

融合实体和上下文信息的篇章关系抽取研究

黄河燕^{1,2} 袁长森^{1,2} 冯冲^{1,2}

摘要 篇章关系抽取旨在识别篇章中实体对之间的关系。相较于传统的句子级别关系抽取，篇章级别关系抽取任务更加贴近实际应用，但是它对实体对的跨句子推理和上下文信息感知等问题提出了新的挑战。本文提出融合实体和上下文信息 (Fuse entity and context information, FECCI) 的篇章关系抽取方法，它包含两个模块，分别是实体信息抽取模块和上下文信息抽取模块。实体信息抽取模块从两个实体中自动地抽取能够表示实体对关系的特征。上下文信息抽取模块根据实体对的提及位置信息，从篇章中抽取不同的上下文关系特征。本文在三个篇章级别的关系抽取数据集上进行实验，效果得到显著提升。

关键词 篇章关系抽取，实体信息，上下文信息，提及位置信息，跨句子推理

引用格式 黄河燕, 袁长森, 冯冲. 融合实体和上下文信息的篇章关系抽取研究. 自动化学报, 2024, 50(10): 1953–1962

DOI 10.16383/j.aas.c220966 **CSTR** 32138.14.j.aas.c220966

Document-level Relation Extraction With Entity and Context Information

HUANG He-Yan^{1,2} YUAN Chang-Sen^{1,2} FENG Chong^{1,2}

Abstract Document-level relation extraction aims to identify the relations among entities from the document. Compared with traditional sentence-level relation extraction, document-level relation extraction is more realistic and poses new challenges of cross-sentence inference and context information understanding. In this paper, we propose a novel method for document-level relation extraction by fusing entity and context information (FECCI), which contains two modules: Entity information extraction module and context information extraction module. Entity information extraction module automatically extracts crucial relation features about entity pair. Context information extraction module extracts different context relation features from the document according to mentions' position information of entity pair. We have conducted experiments on three document-level relation extraction datasets, and the effect has been significantly improved.

Key words Document-level relation extraction, entity information, context information, mentions' position information, cross-sentence inference

Citation Huang He-Yan, Yuan Chang-Sen, Feng Chong. Document-level relation extraction with entity and context information. *Acta Automatica Sinica*, 2024, 50(10): 1953–1962

篇章级别的关系抽取任务是从篇章中识别实体对的关系，它在现实生活中有许多实际的应用，例如，问答系统^[1-2]和大规模知识图谱^[3-5]构建。尽管传统的句子级别关系抽取方法^[6-10]取得了巨大的成功，但它们受到自身任务的限制，在捕获跨句子实体对之间的关系时，抽取效果难以达到人们的期望。因此，关系抽取的研究从句子级别转移到篇章级别是非常有必要的。

最近，Yao 等^[11]提出一个大规模的人工标注的

篇章级别关系抽取数据集 DocRED，其中包含海量的关系事实和文档。但是篇章级别关系抽取任务存在两个主要问题：第一，实体对通常存在于不同的句子中，它们之间的关系需要考虑一个句子或者多个句子的内容；第二，每个实体会存在多个提及，并且它们也有可能出现在不同的句子中，实体信息分散在整个篇章中。为了解决这些问题，Zhou 等^[12]提出两种新方法，即自适应阈值化和局部上下文池化。Zeng 等^[13]提出图聚合的推理网络和异构图模型来解决篇章关系抽取问题。但是，如果不考虑以下两个问题，它们的抽取结果仍不尽人意：

1) 多实体多标签问题。在篇章级别的关系抽取任务中，篇章中存在多个实体，并且同一个实体可能与多个实体之间存在不同的关系。如图 1 所示，一篇文档中存在着多个实体，任意两个实体之间都存在着某种关系（在此认为没有关系“NA”也是一种关系）。实体对 *Emancipation Proclamation* 和

收稿日期 2022-12-12 录用日期 2023-03-29
Manuscript received December 12, 2022; accepted March 29, 2023
本文责任编辑 张民
Recommended by Associate Editor ZHANG Min
1. 北京理工大学计算机学院 北京 100081 2. 北京理工大学自然语言处理实验室 北京 100081
1. School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081 2. Natural Language Processing Laboratory, Beijing Institute of Technology, Beijing 100081

<p>[0] Washington Place (William Washington House) is one of the first homes built by freed slaves after the <i>Emancipation Proclamation</i> of 1863 in <i>Hampshire County, West Virginia, United States</i> [2] <i>William Washington</i> later acquired other properties on ... Route 28 and became the first African - American land developer in the state of <i>West Virginia</i> . [3] One of his subdivisions is the " Blacks Hill " neighborhood of Romney , adjacent to the Washington[4] Washington Place was bought and restored by <i>Ralph W. Haines</i>, a local attorney and historic preservationist .</p>
<p>Subject: <i>Emancipation Proclamation</i></p> <p>Object: <i>United States</i></p> <p>Relation: <i>Country</i></p>
<p>Subject: <i>West Virginia</i></p> <p>Object: <i>United States</i></p> <p>Relation: <i>Contain, Located in the administrative territorial entity</i></p>
<p>Subject: <i>William Washington</i></p> <p>Object: <i>United States</i></p> <p>Relation: <i>Country of citizenship</i></p>

图 1 篇章级别关系抽取数据集 DocRED 中的一个实例
Fig.1 An example of document-level relation extraction dataset DocRED

United States 能够表示关系 *Country*. 实体对 *William Washington* 和 *United States* 表示的关系则是 *Country of citizenship*. 因此, 直观上认为在篇章关系抽取中实体对的两个实体对关系的影响是相同的这种思想具有一定局限性, 应该考虑实体对在表示关系的过程中哪一个实体包含更多的关系信息.

2) 复杂的上下文问题. 与传统的句子级别关系抽取相比, 篇章级别关系抽取需要考虑具有特定实体对的上下文信息. 主要有以下三个原因: a) 有些实体对的关系信息位于实体的上下文信息中, 仅仅使用实体信息很难预测实体对的关系. 例如, 在识别 *Hampshire County* 和 *West Virginia* 时, “in” 可以作为提示词, 很容易识别出这两个实体之间的关系. b) 篇章中包含多种关系信息, 直接使用篇章信息作为实体对的上下文信息会引入噪声信息, 导致模型的抽取结果变差. c) 在篇章关系抽取数据集的标注过程中, 假设两个实体的提及存在关系, 则这两个实体存在关系. 如图 1 所示, 实体对 *West Virginia* 和 *United States* 存在两种关系, *West Virginia* 存在两个提及, 分别在句子 0 和句子 2 中, 但是能够表示关系信息的上下文信息主要位于句子 0 中 (*West Virginia* 的第一个提及). 因此, 在考虑篇章级别关系抽取的实体对的上下文信息时, 应该从

提及层次抽取实体对的丰富上下文信息, 并对其筛选.

本文提出使用实体和上下文信息来捕获实体对的关系信息. 为了解决多实体多标签问题, 假设实体对中的两个实体对关系的贡献程度是不同的. 因此, 本文对实体的信息进行筛选, 抽取出能够表示关系信息的实体信息. 为了解决复杂的上下文问题, 将实体对的提及进行两两组合, 根据提及的位置对篇章进行切分, 使用 max-pooling 抽取出不同的上下文信息, 然后使用注意力机制抽取出重要的上下文信息.

本文的主要贡献包括: 1) 提出对实体进行筛选, 筛选出重要的关系信息, 进而解决多实体多标签问题; 2) 提出使用提及的位置信息, 从篇章中抽取出复杂的多样性的上下文信息; 3) 对数据集进行广泛实验并开展详细分析, 结果表明本文模型取得了更好的效果.

本文内容安排如下: 第 1 节介绍相关工作; 第 2 节给出模型的具体架构; 第 3 节进行实验分析; 第 4 节进行总结与展望.

1 相关工作概述

关系抽取的任务是从纯文本中识别实体对之间的关系. 传统方法侧重于句子级别的关系抽取, 捕获单个句子中实体对的关系. 关系抽取利用卷积神经网络^[4] (Convolutional neural networks, CNN)、长短期记忆网络 (Long short term memory, LSTM) 或循环神经网络^[15-16] (Recurrent neural network, RNN) 来编码句子的特征, 进而捕获实体对的关系特征向量. Miwa 等^[8] 提出将外部句法信息引到模型中, 构建了树形 LSTM 模型. Christopoulou 等^[17] 在实体对之间构成 l -长度的步长, 以区分两个实体之间的不同关系路径. 但是, 这些方法都是句子级别的关系抽取方法, 在篇章级别的关系抽取中难以取得较好的效果.

近些年, 研究者们开始探索篇章级别关系抽取, 其中一个重要原因就是篇章关系抽取更加贴合实际应用, 大量的关系事实蕴藏在篇章之中. 实体对的关系由多个句子共同支撑, 而不是依靠单一句子就能够识别实体对的关系. 传统的句子级别的关系抽取模型在篇章级别的关系抽取任务中不尽人意. 在篇章级别关系抽取的工作中, Christopoulou 等^[18] 通常依赖于明确的共同引用或篇章中的单个事件^[19-20]. Jia 等^[21] 提出将文档和子关系的多层次结合起来进行文本表示学习. Verga 等^[22] 提出一种基于 Transformer 的模型, 通过融合多个提及信息进行多实例

学习的篇章级别关系抽取. Yao 等^[11] 提出一个通用领域的篇章级别关系抽取数据集和一些基线模型. 然而, 本文方法是基于来自同一文档的不同实体对来捕获不同的特征.

Nan 等^[23] 构造了潜在结构细化模型, 该模型通过自动引入潜在篇章级别的图来增强句子间的关系推理, 并提出一种细化策略使模型能够聚合相关关系信息. Wang 等^[24] 根据实体全局和局部表示以及上下文关系表示对文档信息进行编码, 进而预测篇章中实体对的关系. Zhou 等^[12] 提出使用自适应阈值和局部的上下池化来进行多标签分类, 其中局部的上下池化使用 BERT (Bidirectional encoder representation from Transformers) 模型的注意力矩阵来进行计算. Zeng 等^[13] 构建一个异构的提及和图来模拟不同提及之间的复杂交互, 并构建一个实体级别的图结构来推断实体之间的关系.

相较于前人的工作, 为了解决多实体多标签问题, 本文通过增强实体对中某一个实体的信息来捕获重要的关系信息, 最主要的假设就是实体对中实体的重要程度是不同的. 为了解决复杂的上下文问题, 将如何捕获实体对的上下文下沉到捕获提及对的上下文, 捕获实体对更加丰富的关系信息, 以增强实体对的关系表示.

2 模型

2.1 概述

模型的主要目的是识别篇章中实体对的关系. 该模型的主要模块有两个, 分别是实体信息抽取模块和上下文信息抽取模块, 如图 2 所示.

实体信息抽取模块的目的是使用关系特征对实体对的信息进行区分, 自动选择实体对中重要的实体. 之所以对实体对中的实体进行区分, 是因为某些时候同一个实体与不同实体组合会表示不同的关系, 说明该实体存在较少的关系特征, 关系信息主要存在于另外一个实体中. 而针对不同的关系对实体对进行筛选, 可以解决多实体多标签问题. 因此, 需要增加实体信息从而降低多实体多标签问题对关系抽取模型的影响.

上下文信息抽取模块的目的是抽取特定实体对在篇章中的上下文信息. 具体做法是将实体对中实体的提及进行对应组合, 然后使用提及的位置信息对篇章进行切分, 抽取不同的上下文关系特征. 原因是实体对的关系信息并不一定都存在于实体对中, 有时候在实体对的上下文中, 仅使用实体对的信息会忽略掉一些重要的关系特征.

2.2 BERT 编码器

BERT^[25-26] 是一种语言模型, 它利用双向注意力机制和大规模无监督语料库来获得每个词的优秀上下文向量表示. BERT 在多层上重复编码给定的输入表示, 并由掩码多头自注意力和位置前馈操作组成. 当使用 BERT 时, 将如下表示的序列输入 BERT 编码器中: “[CLS] He lived in Beijing [SEP] And...”. [SEP] 标记表示每个序列的结束位置, [CLS] 表示整个序列的开始位置和隐藏层的信息.

首先, 为有效利用子词 (Sub-word) 信息, 模型使用 BPE^[27] (Byte pair encoding) 对整个篇章 Doc 的单词进行切分, $Doc = [c_1, c_2, \dots, c_n]$, 其中, 包含 n 个标识 (Token). 其次, 模型使用 BERT

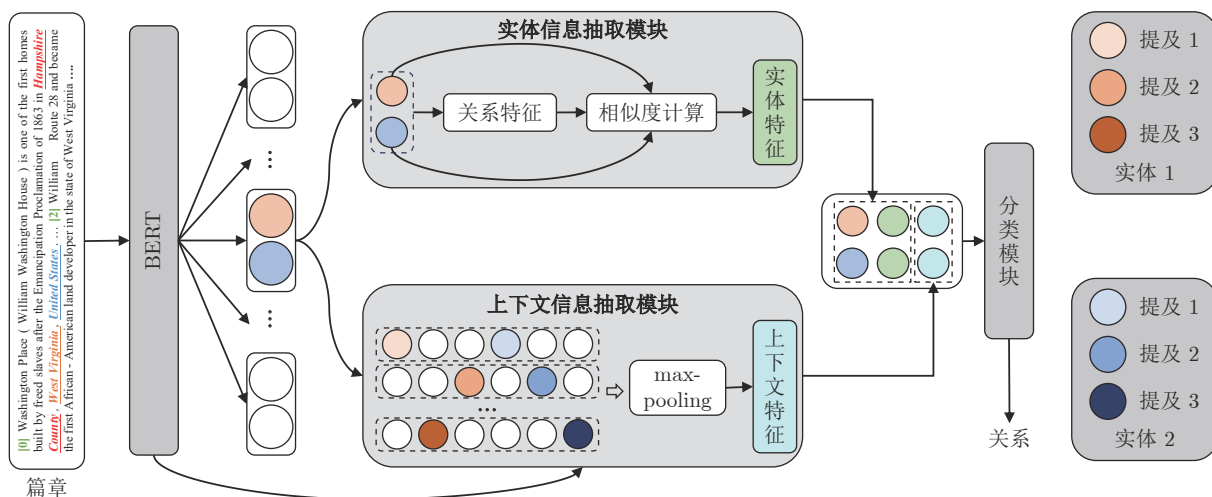


图 2 模型框架图主要有两个部分, 分别是实体信息抽取模块和上下文信息抽取模块

Fig. 2 Architecture of the proposed model, which contains two parts: Entity information extraction module and context information extraction module

编码器对整篇文章进行编码, 抽取出其中重要的关系特征, 具体操作如下

$$[CLS, C] = \text{BERT}([c_1, c_2, \dots, c_n]) \quad (1)$$

$$H = \sigma(W_c C + b_c) \quad (2)$$

其中, c_i 表示篇章中每个单词的表示, CLS 表示整个篇章的特征, C 表示篇章中每个标识经过 BERT 编码器之后的隐藏层向量, W_c 和 b_c 分别表示权重和偏置量. $\sigma(\cdot)$ 表示激活函数, 通常使用 \tanh 激活函数. H 表示单词的标识经过线性变换以后的隐藏层向量, 其中 $H = [h_1, h_2, \dots, h_n]$, $H \in \mathbf{R}^{n \times d}$.

本文使用 Yao 等^[11] 提出的方法来抽取实体特征和提及特征. 包含 k 个提及的实体 e_i 使用均值来计算实体的特征, 计算方法如下

$$e_i = \frac{1}{k} \sum_{i=1}^k m_i \quad (3)$$

其中, m_k 表示实体的第 k 个提及. m_k 在篇章中是从第 s 个标识到第 t 个标识, 计算方法如下

$$m_k = \frac{1}{t-s+1} \sum_{j=s}^t h_j \quad (4)$$

2.3 实体信息抽取模块

实体信息抽取模块的主要任务是对实体对的信息进行筛选, 从实体对中抽取出重要的关系信息. 在篇章级别关系抽取任务中, 通常某个实体与多个实体存在着多种关系或者与多个实体存在着相同的关系. 这种现象表明实体对中两个实体包含关系信息量是不同的, 它们之间存在着差异. 因此, 使用实体对的关系信息对实体进行筛选, 从实体对中抽取出重要的关系特征.

首先, 模型使用实体对关系信息进行粗略表示. 计算方式如下

$$r_e = \tanh(e_i^T W_e e_j + b_e) \quad (5)$$

其中, e_i 和 e_j 表示实体对的隐藏层向量, 计算方法如式 (3) 所示. r_e 表示实体对的关系信息, $W_e \in \mathbf{R}^{d \times d}$ 和 $b_e \in \mathbf{R}^d$ 分别表示矩阵的权重和偏置量. 然后, 使用余弦相似度计算实体对 (e_i, e_j) 与关系特征 r_e 之间的相似度, 本文认为相似度高的实体包含更多的关系信息. 因此, 将相似度较低的实体舍去. 计算方式如下

$$\begin{cases} \alpha_i = \frac{r_e \cdot e_i}{\|r_e\| \|e_i\|} \\ \alpha_j = \frac{r_e \cdot e_j}{\|r_e\| \|e_j\|} \end{cases} \quad (6)$$

其中, “ \cdot ”表示欧几里得点积. 最后, 抽取出重要的

实体关系向量 e_r , 计算方式如下

$$e_r = \begin{cases} e_i, & \alpha_i \geq \alpha_j \\ e_j, & \alpha_j > \alpha_i \end{cases} \quad (7)$$

2.4 上下文信息抽取模块

上下文信息抽取模块的主要任务是从整个篇章信息中抽取出特定实体对的关系特征. 原因是篇章的整体特征包含多个实体对的关系特征, 需要对篇章信息进行过滤, 删除掉无关的噪声信息. 因此, 本文提出使用实体对的上下文特征抽取模块.

首先, 将实体对 (e_i, e_j) 中的每个提及进行组合. 假设实体 $e_i = [m_i^1, m_i^2, \dots, m_i^{k_1}]$, 包含 k_1 个提及; 实体 $e_j = [m_j^1, m_j^2, \dots, m_j^{k_2}]$, 包含 k_2 个提及. 提及的组合存在 $k_1 \times k_2$ 个, 例如, 提及 m_i^1 和 e_j 中的每个提及进行组合, 生成能够表示关系特征的提及对. 其次, 利用提及对在篇章中的位置将篇章切分成三个部分, 然后使用 max-pooling 抽取出提及对的关系信息. 计算方法如下

$$c'_i = [\max(h_{0:p}), \max(h_{p:q}), \max(h_{q:n})] \quad (8)$$

$$c_i = \tanh(W_{c1} c'_i + b_{c1}) \quad (9)$$

其中, p 和 q 分别表示提及对的位置, c_i 表示第 i 个提及对的上下文关系特征, $W_{c1} \in \mathbf{R}^{3d \times d}$ 和 $b_{c1} \in \mathbf{R}^d$ 分别是矩阵参数和偏置量. 由于提及对在篇章中位于不同位置, 因此, 可以抽取出不同的上下文特征. 然后使用注意力机制的方法抽取出提及对的上下文关系特征, 计算方法如下

$$\begin{cases} c = \sum_{i=1}^{k \times (k-1)} \beta_i c_i \\ \beta_i = \text{softmax}(W_{c2} c_i) \end{cases} \quad (10)$$

其中, β_i 表示提及对的关系特征的权重, W_{c2} 表示权重参数, k 表示实体的提及个数, c 表示实体对的上下文关系特征. 最后将实体信息和上下文信息进行融合, 获得实体对的关系特征, 计算方式如下

$$\begin{cases} z_i = \text{FFNN}([e_i, e_r, c]) \\ z_j = \text{FFNN}([e_j, e_r, c]) \end{cases} \quad (11)$$

其中, $z_i \in \mathbf{R}^d$ 和 $z_j \in \mathbf{R}^d$ 表示实体对的关系特征向量, FFNN 表示前馈神经网络.

2.5 分类模块

给定实体对的关系特征 (z_i, z_j) , 模型使用双线性变换计算实体对的关系概率 P , 并使用 Sigmoid 函数作为激活函数, 计算过程如下

$$P(r_e | e_i, e_j) = \sigma(z_i^T W_r z_j + b_r) \quad (12)$$

其中, $\mathbf{W}_r \in \mathbf{R}^{d \times d}$, $\mathbf{b}_r \in \mathbf{R}^d$ 是模型的参数. 为了减少双线性变换的参数数量 (式 (5) 和式 (12)), 使用双线性组进行计算, 它将隐藏层维度切分成 l 个大小相等的组, 然后再使用双线性变换计算, 具体计算过程如下

$$\begin{cases} \mathbf{z}_i = [\mathbf{z}_i^1, \mathbf{z}_i^2, \dots, \mathbf{z}_i^l] \\ \mathbf{z}_j = [\mathbf{z}_j^1, \mathbf{z}_j^2, \dots, \mathbf{z}_j^l] \\ P(\mathbf{r}_e | \mathbf{e}_i, \mathbf{e}_j) = \sigma \left(\sum_{n=1}^l (\mathbf{z}_i^n)^\top \mathbf{W}_r^n \mathbf{z}_j^n + \mathbf{b}_r^n \right) \end{cases} \quad (13)$$

其中, $\mathbf{W}_r^n \in \mathbf{R}^{(d/l) \times (d/l)}$ 和 $\mathbf{b}_r^n \in \mathbf{R}^{d/l}$ 表示模型的参数. Zhou 等^[12] 提出使用自适应阈值来表示损失函数, 基于此, 本文使用该损失函数来计算损失值, 并使用 Dropout 函数防止过拟合.

3 实验

在本节中, 首先描述数据集和实验细节, 然后给出实验结果和分析, 最后进行案例研究.

3.1 数据集

本文在三个公开的篇章级别关系抽取数据集上评估所提模型, 数据集的统计结果如表 1 所示.

表 1 数据集的统计
Table 1 Statistics of the datasets

统计	DocRED	CDR	GDA
训练集	3053	500	23353
开发集	1000	500	5839
测试集	1000	500	1000
关系种类	97	2	2
每篇的关系数量	19.5	7.6	5.4

DocRED^[11]: 该数据集是从维基百科中构造的通用类型的篇章级别关系抽取数据集. 它包含 3 053 个人工标注的训练文档, 对于表达关系的实体对 (非 NA 关系), 大约 7% 的实体对具有多个关系标签.

CDR^[28]: 该数据集是人工标注的生物学领域的篇章数据集. 它包含 1 500 个篇章数据集, 并用来预测化学与疾病概念之间的相互关系.

GDA^[29]: 该数据集是生物学领域的大规模数据集. 它包含 23 353 篇用于训练的文章, 任务是预测基因与疾病概念之间的相互关系. 本文遵循 Christopoulou 等^[18] 的研究结果, 将训练集拆分为训练集和开发集.

3.2 实验设置

本文是使用 Huggingface 的 Transformers 实现

的. 在 DocRED 数据集上使用 cased-BERT_{BASE}^[25] 或 RoBERTa_{BASE}^[30] 作为编码器, 在 CDR 和 GDA 数据集上使用 SciBERT_{BASE}^[31]. 本文使用基于 Apex 库的混合精度^[32] 训练. 模型使用 AdamW^[33] 作为优化器, 学习率为 1×10^{-4} , 对前 6% 的步长使用衰减率为 0 的 linear warmup. 其中, 式 (13) 中 l 的大小为 64. 所有超参数都在开发集上进行了调整. 所有实验是在 TITAN XP 上运行的. 表 2 中列出了重要数据集上的超参数.

表 2 模型的超参数
Table 2 Hyper-parameters of model

参数名称	DocRED	CDR	GDA
批次大小	4	4	4
迭代次数	30	30	10
学习率 (编码)	5×10^{-5}	5×10^{-5}	5×10^{-5}
学习率 (分类)	1×10^{-4}	1×10^{-4}	1×10^{-4}
分组大小	64	64	64
Dropout	0.1	0.1	0.1
梯度裁剪	1.0	1.0	1.0

本文采用 F1 和 Ign F1 作为评价指标, 其中 Ign F1 表示不包括训练集与开发集/测试集共享的实体对的 F1 值, 原因是 DocRED 数据集中存在着开发集/测试集与训练集部分实体对重叠的现象.

3.3 基线模型

本文将提出的模型与以下模型进行对比实验.

Yao 等^[11] 提出了 4 个基线模型, 分别是 CNN、LSTM、Bi-LSTM、Context-Aware. 此外, 本文还将 5 种基于图卷积的关系抽取模型应用于 DocRED, 包括 GAT^[34]、BRAN^[22]、GCNN^[9]、EoG^[19] 和 AG-GCN^[10].

BERT-Two-Step^[35] 虽然类似于 BERT-RE_{BASE}, 但它首先预测两个实体是否有关系, 然后预测特定的目标关系.

HIN-GloVe/HIN-BERT_{BASE}, 由 Tang 等^[36] 提出. 分层推理网络 (Hierarchical inference network, HIN) 充分利用了实体级、句子级和篇章级的丰富信息, 并使用 GloVe^[37] 或 BERT 进行词嵌入.

LSR-GloVe/LSR-BERT, 由 Nan 等^[23] 提出潜在结构细化以端到端的方式构建用于推理的篇章级图, 而不依赖于共同引用或规则. BERT-E_{BASE}, 由 Zhou 等^[12] 使用双线性变化组来捕获关系特征并降低模型参数.

其他预训练模型, 如 BERT^[25]、GAIN^[13]、RoBERTa^[30]、CorefBERT^[38], 也用 Wang 等^[35] 的方法

作为编码器处理篇章级别关系抽取任务。

3.4 对比实验分析

本节使用本文方法 (融合实体和上下文信息的篇章关系抽取方法 $FECI_{BASE}$) 与基线方法进行对比, 实验结果见表 3 和表 4。

表 3 在 DocRED 开发集和测试集上的实验结果 (%)
Table 3 Experiment results on the development and test sets of DocRED (%)

模型	开发集		测试集	
	Ign F1	F1	Ign F1	F1
CNN	41.58	43.45	40.33	42.26
LSTM	48.44	50.68	47.71	50.07
Bi-LSTM	48.87	50.94	48.78	51.06
Context-Aware	48.94	51.09	48.40	50.70
HIN-GloVe	51.06	52.95	51.15	53.30
GAT-GloVe	45.17	51.44	47.36	49.51
GCNN-GloVe	46.22	51.52	49.59	51.62
EoG-GloVe	45.94	52.15	49.48	51.82
AGGCN-GloVe	46.29	52.47	48.89	51.45
LSR-GloVe	48.82	55.17	52.15	54.18
BERT-RE _{BASE}	—	54.16	—	53.20
RoBERTa _{BASE}	53.85	56.05	53.52	55.77
BERT-Two-Step _{BASE}	—	54.42	—	53.92
HIN-BERT _{BASE}	54.29	56.31	53.70	55.60
CorefBERT _{BASE}	55.32	57.51	54.54	56.96
LSR-BERT _{BASE}	52.43	59.00	56.97	59.05
BERT-E _{BASE}	56.51	58.52	—	—
GAIN _{BASE}	59.14	61.22	59.00	61.24
$FECI_{BASE}$	59.74	61.38	59.81	61.22

表 4 在 CDR 和 GDA 数据集上测试集 F1 值 (%)
Table 4 F1 values of test set on CDR and GDA datasets (%)

模型	CDR	GDA
BRAN	62.1	—
CNN	62.3	—
EoG	63.6	81.5
LSR-BERT	64.8	82.2
SciBERT _{BASE}	65.1	82.5
SciBERT-E _{BASE}	65.9	83.3
$FECI_{BASE}$	69.2	83.7

表 3 中, 在测试集和开发集上, 除 LSR-GloVe 模型外, GloVe 序列基线模型和图基线模型的性能明显低于最差的基于 BERT 的模型, 这是因为基于 BERT 的模型可以使模型学习到长距离的信息,

而且不会造成太多的损失. 在使用 BERT 或 BERT 变体的模型中, 与基线模型 LSR-BERT_{BASE} 相比, $FECI_{BASE}$ 在开发集上 F1 和 Ign F1 分别提高了 2.38% 和 7.31%; 在测试集上 F1 和 Ign F1 分别提高了 2.84% 和 2.17%. 这表明 $FECI_{BASE}$ 在篇章级别关系抽取任务中更有效, 因为 $FECI_{BASE}$ 不仅可以从实体对中抽取重要的关系信息, 而且可以从提及层次抽取实体的上下文信息. 另外, 在表 3 中, 除了测试集的 F1 指标外, $FECI_{BASE}$ 已经优于所有方法. 与 HIN-BERT_{BASE}、CorefBERT_{BASE} 和 BERT-E_{BASE} 相比, $FECI_{BASE}$ 在开发集中将 F1 分别提高了 5.07%、3.87% 和 2.86%. 结果表明, $FECI_{BASE}$ 可以从篇章中学习到关键的关系信息, 提取有关信息的更多详细信息以提高模型的性能.

两个生物医学领域的篇章关系抽取实验结果如表 4 所示. Verga 等^[22] (BRAN) 和 Nguyen 等^[39] (CNN) 都是基于序列的模型, 分别使用自注意力网络和 CNN 作为编码器. Christopoulou 等^[18] (EoG) 和 Nan 等^[23] (LSR-BERT) 使用基于图的模型, 通过启发式或结构化注意力构建篇章级别图, 并使用图神经网络进行推理. 在本文实验中, 使用 SciBERT_{BASE} 替换了 BERT_{BASE} 编码器, SciBERT_{BASE} 在科学出版物的多领域语料库上进行了预训练, SciBERT_{BASE} 基线已经优于所有现有方法. 而本文的方法进一步将 CDR 和 GDA 的 F1 值分别提高了 3.3% 和 0.4%.

3.5 消融实验

为进一步分析 $FECI_{BASE}$ 模型, 本节进行消融实验以说明模型中不同模块的有效性. 表 5 展示了相关实验结果 (P 表示模型的参数规模, T 表示每个 epoch 的运行时间), 结果表明模型的各个模块表现出了较好的效果. 原因是 $FECI_{BASE}$ 可以捕获实体对内部的关系信息和关键的上下文信息. w/o Entity 表示取消实体信息抽取模块, w/o Context 表示取消上下文特征抽取模块.

表 5 $FECI_{BASE}$ 在开发集上的消融研究结果
Table 5 Ablation study results of $FECI_{BASE}$ on the development set

模型	开发集			
	Ign F1 (%)	F1 (%)	P (M)	T (s)
$FECI_{BASE}$	59.74	61.38	133.4	2962.4
w/o Entity	58.16	60.07	132.2	2831.7
w/o Context	58.67	60.89	130.5	482.3

实体信息抽取模块. 从表 5 中可以发现, 当没有使用实体信息抽取模块时, 关系抽取效果变得比

较差. 具体来说, F1 和 Ign F1 的实验结果分别降低了 1.31% 和 1.58%. 由于同一个实体对可能有多个关系和多个提及, 因此, 识别实体对中的关键关系信息能够提高关系抽取模型的性能.

上下文特征抽取模块. 从表 5 中可以发现, 当删除了上下文特征抽取模块时, F1 和 Ign F1 的实验结果分别降低了 0.49% 和 1.07%. 主要原因是实体对的关系特征在篇章中的位置是不固定的, 并不是都存在于实体对的信息中. 因此, 忽略实体对的上下文信息会使篇章关系抽取模型的效果变差.

此外, 从表 5 中的运行时间上看, $FECI_{BASE}$ 和 w/o Entity 的运行时间相差不大, $FECI_{BASE}$ 和 w/o Context 相差巨大. 原因是在进行上下文抽取的过程中, 需要设置循环来抽取提及对的上下文信息. 当程序在 GPU 上运行时, 循环会导致运行时间变长. 当去除上下文特征抽取信息模块时, 程序的运行时间大幅度降低. 从参数量上看, 实体信息抽取模块和上下文特征抽取模块所需的参数量并不大, 不会极大地增加 GPU 的运行负担.

3.6 噪声实体和噪声上下文的影响

为了证明 $FECI_{BASE}$ 能够降低噪声实体和噪声上下文的影响, 设计如下两个实验.

噪声实体影响. 设计了 4 种实体信息表示: 1) 只使用头实体信息 (Head entity) 作为实体信息表示; 2) 只使用尾实体信息 (Tail entity) 作为实体信息表示; 3) 实体对信息 (Entity pair) 作为信息表示; 4) 对实体对进行筛选 (本文方法 $FECI_{BASE}$). 从表 6 中可以观察到:

1) 当仅仅使用头实体或者尾实体信息时, 关系抽取效果明显下降. 当使用 $FECI_{BASE}$ 时, 实验效果上升. 头实体、尾实体和 $FECI_{BASE}$ 的共同点是只使用了单一实体信息, 不同点是 $FECI_{BASE}$ 经过了实体筛选, 模型能够自动对实体进行选择.

2) 与使用实体对信息相比, $FECI_{BASE}$ 依然取得了不错的效果. 但是使用实体对信息明显比单独使用头实体信息或尾实体信息的效果好. 因此, 证明实体对中必然存在噪声信息, 仅使用实体对中的单一实体信息 (头实体信息或尾实体信息) 并不能够获得好的实验效果, 而使用 $FECI_{BASE}$ 能够自动地对噪声信息进行过滤, 提升实验结果.

噪声上下文影响. 设计了 4 种噪声上下文信息表示: 1) 传统上下文信息 (Tradition) 是指在使用 BERT 编码器以后, 直接将编码器输出的第一个 Token (CLS) 信息作为实体对的上下文信息; 2) 共现上下文 (Co-occurrence) 是指利用实体对所在的

句子作为上下文信息; 3) 非共现上下文 (Non co-occurrence) 是指使用不包含实体对的句子作为上下文信息; 4) 本文方法 $FECI_{BASE}$ 使用上下文特征抽取. 从表 6 中可以观察到:

1) 当使用传统上下文信息和非共现上下文信息时, 实验效果最差. 该实验结果证明了篇章中是存在噪声上下文信息的, 假设没有对其进行处理, 会影响关系抽取的效果.

2) 使用共现上下文信息比使用传统上下文信息和非共现上下文信息效果好. 原因是共现上下文信息对篇章中的噪声信息进行了过滤, 提供了有效的信息, 因此实验效果提升.

3) 共现上下文的方法比 $FECI_{BASE}$ 的效果差. 原因有两点: a) 实体对的关键信息可能位于实体对所在的句子中, 也可能在篇章的其他位置 (这个信息可以从每个实体对的证明句 (Evidence sentence) 中观察到, 训练集与开发集提供, 但测试集不提供, 因此, 不能直接使用). b) $FECI_{BASE}$ 在提及级别进行了精细化的上下文权重计算, 降低了实体对中无效提及对上下文信息的影响.

表 6 $FECI_{BASE}$ 在开发集上噪声实体和噪声上下文的实验结果 (%)

Table 6 The experiment results of noisy entity and noisy context of $FECI_{BASE}$ on the development set (%)

模型	开发集	
	Ign F1	F1
$FECI_{BASE}$	59.74	61.38
Head entity	58.42	60.14
Tail entity	57.97	60.08
Entity pair	58.91	60.85
Tradition	57.42	59.72
Co-occurrence	58.27	61.01
Non co-occurrence	56.72	58.86

3.7 不同上下文对关系抽取的影响

为了证明所提出的抽取上下文方法的有效性, 设计 3 种不同的上下文信息替代本文提出的上下文抽取信息, 它们分别是随机上下文信息、平均上下文信息和传统上下文信息. 随机上下文信息 (Random) 指在捕获提及对的上下文以后并没有使用注意力机制对上下文信息进行筛选, 直接随机抽取其中一个提及对的上下文信息作为实体对的上下文信息, 其目的是为了验证不同上下文信息对关系抽取的影响. 平均上下文信息 (Mean) 指在捕获提及对的上下文以后使用平均加权的方法融合所有提及对

的上下文信息,其思想是认为所有提及对的上下文信息具有相同的重要程度,没有对噪声上下文信息进行筛选.详细实验结果见表7.

表7 FE_{CI}_{BASE}在开发集上不同上下文信息的实验结果(%)

Table 7 The experiment results of different context information of FE_{CI}_{BASE} on the development set (%)

模型	开发集	
	Ign F1	F1
FE _{CI} _{BASE}	59.74	61.38
Random	58.47	60.61
Mean	59.56	60.94
Tradition	58.19	60.06

从表7中可以观察到:

1) 当使用随机上下文信息时,相较于均值上下文和 FE_{CI}_{BASE},实验效果有所降低,但是比传统上下文信息效果好.原因是随机上下文信息有可能包含某些实体对的关系信息(例如,实体对和提及对完全相同的情况,或者随机抽取的上下文信息恰巧包含了重要的关系信息),但是在大部分情况下,重要的关系信息会被忽略掉,降低了实验的总体效果.

2) 当使用平均上下文信息时,实验效果是除了 FE_{CI}_{BASE} 以外最好的.原因是平均上下文信息能够捕获到重要的信息,但是没有对捕获到的信息进行筛选.因此,相较于 FE_{CI}_{BASE},平均上下文信息引入了噪声信息导致实验效果变差.

3) 当使用传统上下文信息时,实验效果是最差的.原因是对于篇章中所有的实体对都拥有相同的上下文信息(CLS),没有区分度,导致实验效果降低.

3.8 效率研究

为了探索 FE_{CI}_{BASE} 的计算效率,从参数规模、训练时间和解码时间(Train T表示每个 epoch 的训练时间,Decoder T表示解码时间)进行比较.从表8中,可以观察到:

1) FE_{CI}_{BASE} 的参数规模相对来说并没有特别的大,由此可推断出本文的方法并不需要特别大的

表8 不同方法在开发集上的效率
Table 8 Efficiency of different methods on the development set

模型	开发集		
	P (M)	Train T (s)	Decoder T (s)
LSR-BERT _{BASE}	112.1	282.9	38.8
GAIN _{BASE}	217.0	2271.6	817.2
FE _{CI} _{BASE}	133.4	2962.4	829.0

显存就可以运行.

2) FE_{CI}_{BASE} 的训练时间和解码时间都比较长(时间为每个 epoch 的运行时间).原因是进行上下文抽取的过程中,需要设置循环来抽取提及对的上下文信息,而循环会造成程序在 GPU 上的运行时间增加.

3.9 案例分析

图3中展示了开发集中的一个实例,以便更好地了解 FE_{CI} 与 BERT、LSR 和 GAIN 的比较效果(这4个模型都是以 BERT_{BASE} 为基础进行的实验),其中相同颜色的单词表示相同的实体.从图中能够看到:

1) 从关系抽取效果上看,FE_{CI} 方法关系抽取的准确率明显高于 BERT、LSR 和 GAIN.

2) 对于 (Reichsgaue, German) 和 (Josef Grohé, Zivilverwaltung) 的关系, BERT 和 LSR 都没有识别出来,而 FE_{CI} 识别出来了.原因有两点: a) 实体对 (Reichsgaue, German) 中 German 的提及在篇章中分布广泛,有4个提及,导致 BERT 和 LSR 没有识别出实体对的关系类型; b) 实体对 (Josef Grohé, Zivilverwaltung) 中的两个实体在不同的句子中,并且距离较远,导致没有识别出来.而 FE_{CI} 能够很好地解决这些问题,识别出实体对的关系.

3) 可以明显观察到实体对 (Allies, World War II) 和 (World War II, Allies) 具有不同的关系.而这4种模型都没有预测出 (Allies, World War II) 的关系.这说明实体对的方向性对于篇章关系抽取很重要.但是目前仅仅依靠最后的双线性变换来区分实体对的方向性是不够的,后续需要加强这方面的研究.

4) 实体对 (Reichsgau Flandern, Reichsgaue) 的关系只有 GAIN 预测错误,原因可能是 GAIN 利用图矩阵表示提及和实体等信息之间的联系,而图矩阵过于复杂,致使模型学习到过多的噪声信息.

5) 以实体对 (Reichsgaue, German) 为例展示本文方法对不同上下文的抽取情况.在真实数据中,第四个句子是证明句,能够支撑这个实体对的关系.因为 German 有4个提及,Reichsgaue 只有一个提及,所以具有4个不同的上下文,它们之间具有不同的权重.其中最后两个 Germany3 和 German4 的权重较高,它们都位于第四个句子中,这也证明了本文方法的有效性.

4 结束语

本文采用一种新颖的实体和上下文信息抽取模

[0] The **Reichskommissariat of Belgium and Northern France** was a **German** civil administration (**Zivilverwaltung**) which governed **German** - occupied Belgium during **World War II** . [1] The **Reichskommissariat** replaced an earlier military government , the Military Administration in Belgium and Northern France , established in the same territory in 1940 . [2] On 18 July 1944 , the first Gauleiter , **Josef Grohé** , was named the first **Reichskommissar** of the.... [3] The territory was mostly liberated by the **Allies** in September 1944 , in the aftermath of the Normandy... [4] Following the liberation , the territory was retrospectively annexed directly into **Germany** (although no longer under de facto **German** control) as three separate **Reichsgaue** : **Reichsgau Flandern** , **Wallonien** and the Brussels district.

Entity Pair	BERT	LSR	GAIN	FECI	Context	Weight
Reichsgaue, German	✗	✗	✓	✓	German1 →Reichsgaue	0.13
Josef Grohé, German	✓	✓	✓	✓	German2 →Reichsgaue	0.21
Josef Grohé, Zivilverwaltung	✗	✗	✓	✓	Germany3 →Reichsgaue	0.36
Allies, World War II	✗	✗	✗	✗	German4 →Reichsgaue	0.30
World War II, Allies	✓	✗	✓	✓		
Wallonien, Reichsgaue	✓	✓	✓	✓		
Reichsgau Flandern, Reichsgaue	✓	✓	✗	✓		
...		

German1, German2, Germany3, German4 分别代表实体 German 的 4 个提及, 而 Weight 表示根据这 4 个提及对抽取出来的上下文信息的权重.

图 3 篇章级别关系抽取开发集中的一个实例分析

Fig. 3 An example analysis on the document-level relation extraction development set

型 (FECI), 该模型可以捕获实体对中重要的实体信息, 并抽取出实体对的上下文关系信息. 通过在三个数据集上进行实验, 结果表明本文方法是有效的, 并且明显优于基线模型. 通过大量的实验证明实体信息可以增强关系表示, 提高模型的效果. 而提取实体对的上下文信息能够进一步提高模型性能. 此外, 同一个提及、同一个实体、同一个证明句有多个关系是一个亟待解决的问题, 需要更多的关注和探索.

References

- 1 Yu M, Yin W P, Hasan K S, Santos C D, Xiang B, Zhou B W. Improved neural relation detection for knowledge base question answering. arXiv preprint arXiv: 1704.06194, 2017.
- 2 Chen Z Y, Chang C H, Chen Y P, Nayak J, Ku L W. UHop: An unrestricted-hop relation extraction framework for knowledge-based question answering. arXiv preprint arXiv: 1904.01246, 2019.
- 3 Yu H Z, Li H S, Mao D H, Cai Q. A relationship extraction method for domain knowledge graph construction. *World Wide Web*, 2020, **23**(2): 735–753
- 4 Ristoski P, Gentile A L, Alba A, Gruhl D, Welch S. Large-scale relation extraction from web documents and knowledge graphs with human-in-the-loop. *Journal of Web Semantics*, 2020, **60**: Article No. 100546
- 5 Macdonald E, Barbosa D. Neural relation extraction on Wikipedia tables for augmenting knowledge graphs. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. New York, USA: ACM, 2020. 2133–2136
- 6 Mintz M, Bills S, Snow R, Jurafsky D. Distant supervision for relation extraction without labeled data. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. Suntec, Singapore: ACL, 2009. 1003–1011
- 7 Lin Y K, Shen S Q, Liu Z Y, Luan H B, Sun M S. Neural relation extraction with selective attention over instances. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany: ACL, 2016. 2124–2133
- 8 Miwa M, Bansal M. End-to-end relation extraction using LSTMs on sequences and tree structures. arXiv preprint arXiv: 1601.00770, 2016.
- 9 Zhang Y H, Qi P, Manning C D. Graph convolution over pruned dependency trees improves relation extraction. arXiv preprint arXiv: 1809.10185, 2018.
- 10 Guo Z J, Zhang Y, Lu W. Attention guided graph convolutional networks for relation extraction. arXiv preprint arXiv: 1906.07510, 2019.
- 11 Yao Y, Ye D M, Li P, Han X, Lin Y K, Liu Z H, et al. DocRED: A large-scale document-level relation extraction dataset. arXiv preprint arXiv: 1906.06127, 2019.
- 12 Zhou W X, Huang K, Ma T Y, Huang J. Document-level relation extraction with adaptive thresholding and localized context pooling. arXiv preprint arXiv: 2010.11304, 2020.
- 13 Zeng S, Xu R X, Chang B B, Li L. Double graph based reasoning for document-level relation extraction. arXiv preprint arXiv: 2009.13752, 2020.
- 14 Santos C N D, Xiang B, Zhou B W. Classifying relations by ranking with convolutional neural networks. arXiv preprint arXiv: 1504.06580, 2015.
- 15 Cho K, Merriënboer B V, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv: 1406.1078, 2014.
- 16 Liu Y, Wei F R, Li S J, Ji H, Zhou M, Wang H F. A dependency-based neural network for relation classification. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Short Papers). Beijing, China: ACL, 2015. 285–290
- 17 Christopoulou F, Miwa M, Ananiadou S. A walk-based model on entity graphs for relation extraction. arXiv preprint arXiv: 1902.07023, 2019.
- 18 Christopoulou F, Miwa M, Ananiadou S. Connecting the dots:

- Document-level neural relation extraction with edge-oriented graphs. arXiv preprint arXiv: 1909.00228, 2019.
- 19 Yang B S, Mitchell T. Joint extraction of events and entities within a document context. arXiv preprint arXiv: 1609.03632, 2016.
- 20 Swampillai K, Stevenson M. Extracting relations within and across sentences. In: Proceedings of the Recent Advances in Natural Language Processing. Hissar, Bulgaria: DBLP, 2011. 25–32
- 21 Jia R, Wong C, Poon H. Document-level n -ary relation extraction with multiscale representation learning. arXiv preprint arXiv: 1904.02347, 2019.
- 22 Verga P, Strubell E, McCallum A. Simultaneously self-attending to all mentions for full-abstract biological relation extraction. arXiv preprint arXiv: 1802.10569, 2018.
- 23 Nan G S, Guo Z J, Sekulic I, Lu W. Reasoning with latent structure refinement for document-level relation extraction. arXiv preprint arXiv: 2005.06312, 2020.
- 24 Wang D F, Hu W, Cao E, Sun W J. Global-to-local neural networks for document-level relation extraction. arXiv preprint arXiv: 2009.10359, 2020.
- 25 Devlin J, Chang M W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional Transformers for language understanding. arXiv preprint arXiv: 1810.04805, 2019.
- 26 Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017. 6000–6010
- 27 Sennrich R, Haddow B, Birch A. Neural machine translation of rare words with subword units. arXiv preprint arXiv: 1508.07909, 2016.
- 28 Li J, Sun Y P, Johnson R J, Sciaky D, Wei C, Leaman R, et al. BioCreative V CDR task corpus: A resource for chemical disease relation extraction. *The Journal of Biological Databases and Curation*, **2016**: Article No. baw068
- 29 Wu Y, Luo R B, Leung H C M, Ting H, Lam T. RENET: A deep learning approach for extracting gene-disease associations from literature. In: Proceedings of the International Conference on Research in Computational Molecular Biology. Washington, USA: Springer, 2019. 272–284
- 30 Liu Y H, Ott M, Goyal N, Du J F, Joshi M, Chen D Q, et al. RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv: 1907.11692, 2019.
- 31 Beltagy I, Lo K, Cohan A. SciBERT: A pretrained language model for scientific text. arXiv preprint arXiv: 1903.10676, 2019.
- 32 Micikevicius P, Narang S, Alben J, Diamos G, Elsen E, Garcia D, et al. Mixed precision training. arXiv preprint arXiv: 1710.03740, 2018.
- 33 Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv preprint arXiv: 1711.05101, 2019.
- 34 Velickovic P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. arXiv preprint arXiv: 1710.10903, 2018.
- 35 Wang H, Focke C, Sylvester R, Mishra N, Wang W. Fine-tune BERT for DocRED with two-step process. arXiv preprint arXiv: 1909.11898, 2019.
- 36 Tang H Z, Cao Y N, Zhang Z Y, Cao J X, Fang F, Wang S, et al. HIN: Hierarchical inference network for document-level relation extraction. arXiv preprint arXiv: 2003.12754, 2020.
- 37 Pennington J, Socher R, Manning C D. GloVe: Global vectors for word representation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: Association for Computational Linguistics, 2014. 1532–1543
- 38 Ye D M, Lin Y K, Du J J, Liu Z H, Sun M S, Liu Z Y. Coreferential reasoning learning for language representation. arXiv preprint arXiv: 2004.06870, 2020.
- 39 Nguyen D Q, Verspoor K. Convolutional neural networks for chemical-disease relation extraction are improved with character-based word embeddings. arXiv preprint arXiv: 1805.10586, 2018.



黄河燕 北京理工大学计算机学院教授。主要研究方向为语言信息智能化处理, 社交网络, 数据分析和云计算。

E-mail: hhy63@bit.edu.cn

(HUANG He-Yan Professor at the School of Computer Science and Technology, Beijing Institute of

Technology. Her research interest covers intelligent processing of language information, social network, data analysis, and cloud computing.)



袁长森 北京理工大学计算机学院博士后。主要研究方向为知识图谱, 信息抽取。本文通信作者。

E-mail: yuanchangsen@bit.edu.cn

(YUAN Chang-Sen Postdoctor at the School of Computer Science and Technology, Beijing Institute of

Technology. His research interest covers knowledge graph and information extraction. Corresponding author of this paper.)



冯冲 北京理工大学计算机学院教授。主要研究方向为机器翻译, 信息抽取和信息检索。

E-mail: fengchong@bit.edu.cn

(FENG Chong Professor at the School of Computer Science and Technology, Beijing Institute of

Technology. His research interest covers machine translation, information extraction, and information retrieval.)