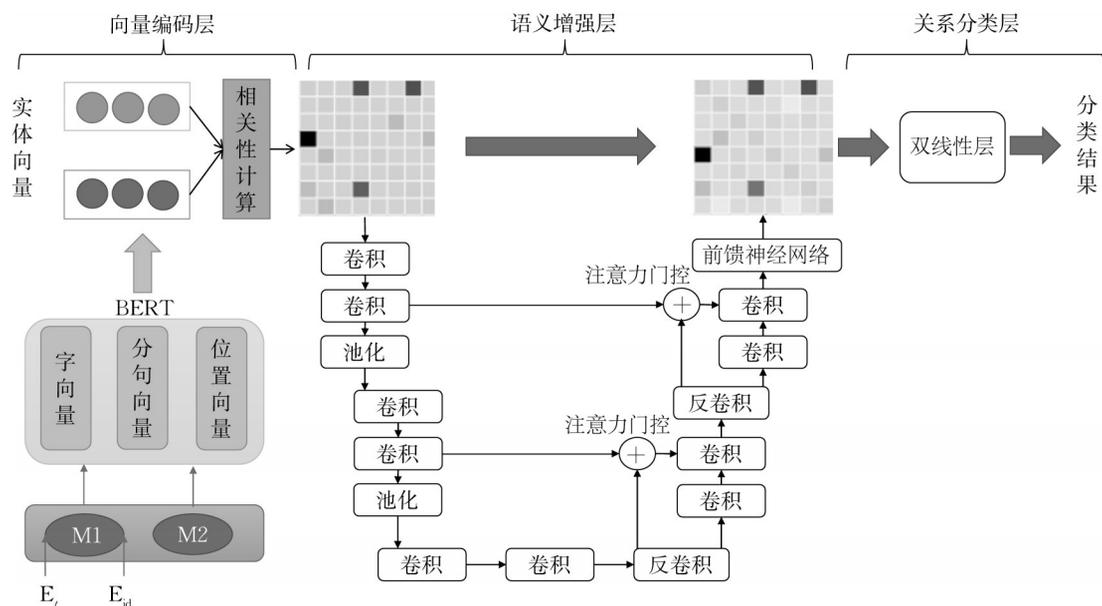


图1 Att-DocuNet模型架构

Fig. 1 Architecture of Att-DocuNet

$e_s(e_o)$ 对文档中每个单词的注意力得分; *表示逐元素相乘; H 是文档的嵌入向量; W_F 是可学习的权重矩阵。

3.2 语义增强层

将实体级关系矩阵 $F \in \mathbb{R}^{N \times N \times D}$ 视为一幅 D 通道图像,进一步将文档级关系抽取作为 F 中的像素级分类问题。其中, N 是从所有数据集样本中计算的最大实体数。为此,本文利用语义分割中的经典模型 U-Net,并添加 Attention Gate 优化分类效果,整合成为 Attention-Unet 语义增强模块。实体级关系矩阵中的每一个元素就是实体间的关系分类特征, Attention-Unet^[22] 可以通过隐式推理促进实体对之间的信息交换。如图1所示,该模块由1个U形结构网络组成,其中包含2个下采样块和2个上采样块,通过跳跃连接和注意力门控进行拼接。一方面,每个下采样块后有2个卷积模块和1个最大池化模块,特征矩阵通过下采样后通道数增加1倍。卷积模块可以提取特征,下采样块可以扩大当前实体对向量的感受野,从而获取更丰富的全局信息。另一方面,模型使用1个反卷积模块和2个卷积模块组成上采样块,通过上采样块还原特征尺寸,减少通道数,提取信息,实现语义信息的汇聚和交互。此外,上采样模块的输入通过跳跃连接和注意力门控进行拼接,借助注意力门控机制可以抑制输入特征矩阵中的不相关区域,同时突出特定局部区域的显著特征,让模型更好地关注重要的实体对。

总而言之,本文使用了一个 Attention-Unet 语义增强模块,以捕获局部和全局信息 Y ,并使用注意力门控机制,帮助模型更好地进行关系分类,如式(6):

$$Y = A_U(W_U F) \quad (6)$$

式中: A_U 表示使用 Attention-Unet 语义增强模块进行编码; $Y \in \mathbb{R}^{N \times N \times D_1}$ 为实体级关系矩阵; W_U 是为了降低 F 的维数而得到的可学习权重矩阵。 D_1 远小于 D 。

3.3 关系分类层

给定实体对嵌入 e_s 和 e_o 与实体级关系矩阵 Y ,使用前馈神经网络将它们映射到隐藏表示 z 。然后,通过双线性函数得到关系的概率。具体计算公式为

$$z_s = \tanh(W_s e_s + Y_{s,o}) \quad (7)$$

$$z_o = \tanh(W_o e_o + Y_{s,o}) \quad (8)$$

$$P(r | e_s, e_o) = \sigma(z_s W_r z_o + b_r) \quad (9)$$

式中: $Y_{s,o}$ 是矩阵 Y 中 (s, o) 的实体对表示; $W_r \in \mathbb{R}^{d \times d}$, $b_r \in \mathbb{R}$, $W_s \in \mathbb{R}^{d \times d}$ 和 $W_o \in \mathbb{R}^{d \times d}$ 都是可学习的参数。

针对类别不平衡问题,使用自适应焦点损失函数^[23]。损失函数包括两部分,第一部分是正类,第二部分是负类。在训练过程中,标签空间被划分为2个子集:正类子集 P_T 和负类子集 N_T 。正类子集 P_T 包含实体对 (e_s, e_o) 中存在的关系,如果 (e_s, e_o) 之间没有关系,则 P_T 为空 ($P_T = \emptyset$)。另一方面,负类子集 N_T 包含不属于正类的关系类, $N_T = R \setminus P_T$ 。每个正类的概率计算式为

$$P(r_i | e_s, e_o) = \frac{e^{f_{r_i}^{(s,o)}}}{e^{f_{r_i}^{(s,o)}} + e^{f_{r_{TH}}^{(s,o)}}} \quad (10)$$

为简单起见, $P(r_i | e_s, e_o)$ 表示为 $P(r_i)$, 因为只讨论 (e_s, e_o) 。对于负类,使用其 Logit 计算阈值类 (TH) 的概率,如式(11):

$$P(r_{TH} | e_s, e_o) = \frac{e^{f_{r_{TH}}^{(s,o)}}}{\sum_{r_j \in N_T \cup \{TH\}} e^{f_{r_j}^{(s,o)}}} \quad (11)$$

同样, $P(r_{TH} | e_s, e_o)$ 被称为 $P(r_{TH})$ 。由于正面标签的分布高度不平衡,利用焦点损失的概念来平衡正面类别的 Logit。最终的损失函数为

$$L_{RE} = \sum_{r_i \in P_T} (1 - P(r_i))^\gamma \lg(P(r_i)) + \lg(P(r_{TH})) \quad (12)$$

其中 γ 是一个超参数。这一损失函数是为了把更多的注意力放在分数较低、更难分的类上。如果 $P(r_i)$ 较低,则相关类别的损失权重将更高,从而可以更好地优化长尾类别。

4 实验

4.1 数据集介绍

在4个公开的文档级关系抽取数据集上对模型进行评估。数据集统计数据如表1所示。

(1) DocRED^[3] 是一个用于文档级关系抽取的大规模数据集。它由维基百科中的文章构建而成,分别包含3 053、1 000、1 000 篇文档用于训练、验证、测试。

(2) CDR^[24] 是一个生物医学领域的关系抽取数据集。它主要用于推断化学成分和疾病概念之间的关系,有500篇训练文档。

(3) GDA^[25] 是一个生物医学领域的大型数据集。它主要用于推断基因和疾病概念之间的关系,有23 353个训练样本。

(4) DWIE^[26] 是一个公共的多任务数据集,可用于信息提取中的4项子任务,本文主要应用于文档

级关系抽取任务。原始数据集共802篇文档,经清洗分别得602、98、99篇文档用于训练、验证、测试。

表1 数据集统计信息

Tab. 1 Statistics of dataset

指标	DocRED	CDR	GDA	DWIE
训练集	3 053	500	23 353	602
验证集	1 000	500	5 839	98
测试集	1 000	500	1 000	99
关系类别数	97	2	2	65
每篇文档实体数	19.5	7.6	5.4	27.4
每个实体提及数	1.4	2.7	3.3	1.98

4.2 实验设置

本文模型基于Pytorch实现,在DocRED和DWIE上使用cased-BERT-base^[27]作为编码器,在CDR和GDA上使用SciBERT-base^[28]作为编码器。使用AdamW^[29]优化模型,在前6%的步骤中进行线

性预热^[30]并将学习率设置为 2×10^{-5} 。将实体级矩阵大小 N 设置为42,并基于训练集调整超参数。在一个NVIDIA RTX3060 12GB GPU上进行了训练,并使用Ign F_1 和 F_1 分数评估模型。

4.3 在DocRED数据集上的实验结果

实验结果如表2所示。在DocRED数据集上,选择以下2种类型算法模型作为基线模型进行对比。

(1)基于图的模型:这类模型先对文档建立图结构,再通过GCN进行推理。主要对比的方法包括GEDA^[13]、LSR^[12]、HeterGSAN^[18]、GAIN^[17]。

(2)基于Transformer的模型:这类模型直接使用预训练模型BERT进行文档级关系抽取。对比的方法包括BERT^[19]、HIN-BERT^[20]、ATLOP^[21]、DocuNet^[4]。

表2 DocRED数据集的结果对比

Tab. 2 Comparison of results on DocRED dataset

模型	验证集		测试集	
	Ign F_1	F_1	Ign F_1	F_1
GEDA-BERT _{base}	54.52	56.16	53.17	55.74
LSR-BERT _{base}	52.43	59.00	56.97	59.05
HeterGSAN-BERT _{base}	58.13	60.18	57.12	59.45
GAIN-BERT _{base}	59.14	61.22	59.00	61.24
BERT _{base}		54.16		53.20
HIN-BERT _{base}	54.29	56.31	53.70	55.60
ATLOP-BERT _{base}	59.22	61.09	59.31	61.30
DocuNet-BERT _{base}	59.86	61.83	59.93	61.86
Att-DocuNet-BERT _{base}	60.35	62.28	60.42	62.32

本文模型在训练集和测试集中分别达到62.28% F_1 和62.32% F_1 ,这分别优于DocuNet模型0.45% F_1 和0.46% F_1 。与基于图的方法的最先进模型GAIN模型相比,Att-DocuNet在训练集上超过它1.06% F_1 ,在测试集上超过它1.08% F_1 。这证明本模型在实体对层次上的推理能力比之前方法更强。

4.4 在生物医学数据集上的实验结果

在生物医学数据集CDR和GDA上,本文将模型与众多基线模型进行对比,包括: EoG^[11]、LSR^[12]、DHG^[14]、GLRE^[16]、ATLOP^[21]和DocuNet^[4]。在2个生物医学数据集上的实验结果如表3所示。

本文模型在2个数据集上分别达到76.8% F_1 和86.1% F_1 ,相较于DocuNet模型提升了0.5%和0.8%,与基于图的模型GLRE相比,在CDR上的性能提升了8.3%。这些结果证明了模型在生物医学领域的良好性能和通用性。

4.5 在DWIE数据集上的实验结果

在DWIE数据集上,选择的基线模型包括:

表3 CDR和GDA数据集的结果对比

Tab. 3 Comparison of results on CDR and GDA dataset

模型	CDR	GDA
EoG	63.6	81.5
LSR	64.8	82.2
DHG	65.9	83.1
GLRE	68.5	
SciBERT _{base}	65.1	82.5
ATLOP-SciBERT _{base}	69.4	83.9
DocuNet-SciBERT _{base}	76.3	85.3
Att-DocuNet-SciBERT _{base}	76.8	86.1

GAIN、BERT、ATLOP、DocuNet。实验结果如表4所示。

本文模型在训练集和测试集中分别达到69.20% F_1 和71.40% F_1 ,相较于DocuNet模型提升了1.30%和1.36%,与基于图的方法的最先进模型GAIN模型相比,Att-DocuNet在训练集上超过它5.82% F_1 ,在测试集上超过它3.83% F_1 。这更全面地证明了本模型在文档级关系抽取任务上的性能优

表4 DWIE数据集的结果对比
Tab. 4 Comparison of results on DWIE dataset

模型	验证集		测试集	
	Ign F_1	F_1	Ign F_1	F_1
GAIN-BERT _{base}	55.63	62.55	62.37	67.57
BERT _{base}	58.40	63.38	62.92	69.12
ATLOP-BERT _{base}	59.03	64.48	62.09	69.94
DocuNet-BERT _{base}	61.88	67.90	64.23	70.04
Att-DocuNet-BERT _{base}	63.91	69.20	65.49	71.40

越性。

4.6 消融实验

为了进一步验证每一个模块的有效性,在DocRED数据集上进行了消融实验,实验结果如表5所示。表中,w/o Entity Type表示不添加实体类型信息;w/o Attention表示在语义分割模块不添加注意力门控单元,仅使用U-net模块;w/o Focal Loss表示将损失函数替换为DocuNet中的自适应阈值损失 Adaptive Loss。

从表5可以看出,w/o Entity Type导致 F_1 下降0.33%,这说明实体类型引入了有效的信息,帮助模型更正确地分类。w/o Attention导致 F_1 下降0.40%,这说明注意力机制语义增强模块能够充分捕获局部依赖信息,帮助模型更好地推理。w/o Focal Loss导致 F_1 下降0.29%,这说明自适应焦点损失函数可以平衡常见关系类别与非常见关系类别的权重,减轻关系分布不均的影响。

表5 消融实验结果

Tab. 5 Results of ablation experiment

模型	Ign F_1	F_1
Att-DocuNet-BERT _{base}	60.35	62.28
w/o Entity Type	60.01	61.95
w/o Attention	59.96	61.88
w/o Focal Loss	60.03	61.99

5 结语

针对文档级关系抽取任务提出Att-DocuNet模型,改进实体标记模块为模型添加实体类型信息,并提出Attention-Unet语义增强模块,引入焦点自适应损失函数,改善文档级关系抽取中的关系相关问题、类别不平衡问题,实现文档级关系信息的全局汇聚与局部捕获。实验结果表明,与基线模型相比,本文模型可以获得更好的性能。未来希望将本方法应用于其他的分类任务,例如嵌套命名实体识别、多模态关系抽取、远程监督关系抽取等任务。

作者贡献声明:

柳先辉:设计框架、技术指导、论文审定。

吴文达:实验研究、论文撰写。

赵卫东:技术指导、论文审定。

侯文龙:技术指导。

参考文献:

- [1] HAO Y, LIU X, WU J, *et al.* Exploiting sentence embedding for medical question answering[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Honolulu: AAAI Press, 2019: 938-945.
- [2] JI S, PAN S, CAMBRIA E, *et al.* A survey on knowledge graphs: representation, acquisition, and applications[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(2): 494.
- [3] YAO Y, YE D, Li P, *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.
- [4] ZHANG N, CHEN X, XIE X, *et al.* Document-level relation extraction as semantic segmentation[C]//IJCAI. Montreal: International Joint Conferences on Artificial Intelligence Organization, 2021: 3999-4006.
- [5] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-assisted Intervention. Cham: Springer, 2015: 234-241.
- [6] ZHOU W, CHEN M. An improved baseline for sentence-level relation extraction[C]//AAACL-IJCNLP 2022. [s. l.]: Association for Computational Linguistics, 2022: 161-168.
- [7] CHEN X, ZHANG N, XIE X, *et al.* Knowprompt: Knowledge-aware prompt-tuning with synergistic optimization for relation extraction[C]//Proceedings of the ACM Web Conference 2022. New York: Association for Computing Machinery, 2022: 2778-2788.
- [8] ZHENG H, WEN R, CHEN X, *et al.* PRGC: Potential relation and global correspondence based joint relational triple 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 1: Long Papers). [s. l.] : Association for Computational Linguistics, 2021; 6225-6235.
- [9] WEI Z, SU J, WANG Y, *et al.* A novel cascade binary tagging framework for relational triple extraction [C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. [s. l.] : Association for Computational Linguistics, 2020; 1476-1488.
- [10] JIA R, WONG C, POON H. Document-Level N-ary relation extraction with multiscale representation learning [C]// Proceedings of NAACL-HLT. Minneapolis: Association for Computational Linguistics, 2019; 3693-3704.
- [11] 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 (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019; 4925-4936.
- [12] NAN G, GUO Z, Sekulić 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. [s.l.] : Association for Computational Linguistics, 2020; 1546-1557.
- [13] LI B, YE W, SHENG Z, *et al.* Graph enhanced dual attention network for document-level relation extraction [C]// Proceedings of the 28th International Conference on Computational Linguistics. Barcelona: International Committee on Computational Linguistics, 2020; 1551-1560.
- [14] ZHANG Z, YU B, SHU X, *et al.* Document-level relation extraction with dual-tier heterogeneous graph[C]//Proceedings of the 28th International Conference on Computational Linguistics. Barcelona: International Committee on Computational Linguistics, 2020; 1630-1641.
- [15] ZHOU H, XU Y, YAO W, *et al.* Global context-enhanced graph convolutional networks for document-level relation extraction [C]//Proceedings of the 28th International Conference on Computational Linguistics. Barcelona: International Committee on Computational Linguistics, 2020; 5259-5270.
- [16] WANG D, HU W, CAO E, *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 (EMNLP). [s.l.] : Association for Computational Linguistics, 2020; 3711-3721.
- [17] ZENG S, XU R, CHANG B, *et al.* Double graph based reasoning for document-level relation extraction [C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). [s.l.] : Association for Computational Linguistics, 2020; 1630-1640.
- [18] XU W, CHEN K, ZHAO T. Document-level relation extraction with reconstruction [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: Assoc Advancement Artificial Intelligence, 2021, 35 (16) : 14167-14175.
- [19] WANG H, FOCKE C, SYLVESTER R, *et al.* Fine-tune Bert for DocRED with two-step process [J]. arXiv e-prints, 2019; arXiv: 1909.11898.
- [20] TANG H, CAO Y, ZHANG Z, *et al.* Hin: Hierarchical inference network for document-level relation extraction [C]// Pacific-Asia Conference on Knowledge Discovery and Data Mining. [s.l.]: Springer, 2020; 197-209.
- [21] ZHOU W, HUANG K, MA T, *et al.* Document-level relation extraction with adaptive thresholding and localized context pooling [C]//Proceedings of the AAAI conference on artificial intelligence. Palo Alto: Assoc Advancement Artificial Intelligence, 2021, 35(16): 14612-14620.
- [22] OKTAY O, SCHLEMPER J, LE FOLGOC L, *et al.* Attention U-Net: Learning where to look for the pancreas [J]. arXiv e-prints, 2018; arXiv: 1804.03999.
- [23] TAN Q, HE R, BING L, *et al.* Document-level relation extraction with adaptive focal loss and knowledge distillation [C]//Findings of the Association for Computational Linguistics: ACL 2022. Dublin: Association for Computational Linguistics, 2022; 1672-1681.
- [24] LI J, SUN Y, JOHNSON R J, *et al.* BioCreative V CDR task corpus: a resource for chemical disease relation extraction [J]. Database, 2016, 1: 10.
- [25] WU Y, LUO R, LEUNG H, *et al.* Renet: A deep learning approach for extracting gene-disease associations from literature [C]//International Conference on Research in Computational Molecular Biology. [s.l.] : Springer, 2019; 272-284.
- [26] ZAPOROJETS K, DELEU J, DEVELDER C, *et al.* DWIE: An entity-centric dataset for multi-task document-level information extraction [J]. Information Processing & Management, 2021, 58(4): 102563.
- [27] DEVLIN J, CHANG M W, LEE K, *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding [J]. arXiv e-prints, 2018; arXiv: 1810.04805.
- [28] BELTAGY I, LO K, COHAN A. SciBERT: A pretrained language model for scientific text [C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019; 3615-3620.
- [29] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization [C]//7th International Conference on Learning Representations (ICLR). New Orleans: [S.n.], 2019; 1-8.
- [30] GOYAL P, DOLLAR P, GIRSHICK R, *et al.* Accurate, large minibatch SGD: training ImageNet in 1 Hour [J]. arXiv e-prints, 2017; arXiv: 1706.02677.