

# 基于多层次语义感知的中文关系抽取研究

付学敬<sup>1</sup>, 丁肖摇<sup>2</sup>

(1. 上海市市场监督管理局信息应用研究中心, 上海 200032;

2. 战略支援部队信息工程大学, 河南 郑州 450000)

**摘要:** 关系抽取是构建知识图谱的基础, 而中文关系抽取也是关系抽取中的难点问题, 现有的中文关系抽取大多采用基于字符特征或者词特征的方法, 但是前者无法捕获字符上下文的信息而后者受制于分词质量, 导致中文关系抽取的性能较低。针对该问题, 提出了基于多层次语义感知的中文关系抽取模型, 该模型利用实体间丰富的语义信息来提高实体对关系预测的性能。多层次语义感知体现在以下三个方面: 首先, 利用 ERNIE 预训练语言模型将文本信息转化为动态词向量; 然后, 利用注意力机制增强实体所在句子的语义表示, 同时通过外部知识尽可能地消除实体词的中文歧义; 最后, 将包含多层语义感知的句子表示放入到分类中进行预测。实验结果表明, 所提模型在中文关系抽取的性能上优于已有模型, 且更具解释性。

**关键词:** 知识图谱; 中文; 关系抽取; 多层次; 语义感知

中图分类号: TP391

文献标识码: A

文章编号: 1673-629X(2024)01-0143-07

doi: 10.3969/j.issn.1673-629X.2024.01.021

## Research on Chinese Relation Extraction Based on Multi-level Semantic Perception

FU Xue-jing<sup>1</sup>, DING Xiao-yao<sup>2</sup>

(1. Information Application Research Center of Shanghai Municipal Administration for

Market Regulation, Shanghai 200032, China;

2. PLA Strategic Support Force Information Engineering University, Zhengzhou 450000, China)

**Abstract:** Relation extraction is the basis of constructing knowledge graphs, and Chinese relation extraction is also a difficult problem in relation extraction. Most existing Chinese relation extraction methods use character-based or word-based features, but the former cannot capture contextual information of characters and the latter is limited by the quality of word segmentation, resulting in lower performance of Chinese relation extraction. In response to this problem, a Chinese relation extraction model based on multi-level semantic perception is proposed. This model uses rich semantic information between entities to improve the performance of predicting relationships between entities. Multi-level semantic perception is reflected in the following three aspects: firstly, text information is transformed into dynamic word vectors using the pre-training language model ERNIE; then, attention mechanism is used to enhance the semantic representation of the sentence where the entity is located, while external knowledge is used to eliminate Chinese ambiguity of entity words as much as possible; finally, the sentence representation containing multi-level semantic perception is put into classification for prediction. Experimental results show that the proposed model outperforms existing models in Chinese relation extraction performance and is more interpretable.

**Key words:** knowledge graphs; Chinese; relation extraction; multi-level; semantic perception

## 0 引言

关系抽取是信息抽取的一个重要子任务, 关系抽取需要在理解语义的基础上从非结构化文本中抽取实体对关系, 其在知识图谱构建中扮演着重要角色。

中文关系抽取作为关系抽取的重要分支, 在中文语义理解、中文知识库构建等下游任务中至关重要, 但是由于中文语义的复杂性以及汉语词意的多样性, 导致专家学者对中文文本关系抽取的研究相对较少, 更多放

收稿日期: 2023-04-14

修回日期: 2023-08-15

基金项目: 2023 年度上海市市场监督管理局科技项目(2023-55); 2022 年度河南省重点研发与推广专项(科技攻关)项目(222102210283)

作者简介: 付学敬(1971-), 男, 硕士, 高级工程师, CCF 会员(O4203M), 通讯作者, 研究方向为自然语言处理和文本数据挖掘; 丁肖摇(1990-), 男, 博士, 研究方向为关系抽取和知识图谱。

在了数据集丰富的英文文本的关系抽取上。但是在现实生活中,所接触的更多的是中文文本的表达,面对海量的中文文本数据,中文关系抽取技术可以更好地服务现实需求,是构建中文知识库的关键模块<sup>[1]</sup>。因此,构建一个高效且稳定的中文关系抽取模型就显得尤为重要。

目前,研究人员更多的是关注英文文本的关系抽取<sup>[2-5]</sup>,但是中文文本的体量也是很庞大的,中文关系抽取的进展关系到中文文本知识图谱的构建水平,也会间接推动中文语料库的发展,所以构建行之有效的中文关系抽取模型也是当前的一项重要任务。鉴于此,该文对中文关系抽取模型开展研究。

与英文关系抽取相比,中文文本与之差别较大,中文关系抽取面临着更为严苛的挑战。首先,中文文本中包含着丰富的语义,而句子形式结构上较英文文本松散;其次,中文文本中有较多的虚词做为句子内或者句子间的连接,而英文文本依靠词汇序列的不同来表达句子语义和句子结构,这两方面的挑战说明理解句子语义对于中文文本的关系抽取至关重要。如图 1 例子所示,首先在“孟德尔研究所有杜鹃”这句话中,实体“孟德尔”和实体“杜鹃”的关系受到分词质量的影响,比如可以按照“孟德尔/研究/所有/杜鹃”这样划分,那么“孟德尔”和“杜鹃”的关系就是“研究”,另一种按照“孟德尔/研究所/有/杜鹃”划分方式,“孟德尔”和“杜鹃”就是“所有者”的关系,如果单独拿出来这个句子,这两种的分词结果都是正确的。其次,这句话中的“杜鹃”也有两种含义,分别是“杜鹃鸟”和“杜鹃花”,这两种含义放在句子中都是成立的,但是在中文这是有歧义的,而且是中文文本中很常见的歧义。这就对中文关系抽取模型提出了更高的要求,首先对于词的分割,要包含动态词向量,也就是说要根据句子的上下文对分词进行改进,其次要利用注意力机制,对实体的词义要包含句子的语义和整个文本的信息,这样实体的词义才是准确的,才能有效避免歧义情况。

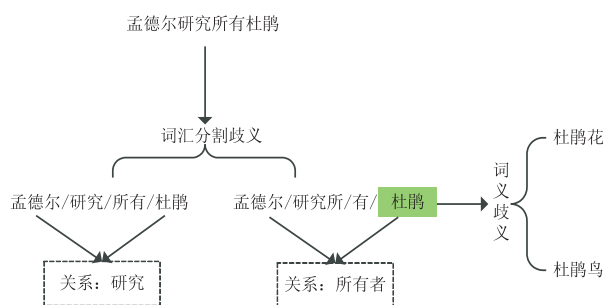


图 1 中文关系抽取例子

该文提出了基于多层语义感知的中文关系抽取模型。首先将含有目标实体的中文句子输入到 ERNIE<sup>[6]</sup> 预训练模型中得到动态词向量,然后利用注

意力机制<sup>[7]</sup>进一步丰富句子中实体的词义,并且生成含有句子信息的向量表示;为了进一步降低中文词的歧义,模型引入外部知识,使句子中的实体在外部知识的引导下提高实体表达的准确性。这几个阶段也是有针对性地解决中文关系抽取中的实体语义缺乏以及实体歧义的问题,使模型能够将含有更为丰富的语义信息的实体传入到分类器中,实现中文关系抽取的更高性能。

所提模型在中文关系抽取数据集上进行了实验,实验结果表明所提模型在性能上优于现有的中文关系抽取模型。

## 1 相关工作

近年来,神经网络和深度学习技术得到了广泛和深入的应用,中文关系抽取在自然语言处理领域得到了迅速发展,相关的中文关系抽取文献数量也在不断增加。关系抽取模型的构建方法通常可以分为两类:基于传统神经网络模型和基于预训练语言模型。

### 1.1 基于传统神经网络的模型

在基于传统神经网络的模型中主要应用卷积神经网络(Convolutional Neural Network, CNN)和循环神经网络(Recurrent Neural Network, RNN)。Liu 等人<sup>[8]</sup>开创性地提出应用 CNN 学习文本语义特征最终实现关系抽取。之后,研究人员对 CNN 进行丰富和扩展,相继提出带有最大池化的 CNN 抽取模型<sup>[9]</sup>、结合注意力机制的 CNN 抽取模型<sup>[10]</sup>和多通道的 CNN 抽取模型<sup>[11]</sup>。虽然 CNN 模型在并行计算上有着独特的优势,但是在中文文本的语义理解和语境建模方面有着明显的劣势。接下来研究人员将目标放在了 RNN 上, Zhang 等人<sup>[12]</sup>率先提出将 RNN 应用在关系抽取模型上并取得了更好的抽取性能。作为 RNN 变体的 LSTM(Long Short-Term Memory)在中文关系抽取模型上也得到了广泛应用, Zhang 和 Yang<sup>[13]</sup>提出 Lattice+LSTM 模型,在 LSTM 上引入了词汇信息可以更好地解决字符和词汇特征融合问题,但是在分割词义上仍然受到分词质量的制约。针对此问题, Li 等人<sup>[14]</sup>提出多粒度格框架模型,将字符和字符序列的潜在词义信息作为输入,在一定程度上解决实体中的歧义问题。但是传统的神经网络在处理实体对长距离依赖上有着天然的劣势,所以更多的研究者尝试利用预训练语言模型来解决中文关系抽取中的实体对长距离依赖问题。

### 1.2 基于预训练语言的模型

近年来,预训练语言模型在自然语言处理领域中大放异彩,其在中文关系抽取任务上的应用实现了更好的性能。针对上述 Li 等人<sup>[14]</sup>使用 LSTM 造成字符

信息丢失以及词汇信息不能共享的问题,Kong 等人<sup>[15]</sup>提出在编码层将 LSTM 和 BERT<sup>[16]</sup>相结合的方式,以实现字符表示包含所有匹配信息词,这样也就缓解了信息丢失问题。Eberts 和 Ulges<sup>[17]</sup>提出以预训练语言模型 BERT 为基础同时引入跨度(Span)思想的关系抽取模型。由于需要计算每个字符之间的跨度,造成时间复杂度较高。Zhong 和 Chen<sup>[18]</sup>提出对实体模型和关系模型分别使用两个不同的编码器来学习不同特征,这里的编码器是 BERT 和 ALBERT<sup>[19]</sup>。采用不同的编码器有助于不同特征信息的表示但是却割裂了实体表示和关系抽取之间的信息共享。Zhou 和 Chen<sup>[20]</sup>提出实体表示改进技术进而提高抽取性能,其中在编码器上分别使用预训练语言模型 BERT 和 RoBERTa<sup>[21]</sup>,RoBERTa 在 BERT 基础上进行了全面优化,包括应用动态掩码技术以及使用整句输入等。Cui 等人<sup>[22]</sup>提出预训练语言模型 MacBERT,在中文关系抽取上获得了比 RoBERTa 更好的效果,需要注意的是 MacBERT 也是对 BERT 的进一步改进。但是该模型仍然缺乏外部知识,对于歧义的词义判断效果不佳。Yang 等人<sup>[23]</sup>提出了以 BERT 作为编码器的混合视角专家框架,框架可以动态学习语义的多视角特征。将不同粒度和视图信息与预训练模型相结合,有利于中文关系抽取。Zhao 等人<sup>[24]</sup>提出一种多义性回馈机制来解决词的多义性问题,模型在编码模块采用了 CNN 和 RoBERTa 相结合的方式有效表示了不同粒度的信息特征,但是可以发现基于词的表示性能要低于基于字符的表示。基于字符和词的融合模型虽然很好地解决了两者的缺点,但是利用上下文信息仍然有限,导致实体词义表示不清,从而产生歧义。2019 年,百度公司推出基于 BERT 的改进模型 ERNIE<sup>[6]</sup>,ERNIE 是一

种针对中文文本的预训练语言模型,它通过对连续的实体词和短语进行掩码训练,使其能够学习到更好的语义知识以提高性能。李天昊等人<sup>[25]</sup>利用 ERNIE 和注意力机制相结合的方式提高中文文本特征的表示。但是对于词的歧义问题仍然没有较好地解决。

基于预训练语言模型的方法实现了更有竞争力的性能,但仍需要解决两个问题。第一个问题是虽然后期的预训练语言模型能够将输入序列转化为动态向量表示,从而一定程度上缓解一词多义问题,但还是无法满足对于句子中词义的深入理解,而这一步骤又是中文关系抽取的基础。第二个问题是对于句子中的目标实体,其所包含的语义信息还相对匮乏,在某些情况下,实体的表示还是会产生歧义,最终影响关系的抽取。对于这两点问题,文中模型都提出了具体的解决办法。

对于第一个问题,目前的 ERNIE 预训练模型可以较好地解决词向量的问题,然而仅仅依靠 ERNIE 模型是不够的,还应该针对下游任务进一步利用注意力机制解决词义的单薄问题,通过将动态向量中重要的词以及词之间的相互依赖性利用权重凸显出来。对于第二个问题,目标实体如果想更好地避免歧义问题,还需要外部知识的助力,外部知识可以在实体存在歧义时给予一定的辅助,使得实体即使在上下文信息很少的句子中,也可以尽可能地理解并补充词义。通过这两个关键问题的解决,最终丰富了句子的语义表示,提高了中文关系抽取的性能。

## 2 多层次语义感知的中文关系抽取模型

该文提出的模型框架如图 2 所示,主要分为三个层次,分别是:编码层、多层次语义感知层、分类层。

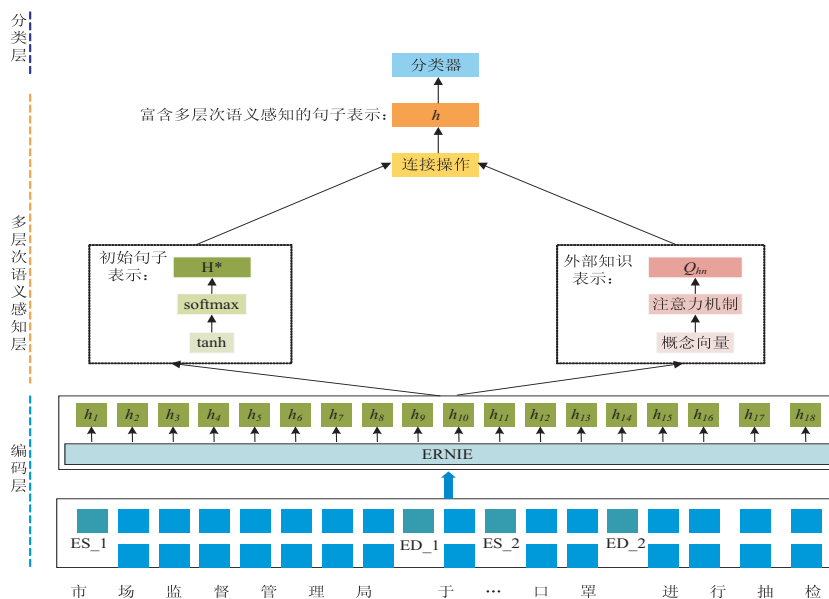


图2 模型框架



在编码层,首先将需要判断实体关系的句子进行实体标记,然后将其放入到 ERNIE 预训练模型中得到输出的动态词向量。这样词向量在编码阶段就具有了实体感知能力和上下文捕捉能力。在多层次语义感知层,首先利用自注意力机制计算句子中的其他词对目标实体对的影响,也就是其他词与目标实体对的交互权重,这样句子中的语义就能够被目标实体对所吸收。然后为了进一步避免目标实体的词义歧义,利用 HowNet<sup>[26]</sup> 知识库作为外部知识对目标实体表示进行补充,这样进一步丰富了目标实体所在的句子表示。在分类层中,将含有多层次语义的句子表示放入到分类器中,计算目标实体对所属的关系类型。

## 2.1 编码层

为了使预训练模型能够精确地识别出需要判断关系的实体对,首先对句子中需要判断关系的实体对进行标记,如句子“市场监督管理局于 2022 年 5 月 20 日对当事人所销售口罩进行抽检”,这里想知道实体“市场监督管理局”和“口罩”的关系,那么在对该句子进行词向量变换之前,就需要对目标实体对(市场监督管理局,口罩)进行标记,标记结果为:“ES\_1 市场监督管理局 ED\_1 于 2022 年 5 月 20 日对当事人所销售 ES\_2 口罩 ED\_2 进行抽检。”其中“ES\_1”和“ED\_1”分别是第一个目标实体“市场监督管理局”的左右两边的标记,“ES\_2”和“ED\_2”分别是第二个目标实体“口罩”的左右两边的标记,将需要判断实体对关系的句子标记完毕后,就可以进行下一步向量化处理。假设句子经过标记后由  $n$  个字符组成,其中  $w_i$  表示句子中的第  $i$  个字符,将句子放入到 ERNIE 预训练模型中得到如下的向量化表示:

$$H = [h_1, h_2, \dots, h_n] = \text{ERNIE}([w_1, w_2, \dots, w_n]) \quad (1)$$

其中,  $H \in \mathbb{R}^{n \times d^h}$ ,  $d^h$  是隐藏状态  $h_i$  的维度。

## 2.2 多层次语义感知层

对于多层次语义感知层,是将实体相关的多粒度的语义感知融入到句子表示中,多粒度具体是以词、句子以及概念为建模对象。

经 ERNIE 预训练模型输出的动态词向量在一定程度上确实包含了实体对所在句子的上下文信息,但是这些语义对于更精确地预测实体对关系是不够的,实体所在的句子表示信息还应该涵盖更为丰富的语义信息,所在多层次语义感知层中通过注意力机制和外部知识共同丰富句子语义信息。

句子中的每一个字符对目标实体的贡献度是不同的,也就是说有些字符更有利于判断实体对的关系,比如在“市场监督管理局于 2022 年 5 月 20 日对当事人所销售口罩进行抽检”句子中,“抽检”相对于“销售”

对判断实体对(市场监督管理局,口罩)关系的影响是不同的,那么对实体对影响大的字符应该被赋予更大的权重,反之则赋予更小的权重,那么利用注意力机制来计算实体所在句子的初始表示:

$$S = \tanh(H) \quad (2)$$

$$\alpha = \text{softmax}(b^T S) \quad (3)$$

$$H^* = H\alpha^T \quad (4)$$

其中,  $b \in \mathbb{R}^{d^h}$  是可训练参数,  $\alpha \in \mathbb{R}^n$  是权重参数。

实体所在句子的句向量初始表示的具体计算过程如图 3 所示。

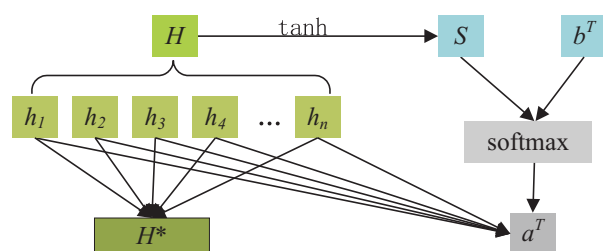


图 3 计算实体所在句子的初始表示的过程

在模型训练过程中,首先对句子的向量化表示进行  $\tanh$  激活处理,从而缓解模型训练中的梯度消失问题;然后将此过程的输出  $S$  与可训练参数  $b$  转置相乘,计算每个字符对于句子中实体的权重  $\alpha$  大小;根据所得权重值更新句子的向量化表示,将基于新的句向量的模型预测结果与数据实际标签进行比较,计算损失函数值,通过将误差以反向传播的方式计算参数梯度,通过梯度下降法更新模型可训练参数,这样就通过注意力机制将字符级特征融入到句子级特征中。

考虑到中文词义的特殊性和复杂性,即同一个词在不同的语义下有着不同的含义,如图 1 中的例子,“杜鹃”这个词在不同的语境下有着不同的词义,所以需要根据不同的语境区分“杜鹃花”或者“杜鹃鸟”,但是之前的模型由于没有先验知识所以无法正确地区分出这个词义,而文中模型引入 HowNet 这个外部知识库,旨在进一步提高模型应对词义歧义的能力。对于 HowNet 知识库,官方提供了 API 接口,这里可以直接调用使用,比如用 API 接口去搜索“口罩”相关的词义,会得到如下“[eat|吃, protect|保护, MakeSound|发声, tool|用具, mouth|口, PutOn|穿戴, viscera|脏, part|部件, AnimalHuman|动物]”与口罩相关的外部知识,这些知识对口罩有着较好的补充丰富的作用,那么假设实体通过 HowNet 遍历得到  $k$  个相关的概念,这  $k$  个概念经过 ERNIE 编码得到  $k$  个动态词向量。同样地,利用注意力机制增强实体所对应的外部知识的语义表示:

$$\beta_j = \frac{\exp(c_k)}{\sum_k \exp(c_k)} \quad (5)$$

$$O_{hn} = \sum_k \beta_j h_i \quad (6)$$

最后将得到的外部知识表示和句子表示相连接得到富含多层次语义感知的句子表示:

$$h = [H^*; O_{hn}] \quad (7)$$

其中, [ ; ] 是连接操作符。

### 2.3 分类层

经过多层次语义感知的作用,得到了实体对所在句子的最终表示,将预定义的中文关系的抽取任务作为二分类任务分别进行处理,利用 sigmoid 函数得到关系  $r$  在关系集合中的概率:

$$p_r = \text{sigmoid}(W_r h + b_r) \quad (8)$$

其中,  $p_r \in \mathbb{R}^{|\mathcal{R}|}$ ,  $W$  和  $b$  是可训练参数。

最后利用二元交叉熵来定义损失函数并且在模型训练中通过 Adam 优化器<sup>[27]</sup>来调节损失函数。损失函数定义如下:

$$\mathcal{L}_{\text{loss}} = - \sum_{r \in \mathcal{R}} (y_r \log(p_r) + (1 - y_r) \log(1 - p_r)) \quad (9)$$

其中,  $y_r \in \{0, 1\}$  表示关系标签  $r$  的真实值。

## 3 实验

### 3.1 数据集

所提模型选取在多个论文中都使用的数据集 SanWen<sup>[28]</sup>和 FinRE<sup>[29]</sup>进行实验。

SanWen 是人工标注的中文数据集,包含 837 篇文档共计 21 240 句话,随机选取其中的 81.1% 作为训练集,8.4% 作为验证集,剩余 10.5% 作为测试集。实体之间有九种关系,包括定位、接近、部分-整体、一般-特殊关系、家庭、社会、所有权、使用、创造。

FinRE 由 2 647 条新浪财经新闻的 18 702 个实例组成。其中的 13 486 条作为训练集,1 489 条作为验证集,剩余的 3 727 条作为测试集。包含了 44 种实体之间的关系,包括竞争、合作、减持等金融领域的特定关系。

### 3.2 评估

对于模型评估,采用此类模型常用的性能指标: F1 值。F1 值的应用主要为了调和精确率和召回率,使其尽可能地兼顾精确率和召回率。三个性能计算方式如下:

$$\text{精确率} = \frac{TP}{TP + FP} \quad (10)$$

$$\text{召回率} = \frac{TP}{TP + FN} \quad (11)$$

$$F1 = \frac{2 * \text{精确率} * \text{召回率}}{\text{精确率} + \text{召回率}} \quad (12)$$

其中, TP 表示模型预测结果为正实例,实际也为正实

例; FP 表示模型预测结果为正实例,但是实际为负实例; FN 表示模型预测结果为负实例,但是实际为正实例。

### 3.3 实验设置

表 1 给出了模型在最好性能下的参数设置。模型在 ERNIE 编码器下进行,所以输出动态向量的维度为 768,这个维度和 BERT 是一致的,在不断调整参数过程中,丢弃率设置为 0.5,初始学习率设置为 1.0,最优学习率为 5e-5,进行 50 轮次可以达到最优训练模型。

表 1 超参数设置

超参数	值
Batch Size	10
Encoder Hidden Size	768
Dropout	0.5
Optimizer	Adam
Learning Rate	0.000 5
Epoch	50

### 3.4 比较模型

为了更好地研究所提模型,在相同数据集下与其他模型进行了比较,比较的模型包括基于传统神经网络的模型和基于预训练语言的模型。

(1) 基于传统神经网络的模型: PCNN + Att<sup>[10]</sup>, Lattice + LSTM<sup>[13]</sup>, Lattice + MG<sup>[14]</sup>, PCNN<sup>[30]</sup>, MGRSA<sup>[31]</sup>。

(2) 基于预训练语言的模型: ERNIE<sup>[6]</sup>, PURE<sup>[18]</sup>, IERE<sup>[20]</sup>, PRM<sup>[24]</sup>, OPT-FLAT<sup>[32]</sup>。

### 3.5 实验结果

#### 3.5.1 SanWen 数据集实验结果

所提模型在 SanWen 数据集上的结果如表 2 所示,可以发现所提模型优于现有模型,比目前最好模型的 F1 值高 0.62%。这显示出对实体所在句子的语义进行处理至关重要,并且仅仅依靠句子本身的语义无法得到良好的性能,必须将外部知识作为补充,因为当模型对产生歧义的词义模糊不清时,外部知识就可以很好地起到应有的效果,再者 ERNIE 预训练模型对实体词的理解起到了丰富作用。

表 2 SanWen 数据集上的实验结果(带有 \* 的表示引用文献[24]中的结果,其他均为原论文中提及的结果)

模型	F1/%
PCNN+Att * <sup>[10]</sup>	60.55
PCNN * <sup>[30]</sup>	61.23
ERNIE * <sup>[6]</sup>	63.25
Lattice+LSTM * <sup>[13]</sup>	63.88

续表 2

模型	F1/%
IERE * <sup>[20]</sup>	63.99
PURE * <sup>[18]</sup>	64.70
Lattice+MG <sup>[14]</sup>	65.61
MGRSA <sup>[31]</sup>	67.12
PRM <sup>[24]</sup>	67.72
OPT-FLAT <sup>[32]</sup>	68.35
Our Model	68.97

### 3.5.2 FinRE 数据集实验结果

所提模型在 FinRE 数据集上的结果如表 3 所示,可以发现所提模型同样优于现有模型,比目前最好模型的 F1 值高 0.09%。同时可以看出所提模型在 FinRE 数据集上的优势要低于在 SanWen 数据集上的优势,经过分析主要是两个数据集在关系类型数量上是不同的, SanWen 数据集包含了 9 种关系类型,而 FinRE 数据集包含了 44 种关系,这一差别造成了模型抽取难度的增大,对于相近的关系类似给模型提出更高的要求,而所提模型的多层次语义感知中的外部知识是将实体抽象为概念,间接地也放大了实体对所属的关系类型。这也给以后改进模型指明了方向,在引用外部知识时尽可能精确化。

表 3 FinRE 数据集上的实验结果(带有\*的表示引用文献[24]中的结果,其他均为原论文中提及的结果)

模型	F1/%
PCNN * <sup>[30]</sup>	45.51
PCNN+Att * <sup>[10]</sup>	46.13
Lattice+LSTM * <sup>[13]</sup>	47.41
ERNIE * <sup>[6]</sup>	47.45
IERE * <sup>[20]</sup>	49.09
Lattice+MG <sup>[14]</sup>	49.26
PURE * <sup>[18]</sup>	49.61
OPT-FLAT <sup>[32]</sup>	50.60
MGRSA <sup>[31]</sup>	52.61
PRM <sup>[24]</sup>	52.97
Our Model	53.06

### 3.6 消融实验

通过表 2 和表 3 可以看出,所提模型在与其他模型的对比中有着较强的竞争力。为了进一步探究模型中三个部分所起到的作用,在 SanWen 数据集上进行消融实验。经过上述理论分析,模型中重要的模块分别是:ERNIE 编码器、句子表示模块以及外部知识表示模块。为了明显地看出三大模块在模型中的作用,

依次关闭其中的某个模块。首先,对于 ERNIE 编码器,使用 BERT-base 中文版本对 ERNIE 进行替换,结果发现模型下降了 1.04%,说明 ERNIE 编码器有助于模型从基础上缓解语义不丰富的问题。当去掉句子表示模块时,模型直接大幅度下降 1.86%,说明模型中的句子表示模块对模型性能的提升起到关键作用,句子表示模块通过利用注意力机制将句子中对实体词有用的字符进一步增强,实体词的语义得到了有针对性的丰富。最后去掉外部知识表示模块,发现模型性能有着小幅度的降低,表明外部知识有利于缓解词的歧义,但是也带来了一定的噪声,所以对模型的提升效果不是那么明显。

表 4 消融实验

模型	F1/%
Our Model	68.97
- ERNIE 编码器	67.93
-句子表示模块	67.11
-外部知识表示	67.89

通过表 4 可以看出,模型中的三个模块对模型整体的性能提升都有一定的作用,三者的有机结合最后使模型的性能与之前的模型相比有着较大的提高。

## 4 结束语

该文提出了基于多层次语义感知的中文关系抽取模型,通过实验表明该模型在处理中文关系抽取中可以取得较好的性能。在未来的研究中,将进一步思考外部知识的引入,在尽可能小地引入噪声的前提下使外部知识更有效地提高中文关系抽取模型的性能。

### 参考文献:

- [1] ZHANG J,HAO K,TANG X,et al. A multi-feature fusion model for Chinese relation extraction with entity sense[J]. Knowledge-Based Systems,2020,206:106348.
- [2] TANG H,CAO Y,ZHANG Z,et al. HIN:hierarchical inference network for document-level relation extraction[C]//Proceedings of the 24th Pacific-Asia conference. Singapore: PAKDD,2020:197-209.
- [3] 杨超男,彭敦陆. 融合 BSRU 和胶囊网络的文档级实体关系抽取模型[J]. 小型微型计算机系统,2022,43(5):964-968.
- [4] 吴 婷,孔 芳. 基于图注意力卷积神经网络的文档级关系抽取[J]. 中文信息学报,2021,35(10):73-80.
- [5] 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. Dublin:ACL,2022:1672-1681.
- [6] SUN Y,WANG S,LI Y,et al. Ernie 2.0:a continual pre-

- training framework for language understanding [C]//Proceedings of the AAAI conference on artificial intelligence. New York: AAAI, 2020: 8968–8975.
- [7] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//31st annual conference on neural information processing systems. Long Beach: NIPS, 2017: 5998–6008.
- [8] LIU C Y, SUN W B, CHAO W H, et al. Convolution neural network for relation extraction [C]//Advanced data mining and applications: 9th international conference. Hangzhou: ADMA, 2013: 231–242.
- [9] ZENG D, LIU K, LAI S, et al. Relation classification via convolutional deep neural network [C]//Proceedings of COLING 2014, the 25th international conference on computational linguistics. Dublin: COLING, 2014: 2335–2344.
- [10] SHEN Y, HUANG X J. Attention-based convolutional neural network for semantic relation extraction [C]//Proceedings of COLING 2016, the 26th international conference on computational linguistics. Osaka: COLING, 2016: 2526–2536.
- [11] 梁艳春, 房爱莲. 基于多通道卷积神经网络的中文文本关系抽取[J]. 华东师范大学学报: 自然科学版, 2021(3): 96–104.
- [12] ZHANG D, WANG D. Relation classification via recurrent neural network [J]. arXiv: 1508.01006, 2015.
- [13] ZHANG Y, YANG J. Chinese NER using lattice LSTM [C]//Proceedings of the 56th annual meeting of the association for computational linguistics. Melbourne: ACL, 2018: 1554–1564.
- [14] LI Z, DING N, LIU Z, et al. Chinese relation extraction with multi-grained information and external linguistic knowledge [C]//Proceedings of the 57th annual meeting of the association for computational linguistics. Firenze: ACL, 2019: 4377–4386.
- [15] KONG B, LIU S, WEI F, et al. Chinese relation extraction using extend softword [J]. IEEE Access, 2021, 9: 110299–110308.
- [16] 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. Minneapolis: NAACL, 2019: 4171–4186.
- [17] EBERTS M, ULGES A. Span-based joint entity and relation extraction with transformer pre-training [C]//European conference on artificial intelligence. Santiago: ECAI, 2020.
- [18] ZHONG Z, CHEN D. A frustratingly easy approach for entity and relation extraction [C]//Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics. Mexico City: NAACL, 2021: 50–61.
- [19] LAN Z, CHEN M, GOODMAN S, et al. ALBERT: A lite bert for self-supervised learning of language representations [C]//International conference on learning representations. Addis Ababa: ICLR, 2020.
- [20] ZHOU W, CHEN M. An improved baseline for sentence-level relation extraction [J]. arXiv: 2102.01373, 2021.
- [21] LIU Y, OTT M, GOYAL N, et al. Roberta: a robustly optimized bert pretraining approach [J]. arXiv: 1907.11692, 2019.
- [22] CUI Y, CHE W, LIU T, et al. Revisiting pre-trained models for chinese natural language processing [C]//Findings of the association for computational linguistics. Punta Cana: EMNLP, 2020: 657–660.
- [23] YANG J, JI B, LI S, et al. Dynamic multi-view fusion mechanism for chinese relation extraction [J]. arXiv: 2303.05082, 2023.
- [24] ZHAO Q, GAO T, GUO N. A novel chinese relation extraction method using polysemy rethinking mechanism [J]. Applied Intelligence, 2023, 53(7): 7665–7676.
- [25] 李天昊, 霍其润, 闫跃, 等. 融合 ERNIE 和注意力机制的中文关系抽取模型 [J]. 小型微型计算机系统, 2022, 43(6): 1226–1231.
- [26] DONG Z, DONG Q. HowNet—a hybrid language and knowledge resource [C]//International conference on natural language processing and knowledge engineering. San Francisco: IEEE, 2003: 820–824.
- [27] KINGMA D P, BA J. Adam: a method for stochastic optimization [J]. arXiv: 1412.6980, 2014.
- [28] XU J, WEN J, SUN X, et al. A discourse-level named entity recognition and relation extraction dataset for chinese literature text [J]. arXiv: 1711.07010, 2017.
- [29] NIU Y, XIE R, LIU Z, et al. Improved word representation learning with sememes [C]//Proceedings of the 55th annual meeting of the association for computational linguistics. Vancouver: ACL, 2017: 2049–2058.
- [30] ZENG D, LIU K, CHEN Y, et al. Distant supervision for relation extraction via piecewise convolutional neural networks [C]//Proceedings of the 2015 conference on empirical methods in natural language processing. Lisbon: EMNLP, 2015: 1753–1762.
- [31] ZHONG Z. Chinese entity relation extraction based on multi-level gated recurrent mechanism and self-attention [C]//2021 2nd international conference on artificial intelligence and information systems. Chongqing: ICAIS, 2021: 1–7.
- [32] ZENG X, ZHONG J, WANG C, et al. Chinese relation extraction with flat-lattice encoding and pretrain-transfer strategy [C]//Knowledge science, engineering and management: 14th international conference. Tokyo: KSEM, 2021: 30–40.