计算机科学
COMPUTER SCIENCE

融合证据句子提取的文档级关系抽取

安先跨, 肖蓉, 杨肖

引用本文

安先跨, 肖蓉, 杨肖. 融合证据句子提取的文档级关系抽取[J]. 计算机科学, 2024, 51(6A): 230800081-6.

AN Xiankua, XIAO Rong, YANG Xiao. Document-level Relation Extraction Integrating Evidence Sentence Extraction [J]. Computer Science, 2024, 51(6A): 230800081-6.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230800081

融合证据句子提取的文档级关系抽取

安先跨　肖　蓉　杨　肖
湖北大学计算机与信息工程学院　武汉 430062
(202221116012800@stu.hubu.edu.cn)

摘　要　文档级关系抽取作为自然语言处理领域的一个关键任务,旨在从长文档中准确抽取实体对之间的语义关系。传统的文档级关系抽取方法通常将整个文档作为输入,但事实上,人类只需根据文档中的部分句子即可预测实体对的关系,即证据句子。在现有研究中,很多研究方法都利用了证据句子,但是都存在无法找全以及很难充分利用这些证据句子的优势等问题。针对该问题,引入更加高效且准确的证据句子选取方法,通过融合公式法和删句法的证据句子提取策略,并将证据提取与训练推理过程相融合,使得文档级关系抽取模型更加关注重要的句子,同时仍可以识别文档中的完整信息。实验表明,改进后的模型在公共数据集上的表现优于已有模型。

关键词:文档级;关系抽取;证据句子;双线性层
中图分类号　TP391

Document-level Relation Extraction Integrating Evidence Sentence Extraction

AN Xiankua,XIAO Rong and YANG Xiao
School of Computer and Information Engineering,Hubei University,Wuhan 430062,China

Abstract　As a crucial task in the field of natural language processing,document-level relation extraction aims to accurately extract semantic relationships between entities from lengthy documents. Traditional document-level relation extraction methods typically take the entire document as input. However,in reality,humans can predict relationships between entity pairs based on only a portion of the document,referred to as evidence sentences. In existing research,many methods start to utilize evidence sentences,but they face challenges such as incomplete evidence retrieval and difficulty in fully leveraging the advantages of these evidence sentences. To address this issue,we introduce a more efficient and accurate evidence sentence selection method. This is achieved by integrating a strategy for extracting evidence sentences through a fusion of formula-based and sentence-deletion-based approaches. We seamlessly integrate the evidence extraction with the training and inference processes,directing the document-level relation extraction model to focus more on crucial sentences while still recognizing comprehensive information within the document. Experimental results demonstrate that the improved model outperforms existing models on public datasets.

Keywords　Document-level,Relation extraction,Evidence sentences,Bilinear layer

1　引言

随着信息的爆炸性增长,从大规模文本中提取有意义的关系信息变得越来越重要,文档级关系抽取是自然语言处理领域的一项重要任务,它旨在从给定的文档中识别实体之间的关系,并进行关系分类。文档级关系抽取在许多领域具有广泛的应用,如信息抽取、问答系统、知识图谱构建等。通过自动化地从大规模文本中提取关系信息,可以帮助人们理解和利用文本中蕴含的知识,从而推动信息处理和决策支持的发展。

近几年,文档级关系抽取取得了飞速发展,一方面得益于RNN[1],LSTM[2],GRU[3]等神经网络的应用。多对实体之间可能存在复杂的关系网络,神经网络可以充分捕捉到整个文档中对实体关系的上下文信息,提升模型在长文档中对多实体关系的识别能力。另一方面,在长文档中,关系信息可能被其他无关信息干扰或淹没,使得关系抽取的准确性下降。注意力机制[4]的运用,使得模型更加关注文本中的关键信息,忽略干扰信息,减少噪音,提高关系抽取的性能。除此之外,大规模人工标注的数据集 DocRed[5]的出现提高了模型的泛化能力和性能,大大加快了文档级关系抽取的研究进度。

尽管先前的研究中涉及了各种关系抽取技术,也有很多学者尝试在关系抽取中提取证据句子,如 haung 等的 E2GRE[6]模型首次联合提取关系和底层证据句子,但它们在充分利用证据句子方面仍然存在一些局限性,证据句子作为包含关系信息的重要线索,可以提供更全面准确的关系信息。然而,先前的研究未能充分利用证据句子的优势,导致关系抽取模型在多实体关系识别和长文档处理方面的性能有限。本研究旨在填补这一研究空白,在原有关系抽取模型的基础上

基金项目:科技大数据湖北省重点实验室(中国科学院武汉文献情报中心)开放基金课题资助项目(E1KF291005)
This work was supported by the Hubei Key Laboratory of Big Data in Science and Technology(Wuhan Library of Chinese Academy of Science)(E1KF291005).
通信作者:肖蓉(20040363@hubu.edu.cn)

改进证据句子提取方法，以充分利用文档的上下文信息提取证据句子，从而提高关系抽取的准确性和召回率。

具体来说，我们注意到 Xie 等的 Eider 模型[7]以及 Xu 等的 SIEF 插件[8]。Eider 模型是本文的的基线模型，此模型通过双线性函数来评估句子的重要性，得出句子对实体的相关得分，从中选出相关性得分高的句子作为证据句子。SIEF 提出了一个新的句子重要性估计和聚焦框架，以激励模型更加关注证据句子，以此来预测实体之间的关系。首先把包含此句和不包含此句子的文档分别输入到关系抽取模型中，然后根据两文档输出实体关系概率之间的差异来评估每个句子的重要性。当一个句子被删除时，关系预测的关联概率前后差距较大，则通常表明该句子是此关系的证据句子，随后 SIEF 又提出一个辅助损失，以激励模型无论是将整个文档作为输入还是将删除非证据句子文档作为输入都能产生相同的输出分布。本文的创新点在于参考了这两种证据句子提取方法，根据 Eider 模型所提供的思路初步提取出证据句子，将此方法记为公式法，随后依据 SIEF 插件所提供的思路对证据句子做进一步的增强，分配权值，此方法记为删句法，后面章节会详细介绍该方法的具体步骤。

融合这两种方法的优势体现在其能够更好地捕捉文档级关系的上下文信息，这两种方法分别关注不同的特征和模型架构，融合后的方法能够综合两者的优势，更全面地考虑实体关系在文档中的分布和上下文语境。这使得模型在文档级关系抽取中能够更准确地理解和刻画文档里的实体关系。其优势还体现在模型的鲁棒性和泛化能力方面。通过融合这两种方法，模型能够对多样性的文档进行更好的适应，并更好地处理文档中的噪声和复杂关系，使模型在应对真实场景中的文档级关系抽取任务时更具有优势。除此之外，通过学习不同证据句子的权重然后将它们的特征结合起来，可以更好地利用两种方法的优点，提高关系分类的准确性和稳定性。实验表明，相对于基线模型，运用该融合方法可以使结果在公共测试集上的 F1 值提高 0.4。

本文第 2 章将分别从深度神经网络技术的更替和证据句子的使用两方面阐述文档级关系抽取的发展历程；第 3 章将详细介绍本模型的方法步骤，首先阐述文档级关系抽的公共做法，随后详细说明本模型的两个证据句子提取方法；第 4 章将从两个方面的对比实验证明本模型的有效性；最后总结全文。

2 相关工作

近年来，关系抽取的研究重点已从句子级别转移到文档级别。根据 Zhu 等的文档级关系抽取研究综述[9]，最初的文档级关系抽取方法主要基于循环神经网络(RNN)架构，如长短时记忆网络(LSTM)和双向长短时记忆网络(BiLSTM)。这些神经网络能够更好地提取长文档中实体之间的远程交互关系，从而识别更丰富的文档语义信息，如 Cai 等便是运用双向长短期神经网络 BiLSTM 进行关系抽取[10]。随后，一批学者对此类方法进行改进，如 Yang 等融合胶囊网络改进的 BSRU-ATTCapsNet 模型[11]。随着技术的更新，RNN 架构逐渐被 Vaswani 等提出的 Transformer[12]架构所替代，该类方法主要利用 Transformer 和 BERT[13]等模型对文档进行编码，如 Zhou 等[14]便使用 Transformer 架构对文档中实体以及其实体提及之间的复杂关系交互进行建模，提取实体信息，生成实体表示，最后根据实体表示预测实体之间的关系。图神经网络的快速发展使得图被广泛用于文档级关系抽取。在此应用中，文档中的单词、提及、实体或句子等信息被表示为节点，而语法知识、共指、邻接、共现等启发式规则被视为边。通过这种方式，文档被建模为一个文档图，并作为模型的输入。借助文档图中边的信息传递，节点信息得以聚合，从而捕捉文档中不同实体对之间的关系。如 Christopoulou 等首次使用异构图，提出了用于文档级关系抽取的 Eog 模型[15]，随后 Xu 等提出了重构图推理路径、减少模型的错误推理路径的 HeterGSAN 模型[16]。亦有 Zeng 等通过构建文本图将 GNN 应用于关系抽取任务，从而提出了构建双图机制的 GAIN 模型[17]，其考虑了多种类型的交互信息，并且利用了注意力机制进行路径推理，使得关系抽取的准确率大大提高。

尽管以上方法在关系抽取中取得很大成就，但事实上，人类只需根据文档中的部分句子即可预测实体对的关系，即证据句子。Huang 等[18]首次提出只需通过实体的支持证据句子集，即可预测目标实体之间的关系，而不需要整篇文档信息。随后 Xie 等[7]认为，将整篇文档作为输入，会引入很多无关信息，增加噪音，降低模型性能。为使模型更加关注与实体对有关的信息，他们提出了证据增强的关系抽取模型。随后有学者将证据句子用于推理，例如 Wang 等[19]提出证据感知注意力机制提取证据特征，根据证据特征进行推理，学习实体对表示，并对其关系进行分类。尽管很多学者开始逐渐运用证据句子，但是有关证据句子的提取没有统一方法，而且现有方法的效果参差不齐。本文旨在寻找一种最优的证据句子提取方法，以便在关系抽取中提高分类效果。本文所提方法将作为基线模型的增强框架，模型通过删句法确定证据句子，然后通过双线性函数的公式法识别句子重要性，并将两者提取出的证据句子加权融合，以激励模型关注证据句子，提高关系抽取模型的鲁棒性和整体性能。

3 方法

本章将详细描述本模型方法，模型的框架概述如图 1 所示。首先，为了方便交流以及清晰地描述问题，本文将在 3.1 节简要地介绍文档级关系抽取过程和过程中的公式表达，随后在 3.2 节中描述公式法的双线性函数是如何评估句子的重要性，又是如何选出重要的证据句子。然后，在 3.3 节中，描述模型的删句法是如何通过删除某个句子，根据所删除句子对分类结果的影响来确定句子的重要性。最后在 3.4 节将详细讲述这两种方法的融合以及证据句子在模型训练和推理中的运用。

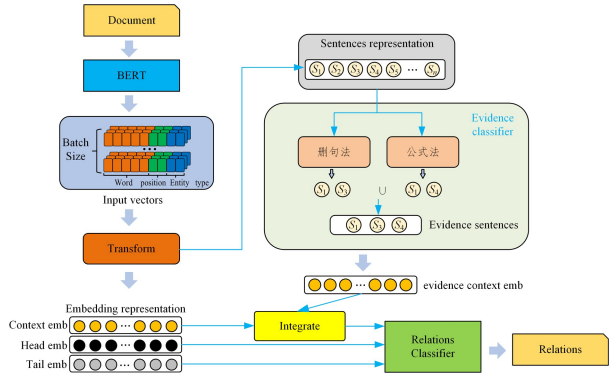

图 1 模型框架

Fig. 1 Frameworkof the proposed model

3.1 问题定义

文档级关系抽取(DocRe)是一项复杂的任务，每位学者都有不同的公式表达。在这一节，为了统一符号的表示，规定文档级关系抽取的输入是包含 n 个句子的文档 $\boldsymbol{D}=\{s_1,s_2,\cdots,s_n\}$，其中每个句子 $\boldsymbol{s}_i$ 是由单词的上下文表示 $\boldsymbol{g}$ 组成的序列，单词的上下文表示 $\boldsymbol{g}$ 由单词编码、位置信息编码以及实体类型的编码融合后经过序列神经网络生成。

$$\boldsymbol{g}_i=[E_w(w);E_l(l);E_t(t)] \tag{1}$$

其中，$E_w(\cdot)$，$E_l(\cdot)$，$E_t(\cdot)$分别是单词编码层、单词位置编码层、实体类型编码层；$\boldsymbol{w},\boldsymbol{l},\boldsymbol{t}$ 分别是单词编码、位置编码、实体类型编码，如果单词位置不是实体的标记位置则为 None。在人工标注的大型数据集 DocRed 中，实体的位置由人工标注，实体的表示是实体所在单词上下文表示的平均值，具体来说，若某个实体在文档中的位置是 $a-b$，则实体的表示为：

$$\boldsymbol{m}_i=\frac{1}{b-a}\sum_{k=a}^{b}g_k \tag{2}$$

但是在一个文档中一种实体通常会多次出现，因此通过 LogSumExp 函数池化所有出现的实体表示，获得此实体的最终表示。

$$\boldsymbol{e}_i=\log\sum_k\exp(\boldsymbol{m}_k^i) \tag{3}$$

其中，k 为实体出现的次数，即实体提及的数量。所以，一个文档级关系抽取模型可以被认为是多关系分类标签模型。

$$P_{ij}=DocRe(c_{ht},e_{i_h},e_{i_5})=\Pr[r_{ht}=1|D,e_{i_h},e_{i_5},e_{i_5}] \tag{4}$$

其中，e_{i_5} 为头实体，e_{i_5} 为尾实体，r_{ht} 是两实体是否具有关系的真值标签。然后模型根据输入提取出实体对之间关系的上下文表示 c_{ht}，分类器通过阈值化概率预测给定实体对的关系，此部分内容遵循之前大多数学者的研究工作。

3.2 公式法

在本模型中我们使用公式法预测某个句子 s_i 是否为实体对 (e_h,e_t) 的证据句子。与实体编码类似，为了获得句子编码 s_i，对 s_i 的所有单词的上下文表示应用 LogSumExp 池化：

$$s_i=\log\sum_k g_k^i \tag{5}$$

此处 k 为句子的单词数量，由此可以得到 s_i 的向量表示，如果 s_i 是 (e_h,e_t) 的证据句，则 s_i 中的实体标记将与关系预测相关，并且对 c_{ht} 贡献更多。因此，模型使用关系上下文编码 c_{ht} 和句子编码 s_i 之间的双线性函数来衡量句子 s_i 对实体对 (e_h,e_t) 的重要性得分：

$$P(s_i|e_h,e_t)=\sigma(s_i\boldsymbol{W}_v c_{ht}+\boldsymbol{b}_v) \tag{6}$$

其中，$\boldsymbol{W}_v$ 和 $\boldsymbol{b}_v$ 均是可以训练的参数矩阵，得分高的句子即被视作证据句子。

由于一个实体对可能有多个证据句子，因此我们使用二元交叉熵作为损失函数来训练证据提取模型：

$$l_{evi1}=-\sum_{h\neq t,NA\notin P_{h,t}^T}\sum_{s_i\in D}y_i\cdot P(s_i|e_h,e_t)+(1-y_i)\cdot\log(1-P(s_i|e_h,e_t)) \tag{7}$$

其中，$h\neq t$，$NA\notin P_{h,t}^T$ 表示不为同一个且具有一定关系的实体对。另外当 s_i 为证据句子时，证据标签 $y_i=1$，否则为 0。

3.3 删句法

该方法预测句子重要性评分的原理是，对于已经训练好的模型，给予两种不同的输入，一种是删除了某个句子的输入文档，另一种是原输入文档。比较这两者的预测概率，如果某个句子被删除，导致预测的关联概率降低，那么这个句子很可能就是证据句子。如果预测的概率没有改变，那么这个句子很可能是非证据句子。此外，当一个句子被删除时，预测的概率有时会增加，在这种情况下，DocRe 模型不是鲁棒的，因为这违反了单调性。

形式上，每次只考虑删除一个句子，并且将删除第 m 个句子的文档表示为：

$$\boldsymbol{D}(-m)=\{s_1,\cdots,s_{m-1},s_{m+1},\cdots,s_n\} \tag{8}$$

对于 DocRE 模型，模型基于式(4)得到原始文档的分类概率 P_{ij} 以及去掉句子 m 后的分类概率：

$$\begin{aligned}P_{ij}^{(-m)}&=DocRe(c_{ht},e_{i_h},e_{i_5})\\&=\Pr[r_{ht}=1|D(-m),e_{i_h},e_{i_5},e_{i_5}]\end{aligned} \tag{9}$$

我们将重要性评分公式设计为：

$$h_{ij}^{(-m)}=P_{ij}\log\frac{p_{ij}}{P_{ij}^{(-m)}} \tag{10}$$

这样设计公式的目的一方面是为了结果更加稳健，另一方面是为了模型结果的单调，消除鲁棒性。对于阈值超参数 β，如果 $h_{ij}^{(-m)}>\beta$，就将句子 m 视为证据句子。经过多次实验，模型选定 $\beta=0.8$，同样使用二元交叉熵作为损失函数来训练证据提取模型：

$$l_{evi2}=-\sum_{h\neq t,NA\notin P_{h,t}^T}\sum_{s_i\in D}y_i\cdot h_{ij}^{(-m)}+(1-y_i)\cdot\log(1-h_{ij}^{(-m)}) \tag{11}$$

其中，$h\neq t$，$NA\notin P_{h,t}^T$ 表示不为同一个且具有一定关系的实体对。另外当被删除的句子 s_m 为证据句子时，证据标签 $y_i=1$，否则为 0。

3.4 融合方法抽取证据句子

通过实验发现两种方法找到的证据句子有很大比例的重合，但也有各自的优点，公式法更能关注到证据句子少、语义重合度高的实体，通过公式计算这类句子通常会取得较高的句子重要性得分。相对而言，若某对关系的证据句子较多，语义重合度少，通过双线性函数得出的句子相关性得分则较低。而删句法是根据删除某个句子所带来的影响来确定证据句子的重要性，这类方法可以识别出更准确的证据句子，但是每个证据句子的重要性很难确定，因此结合两种方法，利用后者来提取更为准确的证据句子，然后再利用前者给证据句子赋予重要性。

假设通过以上两种方法提取的证据句子集分别为 s_{evi1} 和 s_{evi2}，因为两个集合有重复证据句子，所以取两者的并集作为最终的证据句子集：$s_{evi}=s_{evi1}\cup s_{evi2}$。

但是每个证据对实体关系的支撑并不是同等重要，因此我们使用一个带有注意力机制的证据选择器来分配每个证据句子的权值，以便模型可以选择性地关注证据句子里的关系信息，通过对证据句子里关系信息的提取得到实体关系的上下文表示 $c_{ht_{evi}}$：

$$P(s_{evi_i}|e_h,e_t)=\sigma(\boldsymbol{W}\cdot[e_h,e_t,s_{evi_i}]) \tag{12}$$

$$\alpha_i=\frac{P(s_{evi_i}|e_h,e_t)}{\sum_l P(s_{evi_l}|e_h,e_t)} \tag{13}$$

$$\boldsymbol{c}_{ht_{evi}}=\sum_i^l\alpha_i\boldsymbol{s}_{evi_i} \tag{14}$$

其中，$\boldsymbol{W}$ 为可训练参数矩阵，σ 是一个 Sigmoid 函数，此处 l 是证据句子的数量。

假设提取的证据句子已经包含了与实体关系相关的所有信息，那么就不需要使用整个文档进行关系提取。然而，没有

一个系统可以完美地提取证据而不遗漏任何句子。仅仅依赖提取的证据可能会错过文档中的重要信息，导致性能不佳。因此，我们将原始文档和提取的证据的预测结果结合起来。通常证据句子产生的实体关系上下文表示含有更多的关系信息，通过大量实验，模型设定 0.2 的倍率从全文获取全文信息，以弥补证据关系里的信息缺失，将得到的结果作为最终的关系上下文表示，并通过式(4)分类得出实体之间的关系：

$$c_{ht_{evi}} = 0.2 \cdot c_{ht} + c_{ht_{evi}} \tag{15}$$

最终将两种证据句子选取方法的 loss 结合，即 $l_{evi1} + l_{evi2}$，并与关于抽取模型分类的 loss 融合，协同训练。

在推理中使用证据句子，具体而言，如图 1 所示，模型首先从原始文档通过关系抽取层获得一组关系预测分数 $S^{(O)}_{h,t,r}$。然后，模型按照原始文档中出现的顺序将提取的证据句子 s_{evi} 连接起来，为每个实体对构建一个伪文档 $d'_{h,t}$。DocRe 模型对证据文档的预测得分记为 $S^{(E)}_{h,t,r}$。最后，通过一个混合层聚合两组预测值来融合结果：

$$P_{\text{Fuse}}(r|e_h, e_t) = \sigma(S^{(O)}_{h,t,r} + S^{(E)}_{h,t,r} - \tau) \tag{16}$$

之所以选择这种设计是因为它简单，并且只包含一个可学习的参数 τ，从而减轻了开发集中的过拟合。对于参数 τ，我们采用以下损失函数进行优化：

$$l_{\text{Fuse}} = -\sum_{d\in D}\sum_{h\neq t}\sum_{r\in R} y_r \cdot P_{\text{Fuse}}(r|e_h, e_t) + (1-y_r) \cdot \log(1 - P_{\text{Fuse}}(r|e_h, e_t)) \tag{17}$$

如果关系 r 在 (e_h, e_t) 之间成立，则 $y_r = 1$，否则 $y_r = 0$。大量实验表明，使用其他损失函数对性能影响不大。

4 实验结果与分析

4.1 实验数据

DocRed[5] 是由清华大学推出的大型人工标注的文档级数据集，是目前唯一提供证据句子标签作为注释的一部分的数据集。数据集本身涵盖多个领域，根据相关维基百科的统计，至少有 40.7% 的关系事实只能从多个句子中抽取，这也说明目前 DocRed 是文档级关系抽取最权威的数据集，同时该数据集配备官方验证平台，能让各位学者的实验得到统一且准确的验证。

4.2 参数设置

本模型是基于 PyTorch 和 Huggingface 的 Transform (Wolf 等，2019)[20] 实现的。参考前人的研究，本模型使用 cased-BERT-base(Devlin 等，2019)[13] 作为基本编码器，并使用 AdamW[21] 优化器优化本模型，编码器的学习率为 3×10^{-5}，其他参数的学习率为 1×10^{-4}。每批文件的数量设置为 4，关系提取和证据提取损失之间的比率设置为 0.1，两种证据句子选取方法的 loss 权重比为 1:1。模型会根据开发集上的 $F1$ 值，保存当前最优结果对应的参数节点。本模型是在武汉大学超算上进行训练及产生测试结果，训练过程的 epoch 为 40 次。

4.3 基线模型和评估指标

我们使用以下模型作为基准：

1) Huang 等提出的 E2GRE[6] 模型：通过使用大型预训练语言模型作为编码器来联合提取关系和底层证据句子。他们将文档文本与头部实体连接起来，以激励模型更专注于文档中与头部实体更相关的部分。

2) Yang 等的出的 BSRU-ATTCapsNet[11] 模型：提出融合双向简单循环网络与胶囊网络的文档级关系抽取模型。其实现了多个句子间关系融合表示，优化了学习实体关系在空间、方向等多个维度上的关系表示，提高了并行化效率。

3) Wang 等提出使用 BERT[13] 代替 BiLSTM 作为文档级关系抽取的编码器。

4) Xu 等提出的 HeterGSAN[16] 模型：一种新的编码器-分类器-重构器模型，设法从图表示中重构出真实路径依赖关系，增强推理和关系分类。

5) Zeng 等提出的 GAIN[17] 双图模型：同时构建异构提及图和异构实体图，在此基础上，提出了一种新的路径推理机制来推断实体之间的关系。

6) Xie 等提出的 Eider[7] 模型：通过有效的证据提取和推理阶段融合增强文档级关系提取。

根据之前 Yao 等[5] 的研究，我们使用 $F1$ 和 Ign $F1$ 作为关系提取的主要评估指标，其中 $F1$ 值是所预测结果关系正确率和召回率的加权平均，Ign $F1$ 得分是指忽略在训练集和开发集已有关系的 $F1$ 得分。除此之外实验还引入 Evi $F1$ 值，其表示证据句子的 $F1$ 值。本模型和 Eider 的实验结果是在本地复现所得，我们取 5 次实验的平均结果作为测试集(dev)的结果写入表中，随后取最优的一次结果的参数生成测试集对应的结果，并将得到的结果上传到 CodaLab 平台里的 DocRed 专栏得出测试集得分。其他模型的结果均是直接引用其实验结果，表格中的“—”是作者未给出或者是没有对应的结果。

4.4 实验结果与分析

为了验证本文提出的新模型在文档级关系抽取领域的有效性，本文设置了多个对比实验，分别验证本模型所提出的证据句子对推理分类阶段的影响以及去除相应模块对模型性能的影响，并对比了当下流行的模型。

4.4.1 实验 1：对比当下流行的模型

如表 1 所列，通过融合公式法和删句法两种证据句子选取方法，融合后的模型无论是在验证集(dev)还是在测试集(test)上关键性能指标 $F1$ 值均优于独立的模型 Eider 和 SIEF，相对于基线模型 Eider 而言，本模型在测试集上 F1 值有约 0.4 的提升，其他各项指标也有较大提升。相对于目前运用 SIEF 插件最优良的模型 GAIN+SIEF 而言，本模型在测试集上 F1 值也有着 0.3 的提升，体现了本模型在关系抽取上的优势。

表 1 DocRED 数据集上的实验结果

Table 1 Experiment results on DocRED

Model	Dev			Test		
	Ign $F1$	$F1$	Evi $F1$	Ign $F1$	$F1$	Evi $F1$
E2GRE	55.22	58.72	47.12	—	—	—
BSRUATTCapsNet	—	58.76	—	—	58.19	—
BERTbase+SIEF	57.13	59.11	—	57.87	58.93	—
HeterGSAN+SIEF	57.99	60.04	—	57.93	60.02	—
GAIN+SIEF	59.82	62.24	—	59.87	62.29	—
Eider	60.24	62.23	50.29	60.07	62.19	50.54
Our Model	**60.70**	**62.80**	**50.58**	**60.31**	**62.62**	**51.09**

此外证据句子的 $F1$ 值 Evi $F1$ 是验证证据句子最直接的结果，但是之前很少有人重视证据句子，致使很多模型没有 Evi $F1$ 值，这里有 E2GRE 和 Eider 两个模型参考，可见无论是在开发集还是测试集上，本模型所提取的证据句子的 $F1$ 值均优于这两种模型，体现出本模型提取的证据句子有较高

的召回率和正确率，进一步体现了本模型在文档级关系抽取上的优势。

4.4.2 实验2：消融实验

为验证所提取证据句子对推理分类阶段的影响，设置了消融实验。从表2中可以看出，推理分类阶段不使用证据句子，对结果影响较为明显，实验表明，不使用证据句子会使 $F1$ 分值下降0.3左右，这体现出使用证据句子可以提高关系抽取模型的性能。比较验证集Dev和测试集Test结果，推理分类阶段不使用证据句子对测试集上的影响更为明显些，说明本模型的证据句子提取方法更能关注到一些未训练到的关系。除此之外，实验设置了去掉删句法、公式法和两种方法均不使用等消融试验。实验结果如表2所列，无论去掉哪一种证据句子的选取方法，对模型的性能都有较大影响，去除删句法，模型F1值性能降低约1，相对而言，去除公式法性能减少较小，但F1值也下降了约0.8，可见两种方法联合使用才更能发挥出证据句子的效果。使用证据句子，不仅能够加快模型的收敛，而且可以提升模型的性能，通过式(12)－(14)，模型可以从证据句子里提取更多的关系信息，随后与模型提取的上下文关系信息融合，以此来提高关系分类的准确性和完整性，除此之外我们发现本模型在Epoch到33～36次便完成了参数的训练。为了更好地控制变量对比试验结果，表中的实验epoch均是40次，因此可以说明，相对而言本模型的证据句子提取方法可以使模型更快收敛。模型从全局文档中获取的信息通常带有噪音和干扰信息，因此推理阶段使用证据句子是为了消除模型的噪音干扰，但是又不能完全忽略全文的信息，如式(15)设置一定的倍率从全文获取信息。在关系推理中，当从证据句子里提取的关系分类得分高于模型的关系分类得分时，选用证据句子的关系作为最终关系，从表2中可以看出在推理阶段不使用证据句子时，本模型相对于基线模型在各个指标上均有较大幅度的下降。比较两个数据集的结果，验证集的结果影响较小一些，但是测试集的结果影响较大，经分析可能是模型的耦合化较高。总而言之，本模型提出的证据句子选取方法在推理阶段能够发挥应有的作用。

表2 消融实验

Table 2 Ablation experiment

Model	Dev		Test	
	Ign $F1$	$F1$	Ign $F1$	$F1$
-本模型	**60.70**	**62.80**	**60.31**	**62.62**
-推理阶段不使用证据句子	60.57	62.61	60.24	62.31
-去除删句法	59.04	60.87	59.34	61.58
-去除公式法	60.13	62.19	59.56	61.83
-去除证据句子	59.11	61.01	59.30	61.31

结束语 本文提出了一个全新的证据句子提取方法，旨在通过加强证据句子的提取性能，来增强文档级关系抽取模型的性能。该方法采用融合公式法和删句法的证据句子提取策略，充分利用了这两种方法各自的优点。具体而言，本模型以公式法为主提取证据句子，同时通过删句法为证据句子分配权值，最大化且准确地提取证据句子，随后充分利用这些提取出的证据句子的信息，并将证据提取与训练推理过程相融合，以改进文档级关系抽取的性能。

在训练阶段，关系抽取模型与证据提取模型相互提供训练信息并相互增强，从而促进了模型的学习过程。而在推理过程中，模型将原始文档与提取证据的预测结果结合起来，以鼓励模型更加关注重要的句子，同时减少信息损失。实验结果表明，本模型在数据集DocRed上的验证结果明显优于现有的证据句子提取方法。然而，我们也注意到在实验过程中，证据提取模型与关系抽取模型不仅相辅相成，而且彼此之间息息相关，互相制约。优秀的证据句子提取模型只有在优秀的关系抽取模型中才能充分发挥其应有的作用。另外，本模型的证据句子提取方法存在一定的耦合性较大等缺点。因此，希望广大学者能够充分发挥才智，不断改进现有的证据句子提取方法和关系抽取方法，为关系抽取领域贡献自己的力量。

参 考 文 献

[1] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning Internal Representations by Error Propagating [J]. Nature, 1986, 323(6088): 533-536.

[2] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.

[3] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[J]. arXiv: 1406.1078, 2014.

[4] NIU Z, ZHONG G, YU H. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48-62.

[5] 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. 2019: 764-777.

[6] HUANG K, QI P, WANG G, et al. Entity and evidence guided document-level relation extraction[C]// Proceedings of the 6th Workshop on Representation Learning for NLP (RepL4NLP-2021). 2021: 307-315.

[7] XIE Y, SHEN J, LI S, et al. Eider: Empowering Document-level Relation Extraction with Efficient Evidence Extraction and Inference-stage Fusion[C]// Findings of the Association for Computational Linguistics. 2022: 257-268.

[8] XU W, CHEN K, MOU L, 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. 2022: 2920-2929.

[9] ZHU T J, LU J C, ZHOU G, et al. A Survey of Document-level Relation Extraction Techniques[J]. Computer Science, 2023, 50(5): 189-200.

[10] CAI R, ZHANG X, WANG H. Bidirectional recurrent convolutional neural network for relation classification[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016: 756-765.

[11] YANG C N, PENG D L. A Document-level Entity Relation Extraction Model Integrating BSRU and Capsule Network[J]. Journal of Miniaturized Computer Systems, 2022, 43(5): 964-968.

[12] VASWANI A, SHAZEER N, PARMARN, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 6000-6010.

[13] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding

[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019: 4171-4186.

[14] 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. 2021: 14612-14620.

[15] 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). 2019: 4925-4936.

[16] XU W, CHEN K, ZHAO T. Document-level relation extraction with reconstruction[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2021: 14167-14175.

[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). 2020: 1630-1640.

[18] HUANG Q, ZHU S, FENG Y, 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: 998-1004.

[19] WANG H, QIN K, LU G, et al. Document-level relation extraction using evidence reasoning on RST-GRAPH[J]. Knowledge-Based Systems, 2021, 228: 107274.

[20] WOLF T, DEBUT L, SANHV, et al. Huggingface's transformers: State-of-the-art natural language processing[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. 2020: 38-45.

[21] LOSHCHILOV I, HUTTER F. Fixing weight decay regularization in adam[J]. arXiv: 1711.05101, 2019.

AN Xiankua, born in 2000, postgraduate. His main research interests include natural language processing and relation extraction.

XIAO Rong, born in 1980, Ph.D, lecturer. Her main research interests include natural language processing and relation extraction.