计算机科学
COMPUTER SCIENCE

化学物质诱导疾病关系抽取:基于证据聚焦的图推理方法

周雪阳, 傅启明, 陈建平, 陆悠, 王蕴哲

引用本文

周雪阳, 傅启明, 陈建平, 陆悠, 王蕴哲. 化学物质诱导疾病关系抽取:基于证据聚焦的图推理方法[J]. 计算机科学, 2024, 51(10): 351-361.

ZHOU Xueyang, FU Qiming, CHEN Jianping, LU You, WANG Yunzhe. Chemical-induced Disease Relation Extraction:Graph Reasoning Method Based on Evidence Focusing [J]. Computer Science, 2024, 51(10): 351-361.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230800111

化学物质诱导疾病关系抽取:基于证据聚焦的图推理方法

周雪阳[1,2]　傅启明[1,2]　陈建平[2,3]　陆　悠[1,2]　王蕴哲[1,2]
1 苏州科技大学电子与信息工程学院　江苏 苏州 215009
2 苏州科技大学江苏省建筑智慧节能重点实验室　江苏 苏州 215009
3 苏州科技大学建筑与城市规划学院　江苏 苏州 215009
(1213574782@qq.com)

摘　要　针对现有方法在挖掘化学物质与疾病之间的相互作用关系时存在过多地关注全局信息而忽略少量的证据线索和局部提及交互的问题,提出了一种基于证据聚焦的提及水平文档级关系抽取方法(Evidence Focused Mention U-shaped Network,EF-MUnet)。该方法首先基于上下文感知策略建模提及特征,并利用二维卷积捕获邻近提及之间的局部交互;其次为避免无关上下文的干扰,提出两种证据聚焦策略 ATT-EF 和 RL-EF,前者将相似度作为证据线索的衡量指标,后者基于强化学习利用延迟反馈无监督地学习最优证据提取策略;最后使用 U-net 网络捕获实体水平的全局特征,充分挖掘语义关系。实验结果表明,与已有方法相比,EF-MUnet 在生物医学数据集 CDR 上的 F1 评价指标提升了 9.7%,并且对于句间关系的抽取更具有优势。此外,在抽取药物突变相互作用的数据集 DMI 上,EF-MUnet 也取得了最高 98.6%的准确率,证明了它是一种有效的生物医学关系抽取方法并具有较好的泛化能力。

关键词:关系抽取;证据聚焦;强化学习;自注意力机制;生物医学
中图分类号　TP391.1

Chemical-induced Disease Relation Extraction:Graph Reasoning Method Based on Evidence Focusing

ZHOU Xueyang[1,2],FU Qiming[1,2],CHEN Jianping[2,3],LU You[1,2] and WANG Yunzhe[1,2]
1 Electronic and Information Engineering,Suzhou University of Science and Technology,Suzhou,Jiangsu 215009,China
2 Jiangsu Province Key Laboratory of Intelligent Building Energy Efficiency,Suzhou University of Science and Technology,Suzhou,Jiangsu 215009,China
3 School of Architecture and Urban Planning,Suzhou University of Science and Technology,Suzhou,Jiangsu 215009,China

Abstract　To address the problem of existing methods focusing too much on global information while neglecting a small amount of evidence clues and local mention interactions when mining the interaction between chemicals and diseases,a mention level document-level relation extraction method based on evidence focusing(EF-MUnet) is proposed. This method first models mention features based on context aware strategies and captures local interactions between adjacent mentions using two-dimensional convolution network. Secondly,to avoid irrelevant context interference,two evidence focusing strategies ATT-EF and RL-EF are proposed. The former uses similarity as a measure of evidence clues,while the latter uses reinforcement learning to unsupervised learn the optimal evidence extraction policy with the help of delayed feedback. Finally,U-net networks are used to capture global features at the entity level and fully explore semantic relationships. Experimental results show that compared with existing methods,EF-MUnet's F1 score improves by 9.7% on the biomedical dataset CDR,and it has more advantages in extracting inter-sentence relations. In addition,EF-MUnet achieves the highest accuracy of 98.6% on the dataset DMI for extracting interactions between drug and mutation,proving that it is an effective biomedical relation extraction method with good generalization ability.

到稿日期:2023-08-17　返修日期:2024-01-15
基金项目:国家重点研发计划(2020YFC2006602);国家自然科学基金(62102278,62072324);江苏省高等学校自然科学研究项目(21KJA520005);江苏省重点研发计划(BE2020026);江苏省研究生教育教学改革项目;江苏省研究生科研与实践创新计划项目(KYCX23_3321)
This work was supported by the National Key R&D Program of China(2020YFC2006602),National Natural Science Foundation of China(62102278,62072324),University Natural Science Foundation of Jiangsu Province(21KJA520005),Primary Research and Development Plan of Jiangsu Province(BE2020026),Postgraduate Education Reform Project of Jiangsu Province and Postgraduate Research & Practice Innovation Program of Jiangsu Province(KYCX23_3321).
通信作者:傅启明(fqm_1@mail.usts.edu.cn)

Keywords Relation extraction, Evidence focusing, Reinforcement learning, Self-attention mechanism, Biomedicine

1 引言

化学物质与疾病之间的关系在医疗保健和药物研发等生物医学研究领域发挥着重要作用，它们以一种非结构化的形式存储在大量未经发掘的生物医学文献中。通过人工标注的方式从这些非结构化文本中提取化学物质诱导疾病的关系(Chemical-induced Disease，CID)成本高昂且效率低下，无法跟上生物医学文献急剧增长的速度。近年来，利用自然语言处理技术自动提取 CID 关系受到了越来越多的关注。然而，从生物医学文献中抽取关系通常面临着句式冗余、实体提及变体多、提及之间交互复杂等问题，特别是关系事实往往由多个句子共同表达，难以通过简单的模式识别方法获取，需要模型具有理解篇章文本和逻辑推理的能力。

已有研究通常对整个文档中实体之间的交互模式建模以考虑全局信息。然而，大多数关系事实的推断或许只需要几句话[1-3]，并且仅涉及少数的实体提及，阅读整个文档并融合所有提及特征是不必要的，甚至会引入过多的无关信息。表 1 列出了在 3 个公开文档级关系抽取数据集(CDR[4]，GDA[5]，DocRED[6])中原始文档、证据文档和每个实体对所涉及句子数量的均值，分别用“@ Doc”“@ Evi”和“@ Entity pair”表示。其中，CDR 和 GDA 数据集中没有提供证据标签，本文分别随机选取 100 条数据进行人工标注，并将它们的结果用“*”标识。从表 1 中可以看出，对于大多数实体对，仅需原始文档中约 1/3 的句子就能够识别出它们之间的关系事实，而每对实体所涉及的句子也仅占整个文档的一半。具体来看，表 2 中文档包含 5 个句子，然而仅通过句子[S4]就足以识别出关系事实“Molindone-CID-Acute renal failure”，关系事实“Molindone-CID-Rhabdomyolysis”也仅被句子[S1]和句子[S4]所共同表达。此外，从表 2 中可以看出，尽管实体“Rhabdomyolysis”在句子[S1]，[S2]，[S3]和[S4]中被分别提及，但是句子[S2]和[S4]中的提及对于推理关系事实“Molindone-CID-Rhabdomyolysis”是没有帮助的。由此可见，通过编码整个文档，建模实体之间的全局交互势必会受到无关信息的干扰，并且会在一定程度上增加任务的复杂度。

表 1 文档、证据和实体对所涉及句子数量统计

Table 1 Statistics of sentences' number involved in document, evidence, and entity pair, respectively

数据集	@ Doc	@ Evi	@ Entity pair
CDR	9.4	2.5*	5.5
GDA	10.2	4.1*	6.1
DocRED	7.9	1.6	2.6

为此，本文提出了一种基于证据聚焦的提及水平生物医学文档级关系抽取方法(EF-MUnet)。首先建模提及水平特征图，使用二维卷积网络捕获邻近提及之间的局部交互，这类似于从完全特征图中提取子图表示的过程。其次，为了帮助模型聚焦少量的证据句子，设计了两种证据聚焦算法：ATT-EF 和 RL-EF。前者是一种隐式学习的证据聚焦方法，它受自注意力机制的启发，将与实体对特征相近的句子或短语作为支持证据；后者则是一种显式提取的无监督证据聚焦方法，它将对证据句子的选择视为一个序列决策的过程，即使在选择每个句子时并没有明确的监督信号，但是所选证据文档的整体效用却可以被用来评估当前序列决策的好坏。效用反馈总是延迟的，在完成对文档中所有句子的决策后被给出，这启发本文使用强化学习技术来解决这一问题。最后，借助 U-net 网络捕获深层次实体水平的全局特征，并使用多层感知机预测实体之间的语义关系。本文的工作如下：

1)提出了一种面向提及水平建模、基于证据聚焦的生物医学文档级关系抽取方法，该方法能够充分利用邻近提及交互的局部信息，并从少量的证据上下文中有效地识别实体之间的语义关系。

2)受自注意力机制启发，将与实体对特征近似的上下文作为提取它们之间关系事实的支撑线索，提出了一种细粒度证据聚焦策略 ATT-EF。

3)将证据选择建模为一个序列决策的问题，提出了一种基于强化学习的无监督证据提取方法 RL-EF，该方法无需人工监督信号，仅通过延迟的效用反馈学习到聚焦证据的最优策略。

表 2 来自 CDR 数据集的化学物质诱导疾病示例

Table 2 Example of chemical-induced diseases from CDR dataset

文档	[S1] A case of massive ***rhabdomyolysis*** following ***molindone*** administration [S2] ***Rhabdomyolysis*** is a potentially lethal syndrome that psychiatric patients seem predisposed to develop [S3] The clinical signs and symptoms, typical laboratory features, and complications of ***rhabdomyolysis*** are presented [S4] The case of a schizophrenic patient is reported to illustrate massive ***rhabdomyolysis*** and subsequent ***acute renal failure*** following ***molindone*** administration [S5] Physicians who prescribe ***molindone*** should be aware of this reaction		
关系	实体对	语义关系	证据句子
	(Molindone, Rhabdomyolysis)	CID	[1,4]
	(Molindone, Acute renal failure)	CID	[4]

2 相关工作

化学物质与疾病之间的关系在药物发现、药物毒性检测和生物鉴定等研究中具有重要意义。Son 等[7]使用化合物 PRZ-18002 通过对 p-p38 的选择性降解来改善阿尔兹海默病(AD)的生理性质，提出了一种诱导神经退行性疾病相关蛋白质选择性降解的新型治疗靶向蛋白降解模式。Thaisrivongs 等[8]提出了一种有效的选择性 β 位点淀粉样蛋白前体蛋白裂解酶的化学物质抑制剂——维鲁贝司他，作为治疗阿尔兹海默病的潜在疾病改善疗法。Huang 等[9]的研究发现内分泌干扰化学物质(EDCs)对人体免疫功能具有影响，如邻苯二甲酸酯、四氯二苯并二恶英等导致自身免疫性疾病。

自动地从海量文献中抽取化学物质和疾病之间的关系受到了越来越多的关注。BioCreative 社区[2]发布了一项自动化提取篇章级文本中化学物质诱导疾病(CID)关系的挑战，该

任务旨在为生物信息挖掘提供实际效益，自发布以来受到了越来越多的关注。早期的研究通常利用机器学习方法，例如 Lowe 等[10]借助疾病本体库和医学主体词表，利用启发式规则识别句子级 CID 关系。

目前，大多数基于深度学习的 CID 关系提取方法可以分为基于图建模的方法[11-16]和基于 Transformer 的方法[17-22]两类。前者通常依赖于启发式规则将文档建模为以实体为结点的图结构特征表示，并使用图神经网络执行多跳推理，捕获实体之间的语义关系。其中，Nan 等[16]提出了一种细化策略 LSR 以增量地聚合相关信息，以化学或疾病实体在文本中的提及作为结点，通过自动诱导潜在的文档图来增强模型跨句子的 CID 推理能力。Wang 等[13]根据实体的全局和局部表示以及语境关系表征对文档进行编码，以建模实体语义信息并聚合特定实体多次提及的上下文信息，提出了一种联合全局与局部推理的关系抽取方法用于提取 CID 关系。基于 Transformer 的方法通常利用预训练语言模型编码上下文以建模长期依赖，通过隐式消息传递来捕获符号分词、提及和实体之间的交互关系。Zhou 等[20]通过上下文池化将预训练语言模型中的注意力定位到关系事实相关的上下文，同时提出自适应阈值来解决多标签问题。Xu 等[18]显式地建模化学和疾病实体在文本中的提及之间的交互依赖，并将其引入自注意力机制，提出结构化自注意力网络 SSAN 执行上下文推理和结构推理，用于 CID 关系的提取。近年来，将图形建模和 Transformer 建模的方法相结合成为了一个新的研究趋势。Zhang 等[21]利用预训练语言模型编码器捕获实体的上下文信息，同时使用图像风格特征图上的 U 形分割模块来捕捉三元组之间的全局相关性，提出了文档 U 形网络，该方法在 CID 关系抽取任务中获得了优异的性能。Zhang 等[22]提出了一种密集连接交叉注意力网络，分别在实体对矩阵的水平和垂直方向上收集上下文信息，以增强化学实体和疾病实体的表示。

与上述方法不同，部分研究人员注意到，通过编码整个文档提取关系会使得模型缺乏可解释性，并且认为这是导致它与人类识别语义关系存在差距的重要原因之一，因为后者能够以少量的证据句子作为线索。其中，Huang 等[1]认为文档级关系抽取仅需要 3 句话就足以识别所有关系事实，并进一步提出了连续路径、多跳路径和默认路径 3 种启发式方法用于提取证据；Xu 等[23]认为使用整个文档作为线索预测关系是不合理的，并且提出了一种句子重要性评估和聚焦框架 SIEF 以帮助模型更多地关注证据。

目前，生物医学文档级关系抽取是一项值得研究的问题，特别是由于没有明确的监督信号，低成本高效的证据提取仍然是一项具有挑战性的任务。为此，本文提出了一种基于证据聚焦的提及水平文档级关系抽取方法，一方面，面向提及水平建模与推理有助于捕获邻近提及之间的局部交互，而忽略无关提及的影响；另一方面，本文分别设计了两种证据聚焦策略 ATT-EF 和 RL-EF 以帮助模型聚焦少量的证据线索，进一步提升了所提方法的性能。

3 问题描述

本文将生物医学文档级关系抽取任务视为一个给定文本和实体对的多标签分类问题，给定一个生物医学文档 $D_S=\{S_1,S_2,\cdots\}$ 和所包含的实体集合 $\{e_1,e_2,\cdots\}$，其中 S_i 表示文档中的第 i 个句子，e_i 表示第 i 个实体。对于每个句子 $S=\{w_1,w_2,\cdots\}$，它由数量不同的单词组成；对于每个实体 $e=\{m_1,m_2,\cdots\}$，它由数量不同的提及表示，其中 $m=\{w_1,w_2,\cdots\}$ 由 1 个或连续多个单词组成。任务的目标是预测依赖于文档表述的不同实体 e_i 和 e_j 之间的语义关系 $R(e_i,e_j)$，其中 R 是预先定义的关系集合 $\mathfrak{R}$ 中的某类关系，通常来说，e_i 和 e_j 是两种不同类型的实体，如化学实体和疾病实体。

4 基于证据聚焦的提及水平关系抽取方法

本文提出一种基于证据聚焦的提及水平文档级关系抽取方法(EF-MUnet)。首先，使用二维卷积在提及水平特征图上执行推理，捕获邻近提及之间的局部关系；其次，分别基于 ATT-EF 和 RL-EF 策略关注少量且充分的证据线索，避免无关信息的干扰；最后，使用 U-net 网络推理实体水平的全局交互特征，并通过线性网络预测实体之间的语义关系。所提方法框架如图 1 所示。

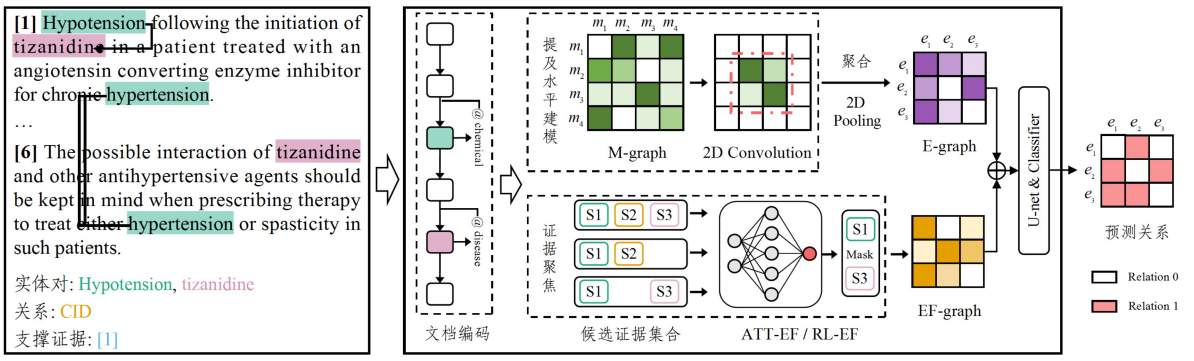


图 1 EF-MUnet 框架图

Fig. 1 EF-MUnet framework diagram

4.1 文档编码

本文使用预训练语言模型将文本转换为词嵌入表示，首先，将原始文档展平为一个由单词组成的文本序列 $D_w=\{w_1,w_2,\cdots\}$；其次，在每个实体提及的前后分别插入特殊标识符“*”和“#”以标注实体在序列中的位置；最后，将其输入语言模型编码器获得上下文相关的嵌入表示。

$$\boldsymbol{H},\boldsymbol{A}=\{\boldsymbol{h}_i\}_{i=1}^{|D|_{\text{token}}}=Encoder(D) \tag{1}$$

其中，H 是输入序列分词级的嵌入表示，$|D|_{\text{token}}$ 是输入序列符号水平分词的数量，$\boldsymbol{A}\in\mathbb{R}^{|D|_{\text{token}}\times|D|_{\text{token}}}$ 是最后一层 Transformer 块中交叉注意力的平均值。此外，为了标识不同单词所属的实体类型，进一步引入实体类型嵌入表示。

$$\boldsymbol{H}=\{[\boldsymbol{h}_i;\boldsymbol{h}_i^{\mathrm{T}}]\}_{i=1}^{|D|_{\text{token}}} \tag{2}$$

其中，h_i^{T} 是第 i 个分词的实体类型嵌入。最后，通过将起始标识符“*”的嵌入作为相应提及的特征，获得提及水平特征表示。

$$\boldsymbol{H}_M=\{\boldsymbol{h}_{m_i}\}_{i=1}^{|D|_m} \tag{3}$$

其中，$\boldsymbol{H}_M\in\mathbb{R}^{|D|_m\times d_h}$，$|D|_m$ 是文档中的提及数量，d_h 是词嵌入的维度与实体类型嵌入的维度之和。

4.2 证据聚焦

文档级关系抽取任务的主要困难在于从整个文档中建模长期依赖以捕获相对位置较远的实体之间的语义关系。然而，已有研究表明并非所有句子都是有效的，且大量关系事实仅与 3 个甚至更少的句子相关[1]，这表明仅需要推理文档的小部分就足以识别出大多数关系。但是，准确地抽取充分且非冗余的证据句子似乎与直接标注关系事实具有相同的任务复杂度，因此无监督证据提取仍然具有挑战性。为此，本文分别基于自注意力机制和强化学习算法设计了两种证据聚焦策略(ATT-EF 和 RL-EF)帮助模型以少量的证据为线索抽取关系。

4.2.1 ATT-EF

从直观上来说，支持证据应该是与实体对之间关联紧密的短语或句子。本文受自注意力机制[24]的启发，通过计算实体对与不同单词之间的相似度来衡量是否将其作为识别该实体对语义关系的线索上下文，提出了一种证据聚焦算法 ATT-EF，如图 2 所示。

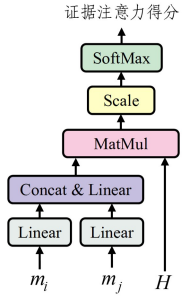


图 2 基于改进自注意力机制的证据聚焦算法(ATT-EF)

Fig. 2 Evidence focusing algorithm based on improved self-attention mechanism(ATT-EF)

具体来说，本文从提及水平建模，因此考虑的是每对提及与其他上下文之间的相似度，即给定一对提及 (m_i,m_j)，首先利用两个独立的前馈神经网络将其相应的特征表示 $\boldsymbol{h}_{m_i}$ 和 $\boldsymbol{h}_{m_j}$ 映射到同一数值空间并降维，随后进行拼接并再次使用一个前馈神经网络自适应地学习对两个提及的融合权重，得到提及对 $\boldsymbol{\chi}(m_i,m_j)$ 的特征表示。最后与上下文的特征表示 $\boldsymbol{H}$ 做 Dot-Product 得到它们之间的相似度得分 $\boldsymbol{\chi}(m_i,m_j)$，如式(4)和式(5)所示：

$$h_m^Q=[FNN_1(h_{m_i});FNN_2(h_{m_j})] \tag{4}$$

$$\boldsymbol{\chi}(m_i,m_j)=FNN_Q(\boldsymbol{h}_m^Q)FNN_K(\boldsymbol{H})^{\mathrm{T}} \tag{5}$$

其中，FNN 是独立的前馈神经网络，FNN_Q 和 FNN_K 类比于自注意力机制中的查询和备查权重参数，$\boldsymbol{H}^{\mathrm{T}}$ 是上下文特征表示经过线性变换后的转置。此外，为避免输入特征维度的大小对网络训练产生影响，缩放后再利用归一化函数将其映射为相似度概率。

$$\boldsymbol{\chi}(m_i,m_j)=\mathrm{Softmax}(\boldsymbol{\chi}(m_i,m_j)/\sqrt{d_k}) \tag{6}$$

其中，d_k 是前馈网络 FNN_Q 或 FNN_K 的输出维度。进一步得到文档中的所有提及对与输入序列的相似度概率并进行维度变换 $\boldsymbol{\chi}\in\mathbb{R}^{|D|_m\times|D|_m\times|D|_{\text{token}}}$，以此作为证据聚焦权重并计算得到提及水平的证据特征表示。

$$\boldsymbol{F}_{EF}=\boldsymbol{\chi}\cdot\boldsymbol{H} \tag{7}$$

其中，$\boldsymbol{F}_{EF}\in\mathbb{R}^{|D|_m\times|D|_m\times d_m}$。ATT-EF 是一种细粒度、单词级别的证据聚焦策略，因此它能够过滤掉一些无关单词而聚焦证据短语，即使它们处在同一句子中。

4.2.2 RL-EF

从给定文档中挑选证据句子本质上可以被视为一个序列决策问题，并且，用于评估所选证据的效用反馈总是在整个决策完成后被给出，如果将它视为延迟奖励，显然强化学习适用于解决这一问题。据此，本文提出了一种证据提取策略 RL-EF，如图 3 所示。具体来说，给定实体对 (e_i,e_j)，代理将根据当前的策略对文档中的句子依次进行动作采样，一旦完成了对整个文档的采样，便将所选句子组合的效用评估作为奖励用于当前策略的学习。

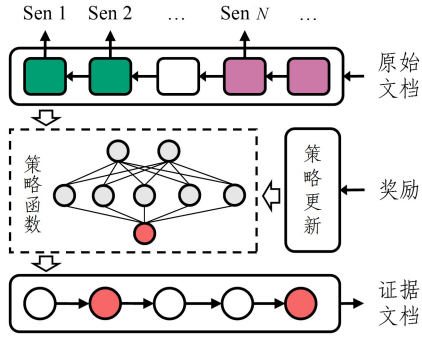


图 3 基于强化学习的证据聚焦算法(RL-EF)

Fig. 3 Evidence focusing algorithm based on reinforcement learning(RL-EF)

1)建模

(1)状态：代理在 t 时刻的观测状态包括当前状态和历史状态两部分，其中，当前状态描述了实体 e_i 和 e_j 以及当前需要的决策句子 S_t 的特征；历史状态则包含了已选择句子的特征。值得注意的是，这里依然使用预训练语言模型获得文本的嵌入表示。不同的是，为了简化训练，这里仅将预训练语言模型作为环境的一部分，不参与策略网络的更新。因此，在 t 时刻的观测状态 s_t 被定义为：

$$s_i=FNN_1[FNN_2(\boldsymbol{h}_{e_i});FNN_3(\boldsymbol{h}_{e_j});FNN_4(\boldsymbol{h}_{S_i});FNN_5(\boldsymbol{h}_{S_{i'}})] \tag{8}$$

其中，$\boldsymbol{h}_e=\mathrm{Max}\{h_{m_i}\}_{m_i\in e}$ 是使用最大池化的方法获得的实体水平特征表示。$\boldsymbol{h}_{S_i}$ 是当前时刻需要决策的句子的特征表示，具体来说，将句子 S_t 拼接首尾标识符“[CLS]”和“[SEP]”后

输入预训练语言模型获得上下文嵌入特征，然后将编码后的起始标识符特征 $\boldsymbol{h}_{[\mathrm{cls}]}$ 作为句子 S_t 的特征表示 $\boldsymbol{h}_{S_t}$。$\boldsymbol{h}_{S_{t'}}$ 是当前时刻已经选择句子的特征表示，为了确保维度不变，它由已选择句子特征的平均值表示，即：$\boldsymbol{h}_{S_{t'}}=1/|D_E|\sum_{S_t\in D_E}h_{S_t}$，其中 D_E 是已选句子的集合，即证据文档。

(2)动作：当前时刻的动作 $a_t\in\{0,1\}$ 被用来指示当前句子是否被选为关于实体对 (e_i,e_j) 的证据句子，形式上当前的动作 a_t 对于 D_E 的改变如下：

$$D_E^t=\begin{cases}D_E^{t-1}\cup S_t & ,a_t=1\\ D_E^{t-1} & ,a_t=0\end{cases} \tag{9}$$

即当 t 时刻动作为 1 时，将句子并入上一时刻的证据文档作为新的证据文档，否则将句子丢弃。

(3)奖励：奖励总是延迟的，其在一个经历(对文档中所有句子的决策)结束后被给出，用来鼓励代理学习选择证据句子的最优策略。研究表明，充分且非冗余的证据能够帮助模型获得最佳的预测效果，过少的句子会造成线索的缺失，过多的句子则会导致信息的冗余。因此，本文利用 MUnet 网络的关系预测结果评估决策的好坏，设计的奖励函数如下：

$$r_T=\frac{1}{|\Re|}\sum_{i=1}^{|\Re|}[y_i\log(P(y_i|D_E))+(1-y_i)\log(1-P(y_i|D_E))] \tag{10}$$

$$r(s_t|D)=\begin{cases}0, & t<|D|_S+1\\ r_T, & t=|D|_S+1\end{cases} \tag{11}$$

其中，$|\Re|$ 是预先定义的关系集合中关系的数量；$|D|_S$ 是文档中句子的数量；y 是真实关系的指示函数；$P(y|D_E)$ 表示根据证据文档，MUnet 网络在关系 y 上的预测概率。非 0 奖励 r_T 被表示为在输入证据文档 D_E 时，MUnet 网络预测出的关系概率分布与真实关系的交叉熵。此外，为避免当 $P(y_i|D_E)$ 等于 0 或 1 时奖励无意义，本文对预测概率 $P(y_i|D_E)$ 做如下限制：

$$P(y_i|D_E)=\mathrm{Min}(\mathrm{Max}(P(y_i|D_E),P_{\min}),P_{\max}) \tag{12}$$

其中，$P_{\min}$ 和 $P_{\max}$ 分别是预先定义的概率最小和最大值，对于超过这一范围的预测概率，在计算奖励时将被截断。

(4)策略：策略指示了代理在状态 s_t 下应该采取什么动作。本文将策略定义为一个关于状态的函数 π_Θ，表达式如下：

$$\pi_\Theta(s_t,a_t)=a_t\cdot\sigma(\boldsymbol{W}s_t+\boldsymbol{b})+(1-a_t)\cdot(1-\sigma(\boldsymbol{W}s_t+\boldsymbol{b})) \tag{13}$$

其中，$\Theta=\{\boldsymbol{W},\boldsymbol{b}\}$ 是可学习参数，$\pi_\Theta(s_t,a_t)$ 表示在状态 s_t 选择动作 a_t 的概率，σ 是激活函数。在训练过程中，遵循当前策略给出的概率对动作采样；在测试过程中，选择概率最大的动作。

2)优化

本文将针对一个实体对的证据句子选择视为一次完整的经历，使用 REINFORCE 算法[25]优化策略，目标是最大化预期累计奖励。形式上来说，目标函数定义如下：

$$\begin{aligned}J(\Theta)&=V_\Theta(s_1|D)\\&=E_{\{s_1,a_1,s_2,\cdots,s_{|D|_S},a_{|D|_S},s_{|D|_S+1}\}\sim\pi_\Theta}\Big[\sum_{t=1}^{|D|_S+1}r(s_t|D)\Big]\end{aligned} \tag{14}$$

其中，$\{s_1,a_1,s_2,\cdots,s_{|D|_S},a_{|D|_S},s_{|D|_S+1}\}$ 是代理根据当前策略 π_Θ 针对文档 D 的一个完整采样序列；$V_\Theta(s_t|D)$ 是初始值，表示遵循策略 π_Θ，从状态 s_1 开始能够获得的预期累计奖励。

根据策略梯度定理[26]，对于任意可微的策略 $\pi_\Theta(s,a)$ 和任意策略目标函数 $J(\Theta)$，策略梯度 $\nabla_\Theta J(\Theta)$ 可以表示为：

$$\nabla_\Theta J(\Theta)=E_{\pi_\Theta}[\nabla_\Theta\log\pi_\Theta(s,a)Q^{\pi_\Theta}(s,a)] \tag{15}$$

其中，$Q^{\pi_\Theta}(s,a)$ 是关于 (s,a) 且遵循策略 π_Θ 的行为价值函数，本文使用 REINFORCE 算法计算策略梯度，如式(6)所示：

$$\nabla_\Theta J(\Theta)=\nabla_\Theta\log\pi_\Theta(s,a)v \tag{16}$$

其中，v 是根据蒙特卡洛算法计算得到的状态行为对 (s,a) 的累计奖励，它被视为是对 $Q^{\pi_\Theta}(s,a)$ 的无偏估计。因此，对于每个状态 s_t，参数 Θ 的更新可以表示为：

$$\Theta\leftarrow\Theta+\alpha\nabla_\Theta\log\pi_\Theta(s_t,a_t)v_t \tag{17}$$

其中，α 是学习率。由于本文只有在终止状态 $s_{|D|+1}$ 时存在一个非 0 的奖励，因此，在一次采样中，所有状态下的 v_t 是相同的，即：对于任意时刻 t，都有 $v_t=r(s_{|D|_S+1}|D)$。

3)证据聚焦掩码

基于 RL-EF 策略构造证据聚焦掩码矩阵 $\boldsymbol{M}_{EF}\in\mathbb{R}^{|D|_e\times|D|_e\times|D|_{\mathrm{token}}}$，它指示了关于实体对的有效证据，其中 $|D|_e$ 是 D 中实体的数量，$|D|_{\mathrm{token}}$ 是 D 的符号水平分词数量。具体来说，给定实体对 (e_i,e_j)，若句子 S_k 被选为证据句子，则 $M_{EF}^{(e_i,e_j)}$ 中对应句子 S_k 位置的元素被置为 1，否则为 0。此外，由于本文基于提及水平建模，而一个实体通常由多个提及组成，因此通过将同一实体的不同提及赋予相同的证据掩码以获得提及水平的证据掩码矩阵 $\boldsymbol{M}_{EF}\in\mathbb{R}^{|D|_m\times|D|_m\times|D|_{\mathrm{token}}}$。最后，将其作为证据聚焦权重来计算提及水平的证据特征。

$$F_{EF}=\boldsymbol{M}_{EF}\cdot H \tag{18}$$

其中，$F_{EF}\in\mathbb{R}^{|D|_m\times|D|_m\times d_h}$。与 ATT-EF 不同的是，RL-EF 是句子水平的证据聚焦策略，它选择的支持证据是整个句子而非某个单词或短语，因此能够保留语义的完整性。

4)RL-EF 网络训练

RL-EF 作为一个独立的网络结构，需要与 MUnet 网络联合训练。为了帮助模型稳定收敛，本文将整个过程划分为 3 个步骤：(1)预训练 MUnet 网络；(2)固定 MUnet 网络，预训练 RL-EF 网络；(3)联合训练 MUnet 网络和 RL-EF 网络至收敛。在步骤(1)中，采用启发式规则[1]，使用连续路径和多跳路径作为对证据文档的估计以预训练 MUnet 网络。其中，连续路径表示两个实体在同一个句子或者连续的两个句子中被提及，将这些句子视为该实体对的证据文档；多跳路径表示两个实体并不直接相连，而是通过第 3 个实体桥接，将这 3 个实体所在的句子视为该实体对的证据文档。在步骤(2)中，固定 MUnet 网络的参数以确保它能够提供稳定的奖赏预训练 RL-EF 网络。

4.3 提及特征推理

在文档级关系抽取任务中，同一实体的不同提及广泛分布在整个文档中并展现出复杂的交互模式。如表 3 所列("@Doc"和"@Evi"分别表示涉及实体在原始文档和证据文档中被提及超过 1 次的关系实例占所有关系实例的比例。对于 CDR 和 GDA 数据集，通过随机抽取 100 条数据进行人工标注的方法获得结果，以"*"标识)，关系实例所涉及的实体在整个文档中被提及次数超过 1 的比例远大于证据文档，这表明

通常仅需要通过少数提及之间的局部语义就能够识别出它们分别所指代的实体之间的关系。

表 3 关系实例涉及实体在原始文档和证据文档中被提及的次数大于 1 的占比

Table 3 Proportion of entities involved in relationship instances mentioned more than 1 times in the original and evidence documents

数据集	@Doc	@Evi
CDR	0.96	0.32*
GDA	0.97	0.46*
DocRED	0.59	0.21

为了解决这一问题，本文提出从提及水平建模交互特征，并使用二维卷积捕获相邻提及之间的局部交互，从而避免无关提及的干扰，如图 4 所示。

(a)提及水平图建模 (b)实体水平图建模

图 4 提及水平和实体水平建模比较

Fig. 4 Comparison between mention-level and entity-level modeling

具体来说，首先使用基于上下文感知的方法获得提及水平的特征图，形式上，给定提及对，它的特征计算如式(19)所示：

$$\boldsymbol{f}(m_i,m_j)=\sum_{k=1}^{|D|_{\text{token}}}\boldsymbol{A}_{m_i,k}\cdot\boldsymbol{A}_{m_j,k}\cdot\boldsymbol{h}_k \tag{19}$$

其中，$\boldsymbol{A}$ 是等式(1)中的交叉注意力矩阵；$A_{m_i,k}$ 和 $A_{m_j,k}$ 分别表示提及 m_i 和 m_j 对于第 k 个分词的注意力得分，即提及 m_i 和 m_j 涉及的分词与第 k 个分词的注意力均值。

此外，值得注意的是图 4(a)中的提及并没有按照它们所指示的实体顺序排列 m_1^1,m_2^1,m_1^2,m_1^3，而是基于它们在文档中所处的位置组合 m_1^1,m_1^2,m_2^1,m_1^3。直观上，相邻提及之间表现出更加紧密的联系，通过二维卷积能够有效捕获邻近提及之间的局部交互，如式(20)所示：

$$\boldsymbol{F}_M=Conv2d(\boldsymbol{F}_M) \tag{20}$$

其中，$\boldsymbol{F}_M\in\mathbb{R}^{|D|_m\times|D|_m\times d_c}$ 可以被看作是从上下文感知特征图中提取的包含提及之间局部交互信息的特征子图，d_c 是二维卷积的输出通道数量。

为进一步聚焦有效的证据线索，本文融合了提及水平的证据特征表示，如式(21)所示：

$$\boldsymbol{F}_M=FNN([\boldsymbol{F}_M;\boldsymbol{F}_{EF}]) \tag{21}$$

其中，FNN 用于自适应地学习上下文感知特征与证据聚焦特征之间的平衡权重，并降低特征的维度。

4.4 关系推理

由于任务的目标是捕获实体之间的语义关系，因此进一步使用二维最大池化将提及水平的特征图聚合到实体水平，即给定实体对 (e_i,e_j)，它的特征表示如式(22)所示：

$$\boldsymbol{f}_E(e_i,e_j)=\text{Max}\{\boldsymbol{f}(m_p,m_q)\}_{m_p\in e_i,m_q\in e_j} \tag{22}$$

由此得到实体水平的特征图 $F_e\in\mathbb{R}^{|D|_e\times|D|_e\times d_c}$。

本文将计算机视觉中的语义分割模型 U-net 网络用于实体水平特征图中的隐式推理[19]，一方面，U 形分割结构能够促进实体对之间的信息交换；另一方面，下采样卷积模块能够扩大信息感受野，从而学习到更加丰富的全局特征。形式上，本文使用 U-net 分割网络捕获深层次的实体水平特征。

$$\boldsymbol{F}_E=U\text{-}net(FNN(\boldsymbol{F}_E)) \tag{23}$$

其中，FNN 用于将特征映射到一个更小的维度作为 U-net 网络输入特征的通道。此时，$F_E\in\mathbb{R}^{|D|_e\times|D|_e\times d_U}$，$d_U$ 是 U-net 网络的输出维度。最后，使用多层感知机来获得实体之间的关系预测分数。

$$P=MLP(\boldsymbol{F}_E) \tag{24}$$

其中，$P\in\mathbb{R}^{|D|_e\times|D|_e\times|\Re|}$ 是实体对关于预先定义的关系集合 $\Re$ 的预测分数。

4.5 损失函数

在文档级关系抽取任务中只有少量实体对具有语义关系，即存在正负样本比例极度不均衡的情况，因此本文使用不对称损失(Asymmetric Loss，ASL)[27]训练网络，如式(25)—式(28)所示：

$$P_n(R|(e_i,e_j))=\text{Max}(P(R|(e_i,e_j))-n,0) \tag{25}$$

$$L_+=(1-P(R|(e_i,e_j)))^{\gamma+}\log(P(R|(e_i,e_j))) \tag{26}$$

$$L_-=(P_n(R|(e_i,e_j)))^{\gamma-}\log(1-P_n(R|(e_i,e_j))) \tag{27}$$

$$L=-\sum_{i,j\in|D|_e}\sum_{R\in\Re}Y_{(R=1)}L_++Y_{(R=0)}L_- \tag{28}$$

其中，$P(R|(e_i,e_j))$是模型关于实体对(e_i,e_j)在关系 R 上的预测概率；$\gamma+$和 $\gamma-$分别是调节正负样本比例的超参数；n 是可调节超参数，用于进一步过滤负样本；Y 是真实标签的指示函数。

5 实验

5.1 数据集

为了证明所提方法的有效性，本文在生物医学文档级关系抽取数据集 CDR[4]上开展实验，该数据集是 BioCreative 针对化学物质、疾病和化学诱导疾病关系(CID)手工注释的，包括来自于 PubMed 数据库的 1 500 篇生物医学文章的标题和摘要，并被平均划分为训练集、验证集和测试集。为了与之前的研究方法进行公平有效的比较，本文遵循该数据集的划分方式。此外，已有研究发现在 CDR 数据集中的疾病或化学物质存在假名关系[28]，如表 4 所列。

表 4 CDR 数据集示例

Table 4 Example of CDR dataset

标题	摘要	实体	关系
Carbamazepine-induced cardiac dysfunction	A patient with sinus bradycardia and atrioventricular block, induced by carbamazepine, prompted an extensive literature review of all previously reported cases	Carbamazepine/carbamazepine(C2); cardiac dysfunction(D4); bradycardia(D5); atrioventricular block(D6)	C2-D5;C2-D6

从表 4 中可以看出，CDR 语料库中仅将 C2-D5 和 C2-D6 标注为关系事实，而将 C2-D4 处理为负样本，这是因为 CID 任务的目标是提取最具体的疾病与化学物质之间的关系，“cardiac dysfunction”只是一个一般性的概念实体。然而，

标题文本“Carbamazepine-induced cardiac dysfunction”是一种典型的关系事实表述模式，因此 C2-D4 被标注为负样本必然会影响模型的训练。为解决这一问题，本文使用 MeSH 词表删除语料中涉及实体假名的负样本实例，即假名过滤（HF）。为了与其他方法进行公平的比较，本文分别测试了有无假名过滤情况下模型在 CDR 数据集上的性能。

此外，本文还在 Peng 等[29]提出的药物突变数据集（Drug-Mutation Interaction，DMI）上测试了 EF-MUnet 的泛化能力。该数据集共包含 6087 个与药物、突变相关的二元关系事实，涉及 5 类具体关系：“resistance or non-response”“sensitivity”“response”“resistance”“None”。同时，遵循 Peng 等的研究，通过将前 4 种关系视为“有关系”，将“None”视为“无关系”，将其转换为二分类问题。在后续表述中，本文分别使用“Five-class”和“Two-class”来表示上述两种设定。

5.2 实验设置

本文所提方法基于 Pytorch 实现，训练在一块 NVIDIA RTX 3090 24GB GPU 上进行，并遵循大多数研究使用 Intra-F1，Inter-F1 以及 F1 来评估模型的性能。在实验过程中，将 SciBERTbase[30]作为文本编码器，使用 AdamW 优化关系分类模型，并对前 6%的训练步骤进行线性预热，其他详细的参数设置如表 5 所列。

表 5 超参数设置

Table 5 Hyperparameter settings

超参数	值
Base	
最大实体数量	22
最大输入长度	512
词嵌入维度	768
实体类型嵌入维度	20
卷积输出通道数	64
U-net 网络输入/输出通道数	3/256
基础学习率/SciBERT 学习率	$4\times10^{-4}/3\times10^{-5}$
Dropout	0.4
RL-EF	
输入线性层输出维度	64
动作空间数量	2
学习率	1×10^{-3}

5.3 实验结果

5.3.1 模型比较

为了证明所提方法的有效性，本文在生物医学文档级关系抽取数据集 CDR 上开展了实验，并与之前的大多数方法进行了比较，结果如表 6 所列。其中 EF-MUnet$_{\text{ATT-EF}}$ 和 EF-MUnet$_{\text{RL-EF}}$ 分别表示基于 ATT-EF 和 RL-EF 策略的 EF-MUnet 方法，“+HF”表示与假名过滤方法结合。

实验结果表明，所提方法 EF-MUnet$_{\text{RL-EF}}$+HF 获得了最优的性能，F1 分数达到了 86.8%，与之前的最优方法相比，提升了 9.7%。此外，还可以注意到 EF-MUnet$_{\text{ATT-EF}}$ 的性能表现略低于 EF-MUnet$_{\text{RL-EF}}$，这可能是因为在模型学习过程中，相较于隐式地学习证据短语，显式地提供证据句子更加简单并且有效。而且相比已有句间关系推理能力表现最好的方法 MRN，EF-MUnet 在 4 种模式下（EF-MUnet$_{\text{ATT-EF}}$，EF-MUnet$_{\text{RL-EF}}$，EF-MUnet$_{\text{ATT-EF}}$+HF 和 EF-MUnet$_{\text{RL-EF}}$+HF）的句间关系推理性能都有显著提升，Inter-F1 分数分别提高了 3.3%，5.5%，19.7%和 21.3%，这对于文档级关系抽取任务来说至关重要。值得注意的是，假名过滤无需引入额外的学习参数，仅仅使用 MeSH 词表删除语料中涉及实体假名的负样本实例使得标注更加合理，以此提升模型的性能。

表 6 不同方法在 CDR 数据集上的性能比较

Table 6 Performance comparison of different methods on CDR dataset

（%）

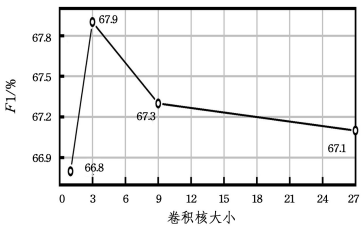
方法	*Intra*-F1	*Inter*-F1	F1
BRAN[31]	—	—	62.1
Zheng 等[32]	67.4	43.9	61.5
EoG[33]	68.2	50.9	63.6
SciBERT[30]	—	—	65.1
LSR[16]	68.9	53.1	64.8
DHG[12]	68.6	54.1	65.9
Tran 等[34]	—	—	66.1
BERT-GT[35]	—	—	65.9
Xu 等[18]	—	—	68.7
ATLOP[20]	—	—	69.4
MRN[36]	70.4	54.2	65.9
DocuNet[21]	—	—	76.3
DocR-BERT+CSA_All[37]	69.4	52.7	62.8
Seq2rel[38]	—	—	67.2
Dense-CCNet[22]	—	—	77.1
Duan 等[39]	—	—	64.2
RDDCP[40]	—	—	71.6
MHGNN[41]	—	—	73.0
EF-MUnet$_{\text{ATT-EF}}$	73.0	57.5	67.9
EF-MUnet$_{\text{RL-EF}}$	73.6	59.7	69.3
EF-MUnet$_{\text{ATT-EF}}$+HF	88.6	73.9	85.8
EF-MUnet$_{\text{RL-EF}}$+HF	**89.3**	**75.5**	**86.8**

综上，本文所提方法通过在提及水平建模，并利用二维卷积网络捕获提及之间的局部关系，再基于 ATT-EF 或 RL-EF 策略关注证据线索，能够有效地提升文档级关系抽取的性能。

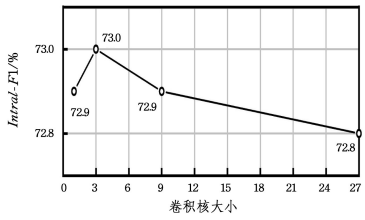
5.3.2 二维卷积窗口大小的影响

为了进一步验证在提及水平图推理中二维卷积网络对捕获提及之间局部交互能力的影响，本文分别比较了在卷积核大小为 1，3，9 和 27 时 EF-MUnetATT-EF 在测试集上的性能，结果如图 5 所示。

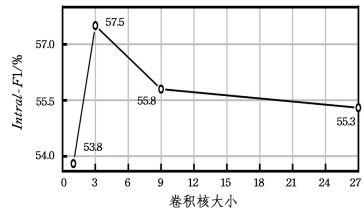
如上文所述，二维卷积用于从提及水平特征图中融合特征并提取特征子图。理论上，卷积窗口的大小直接影响信息聚合时的感受野。当卷积核大小为 1 时，局部特征子图中的每个像素点都是原始特征图中像素点的一一映射，这种情况下模型无法捕获到相邻提及之间的局部交互，不具备捕获远程实体关系的能力。因此从图 5 可以看出，此时的模型整体性能最差，特别是表征句间关系推理能力的 Inter-F1 分数仅为 53.8%。随着卷积核的增加，二维卷积在提取特征子图时能够融合邻近提及的特征，进而有效地捕获提及之间的局部信息。从图 5 可以看出，当卷积核大小增加到 3 时，模型的性能获得了明显的提升，F1 分数提升至 67.9%，Inter-F1 分数提升了 3.7%。然而，随着卷积核大小的进一步增加，模型的性能反而下降，特别是当卷积核大小为 27 时，模型的句内关系推理能力表现最差，Intra-F1 分数仅为 72.8%，这是因为此时特征子图中的每个像素点都融合了过多距离较远的无关信息，反而忽视了包含更多有价值线索的局部信息。



(a)卷积核大小对 EF-MUnet 的整体性能影响



(b)卷积核大小对 EF-MUnet 的句内关系抽取性能影响



(c)卷积核大小对 EF-MUnet 的句间关系抽取性能影响

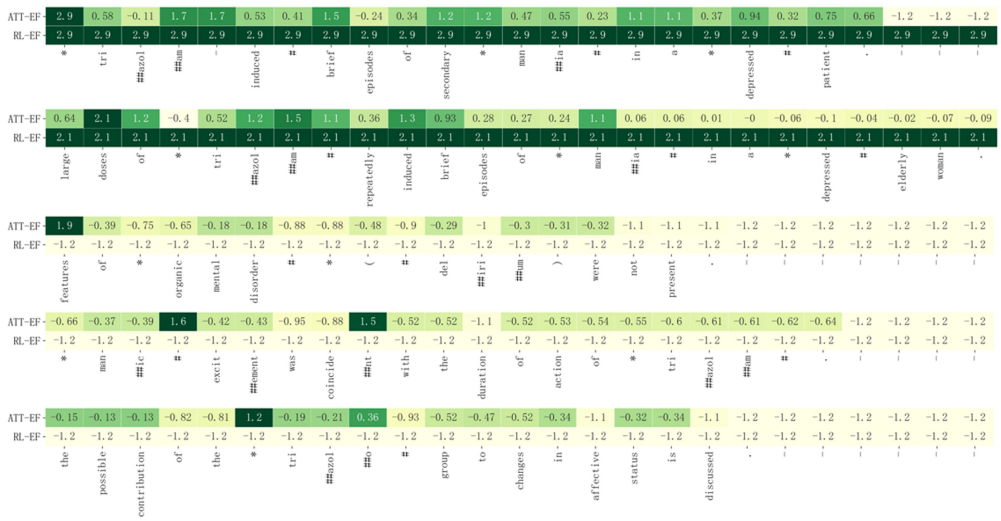
图 5 卷积核大小对于 EF-MUnet 性能的影响

Fig. 5 Effect of convolutional kernel size on EF-MUnet performance

总而言之，使用二维卷积提取特征子图，捕获邻近提及之间的局部交互对于文档级关系抽取任务来说是有帮助的，而卷积核大小影响了邻近提及之间特征的融合。实验结果表明，在本任务中，卷积核大小为 3 时，所提方法取得了最好的性能。

5.3.3 证据关注可视化

本文借助 CDR 数据集中的一个文档示例进一步分析所提 ATT-EF 和 RL-EF 策略关于证据线索聚焦的表现，并且比较了它们之间的差异。如图 6 所示，该文档包含 5 个句子并标注出关系事实“triazolam-CID-mania”，而这一关系事实仅被句子[S1]和[S2]所表述。为了验证 ATT-EF 和 RL-EF 的有效性，分别在图 6 中可视化了两种策略下的证据掩码概率分布，图中颜色的深浅表示关注度的高低。首先，从图中可以看出，两种策略对句子[S1]和[S2]都具有更高的关注度。不同的是，RL-EF 是一种面向句子级的证据聚焦策略，它能够准确地识别出证据句子并且对非证据句子的关系度识别为 0。ATT-EF 是一种细粒度的证据聚焦策略，用来考虑提及对与上下文中每个单词的关注度，从图中可以看出，它在非证据句子中仍然存在着一些关注区域，引入了无关信息，但是它对证据句子具有过滤能力，能够帮助模型过滤一些可能与证据抽取无关的信息，帮助模型只聚焦有价值的上下文，有助于模型理解冗余句式。



[S1] @Chemical **Triazolam**-induced brief episodes of secondary @Disease **mania** in a depressed patient.

[S2] Large doses of @Chemical **triazolam** repeatedly induced brief episodes of @Disease **mania** in a depressed elderly woman.

[S3] Features of organic mental disorder (delirium) were not present.

[S4] Manic excitement was coincident with the duration of action of @Chemical **triazolam**.

[S5] The possible contribution of the @Chemical **triazolo** group to changes in affective status is discussed.

Relation: *triazolam-CID-mania*

Evidence: [1], [2]

图 6 不同策略下证据聚焦分析

Fig. 6 Evidence focusing analysis under different strategies

5.3.4 消融实验

为了进一步分析所提模块的有效性，本文训练和验证了所有模块的可能组合，在测试集上进行了消融实验，结果如表 7 所列。其中，“-提及水平图推理”表示消融二维卷积推理模块；“-上下文感知”表示消融基于交叉注意力矩阵计算的上下文感知模块；“-证据聚焦”表示消融基于 ATT-EF 或 RL-EF 策略聚焦模块；“-U-net”表示消融 U-net 网络。

表 7 消融不同模块后 EF-MUnet(ATT-EF/RL-EF)的性能比较

Table 7 Performance comparison of EF-MUnet(ATT-EF/RL-EF) after ablation of different modules

(%)

方法	*Intra-F*1	*Inter-F*1	*F*1
EF-MUnet	72.7/73.5	57.5/59.7	67.9/69.3
-提及水平图推理	72.2/72.8	56.4/58.8	67.3/68.6
-上下文感知	70.1/72.7	54.8/56.6	65.9/67.7
-证据聚焦	72.9/73.2	55.1/56.8	67.2/68.2
-U-net	72.3/72.9	55.8/57.7	66.9/68.3

从表 7 中可以看出，消融任何一个模块后，EF-MUnet 的整体性能都有所下降，这表明本文所提的每个模块都是有效并且互补的。其中，上下文感知策略能够捕捉到实体对与上下文之间的语义关联，获取丰富的上下文信息，消融该模块对于 EF-MUnet 性能的影响最大，EF-MUnet$_{\text{ATT-EF}}$ 和 EF-MUnet$_{\text{RL-EF}}$ 的 F1 分数分别降低了 2%和 1.6%；证据聚焦模块对于句间关系抽取的性能也具有较大影响，消融后，Inter-F1 分别减少了 2.4%和 2.9%，这是因为有效的证据聚焦能够减小远距离实体之间的跨度，降低句间关系抽取的复杂度；此外，消融 U-net 模块对模型的性能也有较大的影响，这表明捕获实体之间的全局交互对于文档级关系抽取任务来说是有效的。

5.3.5 泛化实验

本文主要针对提取化学物质诱导疾病关系设计并开展实验，为了测试所提方法 EF-MUnet 的泛化能力，将其进一步运用于药物与突变相互作用的关系抽取任务中。为了与之前的大多数方法进行公平且合理的比较，本文采用准确率作为评估指标在 DMI 数据集上开展实验并与现有方法进行对比，同时额外评估了所提方法的 F1 分数，结果如表 8 所列。

表 8 不同方法在 DMI 数据集上的性能比较

Table 8 Performance comparison of different methods on DMI dataset

(%)

方法	Two-class		Five-class	
	Accuracy	*F*1	*Accuracy*	*F*1
AGGCN[42]	85.4	—	76.9	—
BERT[43]	91.0	—	90.0	—
BlueBERT[44]	92.1	—	92.2	—
LF-GCN[45]	87.1	—	—	—
GA LSTM[46]	85.8	—	77.5	—
NGG-LSTM[47]	86.5	—	77.8	—
MA-GCN[47]	91.3	—	—	—
BERT-GT[48]	93.3	—	93.5	—
Chen et al.[49]	90.3	—	87.6	—
EF-MUnet$_{\text{ATT-EF}}$	97.6	95.7	98.2	96.4
EF-MUnet$_{\text{RL-EF}}$	**98.5**	**96.8**	**98.6**	**97.4**

从表 8 中可以看出，两种策略下的 EF-MUnet 都获得了较高的准确率，与之前的最优方法相比，EF-MUnet$_{\text{RL-EF}}$ 在二分类和多分类两种设定下准确率分别提升了 5.2%和 5.1%。这表明本文所提出的提及水平建模推理以及证据聚焦策略是有效的，即使在没有引入远程词典的情况下，在药物与突变关系抽取任务中与已有方法相比依然具有优越的性能，证实了 EF-MUnet 在生物医学领域关系抽取中具有一定的泛化能力，是一个稳定且有效的关系抽取方法。

结束语 本文提出了一种基于证据聚焦的提及水平生物医学文档级关系抽取方法，该方法通过建模提及水平的特征表示，并利用二维卷积捕获提及之间的局部交互模式；同时受益于 ATT-EF 和 RL-EF 策略以聚焦于少量且充分的证据线索；最后借助 U-net 网络捕获实体水平的全局特征以挖掘语义关系。与已有的方法相比，EF-MUnet 在生物医学文档级关系抽取数据集 CDR 上表现出优越的性能，特别是对于句间关系推理的能力获得了显著的提升，表明它是一个有效的文档级关系抽取方法。此外，该方法在提取药物与突变之间相互作用的数据集 DMI 上也取得了最优的性能，说明它在生物医学领域的相关任务中具有泛化性。

然而，随着模型深度和复杂度的增加，单一的关系标签无法为模型的训练提供高质量的监督，而一些与关系抽取任务相关的中间步骤（如命名实体识别、证据抽取）能够为模型提供更加明确的指导。因此，在未来的工作中，将结合多任务学习进一步提升 EF-MUnet 在文档级关系抽取任务中的性能并增强其可解释性。

参 考 文 献

[1] HUANG Q, ZHU S, FENG Y, et al. Three sentences are all you need: Local path enhanced document relation extraction[C]// 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. 2021: 998-1004.

[2] MA Y, WANG A, OKAZAKI N. DREEAM: Guiding Attention with Evidence for Improving Document-Level Relation Extraction[J]. arXiv: 2302.08675, 2023.

[3] 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.

[4] LI J, SUN Y, JOHNSON R J, et al. BioCreative V CDR task corpus: a resource for chemical disease relation extraction [J]. Database, 2016: baw068.

[5] YE W, LUO R B, HENRY C M L, et al. Renet: A deep learning approach for extracting gene-disease associations from literature [C]// International Conference on Research in Computational Molecular Biology. Springer, 2019: 272-284.

[6] YAO Y, YE D M, 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.

[7] SON S H, LEE N R, GEE M S, et al. Chemical Knockdown of Phosphorylated p38 Mitogen-Activated Protein Kinase (MAPK) as a Novel Approach for the Treatment of Alzheimer's Disease [J]. ACS Central Science, 2023, 9(3): 417-426.

[8] THAISRIVONGS D A, MORRIS, W J, SCOTT J D. Discovery and Chemical Development of Verubecestat, a BACE1 Inhibitor for the Treatment of Alzheimer's Disease [J]. ACS Symposium Series, 2018, 1037: 53-89.

[9] HUANG R G, LI X B, WANG Y Y, et al. Endocrine-disrupting chemicals and autoimmune diseases [J]. Environmental Research, 2023, 231: 116222.

[10] LOWE D M, O'BOYLE N M, SAYLE R A. Efficient chemical-disease identification and relationship extraction using Wikipedia to improve recall [J]. Database, 2016, 2016: baw039.

[11] LI B, YE W, SHENG Z, et al. Graph enhanced dual attention network for document-level relation extraction[C]// 28th International Conference on Computational Linguistics. 2020: 1551-1560.

[12] ZHANG Z, YU B, SHU X, et al. Document-level relation extraction with dual-tier heterogeneous graph[C]// 28th International Conference on Computational Linguistics. 2020: 1630-1641.

[13] WANG D, HU W, CAO E, et al. Global-to-local neural networks for document-level relation extraction[C]// 2020 Conference on Empirical Methods in Natural Language Processing. 2020: 3711-3721.

[14] ZENG S, XU R, CHANG B, et al. Double graph based reasoning for document-level relation extraction[C]// 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference. 2020: 1630-1640.

[15] WANG X, KEHAI C, TIEJUN Z. Document-level relation extraction with reconstruction[C]// 35th AAAI Conference on Artificial Intelligence. 2021: 14167-14175.

[16] NAN G, GUO Z, SEKULIĆ I, et al. Reasoning with latent structure refinement for document-level relation extraction[C]// 58th Annual Meeting of the Proceedings of the Conference. 2020: 1546-1557.

[17] YE D, LIN Y, DU J, et al. Coreferential reasoning learning for language representation [C]// 2020 Conference on Empirical Methods in Natural Language Processing. 2020: 7170-7186.

[18] XU B, WANG Q, LYU Y, et al. Entity Structure Within and Throughout: Modeling Mention Dependencies for Document-Level Relation Extraction[C]// 35th AAAI Conference on Artificial Intelligence. 2021: 14149-14157.

[19] HONG W, CHRISTFRIED F, ROB S, et al. Fine-tune Bert for DocRED with Two-step Process[J]. arXiv: 1909. 11898, 2019.

[20] ZHOU W, HUANG K, MA T, et al. Document-Level Relation Extraction with Adaptive Thresholding and Localized Context Pooling[C]// 35th AAAI Conference on Artificial Intelligence. 2021: 14612-14620.

[21] ZHANG N Y, CHEN X XIE X, et al. Document-level relation extraction as semantic segmentation[C]// Proceedings of the 30th International Joint Conference on Artificial Intelligence. 2021: 3999-4006.

[22] ZHANG L, CHENG Y. A Densely Connected Criss-Cross Attention Network for Document-level Relation Extraction[J]. arXiv: 2203. 13953, 2022.

[23] XU W, CHEN K, MOU L, et al. Document-Level Relation Extraction with Sentences Importance Estimation and Focusing [C]// 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2022: 2920-2929.

[24] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Advances in Neural Information Processing Systems 30—Proceedings of the 2017 Conference. Long Beach, CA, United states, 2017: 5999-6009.

[25] SUTTON R S, BARTO A G. Reinforcement learning: An introduction [M]. MIT press, 2018.

[26] SUTTON R S, MCALLESTER D, SINGH S, et al. Policy gradient methods for reinforcement learning with function approximation[C]// Advances in Neural Information Processing Systems. 1999.

[27] RIDNIK T, BEN-BARUCH E, ZAMIR, et al. Asymmetric loss for multi-label classification[C]// 18th IEEE/CVF International Conference on Computer Vision. 2021: 82-91.

[28] GU J, QIAN L, ZHOU G. Chemical-induced disease relation extraction with various linguistic features [J]. Database, 2016, 2016: baw042.

[29] PENG N, 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.

[30] BELTAGY I, LO K, COHAN A. SciBERT: A pretrained language model for scientific text[C]// 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing. 2019: 3615-3620.

[31] VERGA P, STRUBELL E, MCCALLUM A. Simultaneously self-attending to all mentions for full-abstract biological relation extraction[C]// 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2018: 872-884.

[32] ZHENG W, LIN H F, LI Z H, et al. An effective neural model extracting document level chemical-induced disease relations from biomedical literature [J]. Journal of Biomedical Informatics, 2018, 83(2018): 1-9.

[33] CHRISTOPOULOU F, MIWA M, ANANIADOU S. Connecting the dots: Document-level neural relation extraction with edge-oriented graphs[C]// 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing. 2019: 4925-4936.

[34] TRAN H M, NGUYEN M T, NGUYEN T H. The dots have their values: exploiting the node-edge connections in graph-based neural models for document-level relation extraction[C]// ACL 2020: EMNLP 2020: Findings of the Association for Computational Linguistics. 2020: 4561-4567.

[35] LAI P T, LU Z. BERT-GT: cross-sentence n-ary relation extraction with BERT and Graph Transformer [J]. Bioinformatics, 2020, 36(24): 5678-5685.

[36] LI J, XU K, LI F, et al. MRN: A locally and globally mention-based reasoning network for document-level relation extraction [C]// ACL-IJCNLP 2021: Findings of the Association for Computational Linguistics. 2021: 1359-1370.

[37] LI Z G, LIN H F, SHEN C, et al. Document-level Chemical-induced Disease Relation Extraction via Cross Self-attention [J]. Journal of Chinese Information Processing, 2022, 36(7): 98-105.

[38] GIORGI J，BADER G D，WANG B. A sequence-to-sequence approach for document-level relation extraction[C]// BioNLP 2022 @ ACL 2022：Proceedings of the 21st Workshop on Biomedical Language Processing. 2022：10-25.

[39] DUAN J Y，YANG X，WANG H，et al. Document level relationship extraction based on inter sentence information in graph attention convolutional networks [J]. Computer Science，2023，50(S1)：191-196.

[40] DONG Y，XU X. Relational distance and document-level contrastive pre-training based relation extraction model [J]. Pattern Recognition Letters，2023，167：132-140.

[41] WANG N，CHEN T，REN C，et al. Document-level relation extraction with multi-layer heterogeneous graph attention network [J]. Engineering Applications of Artificial Intelligence，2023，123：106212.

[42] GUO Z，ZHANG Y，LU W. Attention guided graph convolutional networks for relation extraction[C]// 57th Annual Meeting of the Association for Computational Linguistics. 2019：241-251.

[43] DEVLIN J，CHANG M W，LEE K，et al. Bert：pre-training of deep bidirectional transformers for language understanding [C]// 2019 Conference of the North American Chapter of the Association for Computational Linguistics：Human Language Technologies. 2019：4171-4186.

[44] PENG Y，YAN S，LU Z. Transfer learning in biomedical natural language processing：an evaluation of BERT and ELMo on ten benchmarking datasets [C] // 18th SIGBioMed Workshop on Biomedical Natural Language Processing. 2019：58-65.

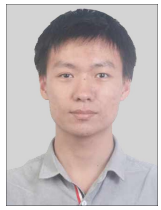
[45] JIANG Y，ZHOU Y，TU K. Learning and evaluation of latent dependency forest models [J]. Neural Computing and Applications，2019，31：6795-6805.

[46] ZHAO L，XU W，GAO S，et al. Cross-sentence N-ary relation classification using LSTMs on graph and sequence structures [J]. Knowledge-Based Systems，2020，207：106266.

[47] ZHAO D，WANG J，LIN H，et al. Biomedical cross-sentence relation extraction via multihead attention and graph convolutional networks [J]. Applied Soft Computing，2021，104：107230.

[48] LAI P T，LU Z. BERT-GT：cross-sentence n-ary relation extraction with BERT and Graph Transformer [J]. Bioinformatics，2020，36(24)：5678-5685.

[49] CHEN X，ZHANG M，XIONG S，et al. On the form of parsed sentences for relation extraction [J]. Knowledge-Based Systems，2022，251：109184.



ZHOU Xueyang，born in 1998，postgraduate. His main research interests include natural language processing and biomedical information mining.

FU Qiming，born in 1985，Ph.D，professor，is a member of CCF(No. 23956M). His main research interests include reinforcement learning，deep learning and intelligent information processing.

（责任编辑：何杨）