计算机科学
COMPUTER SCIENCE

基于注意力机制和提示学习联合训练的上下位关系识别研究

白宇, 王新哲

引用本文

白宇, 王新哲. 基于注意力机制和提示学习联合训练的上下位关系识别研究[J]. 计算机科学, 2024, 51(6A): 230700226-5.

BAI Yu, WANG Xinzhe. Study on Hypernymy Recognition Based on Combined Training of Attention Mechanism and Prompt Learning [J]. Computer Science, 2024, 51(6A): 230700226-5.

相似文章推荐（请使用火狐或 IE 浏览器查看文章）

Similar articles recommended (Please use Firefox or IE to view the article)

基于SAMNV3的滚动轴承智能故障诊断方法

Intelligent Fault Diagnosis Method for Rolling Bearing Based on SAMNV3

计算机科学, 2024, 51(6A): 230700167-6. https://doi.org/10.11896/jsjkx.230700167

基于LSTM和注意力机制的远程会诊需求预测

Forecasting Teleconsultation Demand Based on LSTM and Attention Mechanism

计算机科学, 2024, 51(6A): 230800119-7. https://doi.org/10.11896/jsjkx.230800119

融入类别标签和主题信息的用户兴趣识别方法

User Interest Recognition Method Incorporating Category Labels and Topic Information

计算机科学, 2024, 51(6A): 230500169-8. https://doi.org/10.11896/jsjkx.230500169

基于注意力的多尺度蒸馏异常检测

Attention-based Multi-scale Distillation Anomaly Detection

计算机科学, 2024, 51(6A): 230300223-11. https://doi.org/10.11896/jsjkx.230300223

基于改进Deeplabv3+算法的滚珠丝杠驱动表面点蚀缺陷检测

Detection of Pitting Defects on the Surface of Ball Screw Drive Based on Improved Deeplabv3+ Algorithm

计算机科学, 2024, 51(6A): 240200058-6. https://doi.org/10.11896/jsjkx.240200058

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230700226

基于注意力机制和提示学习联合训练的上下位关系识别研究

白　宇　　王新哲
沈阳航空航天大学人机智能研究中心　沈阳 110136
(wangxinzhed@163.com)

摘　要　专利术语间的上下位关系是一种重要的语义关系,专利文本中术语间的上下位关系识别在专利检索、查询扩展、知识图谱构建等多个领域发挥着重要作用。然而,专利文本领域的多样性、语言表述的复杂性使得术语间的上下位关系识别仍然面临许多挑战。文中提出一种融合提示学习和注意力机制的术语上下位关系识别方法,该方法基于远程监督框架,将术语之间的最短依存路径作为辅助特征融入提示模板,使用图神经网络将术语间的共现信息融入提示学习和注意力机制联合训练过程。在专利文本测试数据集上的实验结果表明,所提方法的 AUC 值、F1 值达到 94.94%和 89.33%,相较于 PARE 模型分别提升了 3.82%和 3.17%。该方法有效地去除了使用远程监督方法标注的数据集的噪声,避免了掩码语言模型的训练目标和下游任务的不匹配问题,充分利用了预训练语言模型中存在的语言知识信息。

关键词:术语关系识别;远程监督;提示学习;注意力机制;上下位关系
中图分类号　TP391

Study on Hypernymy Recognition Based on Combined Training of Attention Mechanism and Prompt Learning

BAI Yu and WANG Xinzhe
Research Center for HCI,Shenyang Aerospace University,Shenyang 110136,China

Abstract　The hypernymy between patent terms is an important semantic relationship. The identification of hypernymy between terms in patent text plays an important role in patent retrieval,query expansion,knowledge graph construction and other fields. However,due to the diversity of patent field and the complexity of language expression,the task of identifying the hypernymy between terms still faces many challenges. This paper proposes a method to recognize the hypernymy of terms by integrating prompt learning and attention mechanism. This method is based on the distantly supervised framework,and uses the shortest dependent path between terms as an auxiliary feature to integrate into the prompt template. Graph neural network is used to integrate the common information between terms into the joint training process of prompt learning and attention mechanism. Experimental results on the patent text test dataset show that the AUC and f1 value of our method reache 94.94% and 89.33%,respectively,which are 3.82% and 3.17% higher than the PARE model. This method effectively removes the noise of the dataset annotated using distantly supervised methods,avoids the mismatch problem between the training target of the masked language model and downstream tasks,and fully utilizes the language knowledge information existing in the pre-trained language model.

Keywords　Term relationship recognition,Distant supervision,Prompt learning,Attention mechanism,Hypernymy

1　引言

据研究表明,专利是世界上最大的技术信息来源,包含了世界上 90%～95%的科技信息[1]。专利中通常包含大量的技术术语和概念,这些术语和概念之间的上下位关系对于理解专利的技术含义和技术进展非常重要。上下位关系作为自然语言处理领域中的核心语义关系,对其准确预测有助于挖掘专利术语之间的内在层次结构,帮助人们更好地理解和组织知识。因此,上下位关系的准确预测在自然语言处理领域中具有重要的意义和价值。根据 WordNet[2] 的定义,给定两个词语 x 和 y,如果句子"x 是一种|类|…|y"是可以接受的,则称 y 是 x 的上位词,x 是 y 的下位词。例如,在句子"本外观设计产品用于汽车衡、轨道衡、平台秤等各种称重的衡器"中,"汽车衡"是一种"衡器"是可以接受的,所以"汽车衡"是"衡器"的下位词,"衡器"是"汽车衡"的上位词。

深度学习中的关系抽取任务主要分为有监督和远程监督两类[3],但是由于有监督方法十分依赖人工标注的数据,而人类注释者几乎不可能注释训练数据集以满足各行各业的需求,因此远程监督的思想被提出,即如果两个实体在知识库中存在某种关系,那么包含这两个实体的任何句子均能表示出这种关系[4]。由于目前没有中文专利领域的上下位关系数据集,所以本文基于远程监督的思想构建了中文专利领域的上下位关系数据集,但因为远程监督假设是一个强假设,所以这种方法会带来一些错误标记的噪声,因此本文使用注意力机

基金项目:辽宁省应用基础研究计划(2022JH2/101300248)
This work was supported by the Liaoning Province Applied Basic Research Program(2022JH2/101300248).
通信作者:王新哲(wangxinzhed@163.com)

制提取句子中的特征，降低噪声样本对模型的影响。由于关系向量通常只包含了关系本身的表示，而没有考虑术语之间的语义信息和上下文关系，导致关系表示缺乏完整的语义信息，因此引入图神经网络将术语向量与关系向量进行拼接，因为共现的术语之间通常存在某种语义关系，所以将术语间的共现信息组成图神经网络，将需要识别关系的术语对的图神经网络向量拼接在关系向量后。同时加入提示学习方法，将术语之间的最短依存路径作为辅助特征添加进提示模板，充分利用预训练语言模型中存在的大量知识。

本文工作主要包括：

(1)通过使用注意力机制和提示学习方法联合训练来优化上下位关系的识别效果。

(2)利用远程监督的思想，构建了中文专利领域的上下位关系数据集，使用融合图神经网络的注意力机制对数据集进行去噪，并将最短依存路径作为辅助特征融入提示学习模板中。

(3)实验结果表明，文中方法在中文专利数据集上达到了良好的效果，AUC 值和 F1 值分别达到了 94.94% 和 89.33%。

2 相关工作

针对有监督的训练过程中需要大量的标注数据的问题，Mintz 等[4]在 2009 年提出远程监督假设帮助解决了此问题，但由于缺少人工参与，所以数据集中往往存在着错误标签的问题。因此后续工作都致力于如何降低噪声对模型的影响。Zeng 等[5]将远程监督看作多实例学习问题，使用 CNN 模型自动学习文本特征。模型选择每个包中置信度最高的一个句子进行后续训练，这样做的好处在于这一个句子有很大的可能性不是噪声，但缺点是许多数据未被利用。针对上述方法的缺点，Jiang 等[6]通过最大池化层将包中所有句子的特征融合，以池化后的输出表示该包的特征。Lin 等[7]首次将注意力机制引入远程监督关系抽取算法，用句子级注意力代替多实例学习，动态降低噪声实例的权重。Qin 等[8]通过对抗生成网络在训练开始前过滤噪声。Yuan 等[9]采用句子级选择注意力机制减少噪声和不匹配句子的影响，同时利用关系间的相关性提高注意权重的质量。Chen 等[10]提出了一种新的对比实例学习框架，将属于同一关系的实例全部放在一个包中进行训练，得到该关系的词向量表示，使包含相同关系三元组的实例在隐藏的语义空间中尽可能地接近，而包含不同关系三元组的实例尽可能地远离。Li 等[11]提出了分层对比学习框架，整合了全局结构信息和局部细粒度信息交互，减少了噪声数据的影响。

3 基于注意力机制和提示学习联合训练的关系识别模型

为了过滤数据集中的噪声和进一步提高上下位词识别的准确性，文中采用了一种基于注意力机制和提示学习的联合训练的方法，这两个模型共享权重，构成了如图 1 所示的模型。左侧部分是基于 BERT＋注意力机制的术语上下位关系识别模型，它使用注意力机制来自动确定句子中哪些部分对两个术语之间的关系识别影响最大。右侧部分是基于 BERT＋提示学习的术语上下位关系识别模型，它使用提示学习的方法来进一步增强模型的鲁棒性和泛化能力。联合训练模型框架如图 1 所示。

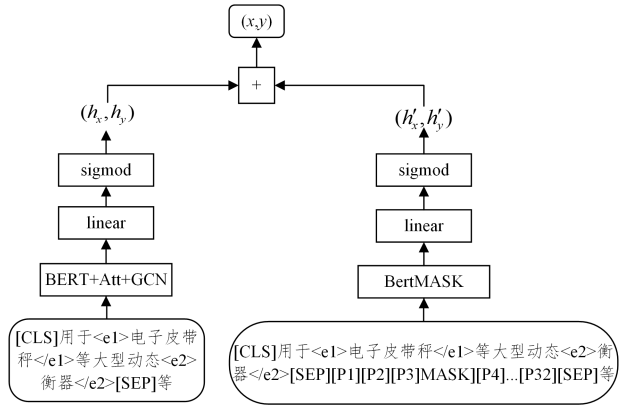

图 1 联合训练模型框架

Fig. 1 Framework of joint training model

模型首先将包含同一术语对的句子构建为一个段落，段落以[CLS]标记开始，然后由[SEP]分隔每个句子，并用[PAD]填充所有剩余的标记，并在段落中提到的头部实体和尾部实体的开始和结束处分别添加 4 个特殊的标记〈e1〉、〈/e1〉和〈e2〉、〈/e2〉。例如原输入为“用于电子皮带秤等大型动态衡器”经过处理后变成“[CLS]用于〈e1〉电子皮带秤〈/e1〉等大型动态〈e2〉衡器〈/e2〉[SEP]”。然后本文通过连接包含术语对 e1 和 e2 的所有句子来构造一个段落 P(e1,e2)，这样可以通过 BERT 对整个段落进行编码，段落中的每个 token 都能够与来自段落中其他句子的其他 token 交换信息，使词嵌入表示含有更多的信息。

在提示学习模块中，我们在上述序列的基础上增加了对应的提示模板，提示模板包括硬提示模板和软提示模板，其中硬提示就是人为定义的固定模板，例如在原有的句子基础上套用硬提示模板“e1 和 e2[MASK]上下位关系吗”，“e1”和“e2”代表上下位词对。所以上述序列被扩展为“[CLS]用于〈e1〉电子皮带秤〈/e1〉等大型动态〈e2〉衡器〈/e2〉[SEP]等[SEP]电子皮带秤和衡器[MASK]上下位关系吗”。软模板提示学习的方法是将提示语句的生成作为一个任务进行学习，相当于把提示语句的生成从人类进行尝试换成机器自己进行学习。然后，为了充分利用术语之间的语义信息，文中利用依存句法分析得出术语对之间的最短依存路径，拼接在提示学习模板中，将最短依存路径作为上下位关系识别的辅助特征，更好地捕捉它们之间的语义关系，提高上下位关系识别的效果。

在上述句子中，最短依存路径最后为“电子皮带秤等衡器”，所以本文把“等”拼接在上述序列后面，最后提示学习模块的输入为“[CLS]用于〈e1〉电子皮带秤〈/e1〉等大型动态〈e2〉衡器〈/e2〉[SEP][P1][P2][P3]MASK][P4]…[P32][SEP]等[SEP]”，其中 pi 为软提示模板的触发词，提示模板长度为 32。

3.1 基于注意力机制的上下位关系识别

为了解决使用远程监督思想自动标记的数据集中存在的噪声问题，文中模型引入了注意力机制。因为在每个段落中可能只有一部分单词对术语之间的关系识别有用，所以文中采用注意力机制对 BERT 输出的字符向量进行加权求和，它能识别段落中哪些部分对术语之间的关系识别有更大的影响，更加准确地找到对术语关系识别有用的信息，减少噪声对

关系识别结果的干扰。

将上述段落 E 输入到 BERT 中获得每个 token 的向量集 $\boldsymbol{H}_e$：

$$\boldsymbol{H}_e=Bert(E) \tag{1}$$

然后，随机初始化数据集中存在的关系为初始关系向量 $\boldsymbol{H}_r$，但由于关系向量通常只包含了关系本身的表示，而没有考虑术语之间的语义信息，因此本文根据专利数据中存在大量的术语的特点并且共现的术语之间通常存在某种语义关系，将数据集中共现过的术语组成图神经网络，将每个句子中需要判断关系的术语对的图神经网络向量拼接在初始关系向量后。本文选用 GCN 将共现关系图信息引入模型。在具有 l 层 GCN 结构的网络中，第($0<h\leqslant l$)层的输入是对应图的邻接矩阵 $\boldsymbol{A}$，以及 $h-1$ 层输出的节点特征的集合 $NF^{(h-1)}=\{f_1,f_2,\cdots,f_n\}$。GCN 层如式(2)、式(3)所示：

$$NF^h=\sigma(\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}NF^{(h-1)}\boldsymbol{W}^{(h-1)}) \tag{2}$$

$$\widetilde{\boldsymbol{D}}=\sum_i\widetilde{\boldsymbol{A}}_{ij},\widetilde{\boldsymbol{A}}=\boldsymbol{A}+I_N \tag{3}$$

其中，$\boldsymbol{W}^{(h-1)}$ 为模型中待训练的参数，矩阵 $\widetilde{\boldsymbol{A}}$ 为添加了自连接的图邻接矩阵，矩阵 $\widetilde{\boldsymbol{D}}$ 为矩阵 $\widetilde{\boldsymbol{A}}$ 的度矩阵，$\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}$ 为归一化之后的邻接矩阵，σ 为激活函数。

图神经网络层最终输出结果如式(4)所示：

$$\boldsymbol{H}_G=GCN(\boldsymbol{Node}_f,\widetilde{A}) \tag{4}$$

其中，$\boldsymbol{Node}_f$ 为输入到模型的每个节点的向量。

将关系向量 $\boldsymbol{H}_r$ 和图神经网络中的术语节点的向量 $\boldsymbol{H}_G$ 顺序拼接为 $\boldsymbol{H}_r'$，利用注意力机制计算每个句子包向量 $\boldsymbol{H}_e$ 和 $\boldsymbol{H}_r'$的注意力分数，应用 softmax 函数，将 $\boldsymbol{H}_e$ 和 $\boldsymbol{H}_r'$向量的乘积归一化，得到注意力结果。注意力机制的计算式如式(5)、式(6)所示：

$$\boldsymbol{H}_r'=\{\boldsymbol{H}_r;\boldsymbol{H}_G\} \tag{5}$$

$$\alpha=\mathrm{softmax}(\boldsymbol{H}_e\boldsymbol{H}_r'^T) \tag{6}$$

$$\boldsymbol{H}_o=\alpha\boldsymbol{H}_e \tag{7}$$

然后将得到的注意力结果经过一个线性层变成一个一维向量，得到每个关系的注意力分数(h_x,h_y)：

$$(h_x,h_y)=w\boldsymbol{H}_o+b \tag{8}$$

其中 w,b 为训练参数。

3.2　基于提示学习的上下位关系识别

预训练语言模型训练在应用到下游任务时，会发生训练与微调时的不匹配问题，因此本文引入提示学习，通过向输入增加“提示信息”，使输入和输出更适合原始语言模型，提升专利术语间上下位关系识别时的准确性。基于直接建模文本概率的语言模型，直接计算模板[MASK]处候选词的概率 $S_p(l|x)$：

$$S_p(l|x)=M(v(l)|P(x)) \tag{9}$$

其中，$v(l)$表示经过“类别－>类标签词”映射得到的词语，$P(x)$表示将原始输入和模板进行拼接之后得到的输入。

首先选择一个句子，选取符号“$\langle P\{i\}\rangle$”作为触发词，具体填充长度可控，之后为模板添加一个[MASK]标记作为最后的答案的标记，将“是，否”作为 MASK 的可选项，将输入语句与提示模板拼接，组成输入序列 E'。训练过程中，逐个遍历当前样本的各个触发词，从词汇表中找出 k 个候选触发词，然后从 k 个候选触发词中找出对模型提升作用最大的那一个，并用它替换当前位置的原有触发词。

之后在 BERT 的词表中增加对应数量的标签$\langle P\{i\}\rangle$，并将序列 E'输入到 BERT 模型中得到每个 token 的向量集。

$$\boldsymbol{H}_e'=Bert(E') \tag{10}$$

得到 $\boldsymbol{H}_e'$之后，本文将[MASK]标签所对应的 token 输出拿出来，经过线性变换，将其维度改变为 BERT 词表大小，将词表中字符“是”和“否”在该向量中所对应的概率取出，经过 sigmod 函数得到该部分的最终输出概率(h_x',h_y')，将其作为判断两个术语是否为上下位词的判断依据之一。

最终输出结果如式(11)所示：

$$(x,y)=\sigma(h_x,h_y)+\sigma(h_x',h_y') \tag{11}$$

其中，(h_x,h_y)为注意力部分的输出概率，(h_x',h_y')为提示学习部分的输出概率。

最后，损失函数使用二值交叉熵损失方法：

$$loss=-w_i[y_i\log x+(1-y_i)\log y] \tag{12}$$

其中，输入的 x 和 y 为预测值，范围在 0～1 之间；y_i 为真实标签，值为 0 或者 1。

4　实验与分析

4.1　数据准备

针对目前没有公开的中文专利上下位关系数据集的问题，因此本文使用了远程监督的思想，即默认如果两个实体在知识库中存在某种关系，则包含这两个实体的非结构化句子均能表示出这种关系。首先在京东网站上爬取存在上下位关系的术语对，然后手动删除不是上下位关系的词对或者没有用的词对，例如(游戏设备，游戏配件)，词对并没有明确指出游戏配件的种类，因此认为它是无用的。然后利用这些词对从中文专利 solr 库中爬取上下位词对共同出现的句子，构建出上下位关系数据集中的正例。

对于负例的构建，文中利用术语识别，识别出包含正例的句子中的所有术语，然后随机选取一个术语，和术语对中的上位词组成非上下位关系。为了保证随机选取术语时，不会选到下位词，文中对数据集进行了一个筛选操作，即如果随机选取的术语是所给数据集中包含的该上位词的其他下位词时，则随机再从文本中选取一个术语。

对于测试集，为了保证结果的准确性，文中采取人工标注的方法，共标注了 2741 条数据作为测试集。

4.2　比较方法

本文选取了 3 种最新的模型进行比较：

CIL：Chen 等[10]提出了一种新的对比实例学习框架，将初始多实例学习视为关系三重编码器，并对每个实例的负对进行约束。

HICLRE：Li 等[11]提出了一个分层对比学习框架，利用多粒度语义重构增强交互层信息，动态梯度对抗扰动学习三层更好的表示。

PARE：Rathore 等[12]探索了一种简单的基线方法，将实体对的所有句子连接成一个段落，并使用 BERT 联合编码。

4.3　评价标准

本文使用 AUC 和 F1 值作为评估指标，AUC 可以衡量此学习器对于正例和负例的分类能力，其他指标可以对实验结果做出更准确的评价。

$$AUC=\frac{\sum I(P_{正样本},P_{负样本})}{M\times N} \tag{13}$$

其中 $I(P_{正样本},P_{负样本})=\begin{cases}1, & P_{负样本}>P_{正样本}\\0.5, & P_{负样本}=P_{正样本}\\0, & P_{负样本}<P_{正样本}\end{cases}$

$$P=\frac{TP}{TP+FP} \tag{14}$$

$$R=\frac{TP}{TP+FN} \tag{15}$$

$$F1=\frac{2\times P\times R}{P+R} \tag{16}$$

4.4 实验参数设置

本文选择使用 BERT 预训练语言模型初始化词向量，初始词向量的维度为 512，隐藏层的维度为 768，采用 sigmod 函数作为激活函数，并使用 Adamw 优化算法学习模型参数，学习率设置为 1×10^{-5}，batchsize 为 2，损失函数使用了二值交叉熵损失方法。实验中采用早停的方法，如果 10 轮内损失没有持续下降就停止训练。

4.5 对比实验

从表 1 可以看出，本文模型的 AUC 和 F1 值分别比实验结果最好的 PARE 模型提升了 3.82%，3.17%。原因在于：本文模型更好地利用了预训练模型中的信息，通过图神经网络将术语的共现信息引入关系表示中，丰富了关系向量的表示，使模型能够更全面地理解关系。如表 1 所列，本文方法获得了最优结果。

表 1 本文模型与现有方法的性能对比

Table 1 Performance comparison between our model and existing methods

(%)

Algorithm	*AUC*	*F*1
CIL	81.84	81.57
PARE	91.12	86.16
HICLRE	91.93	83.59
Our Method	94.94	89.33

4.6 提示模板性能分析

为了探究不同提示对模型性能的影响，文中进行了一系列的实验，比较了 3 种人工创建的硬提示和软提示的实验效果。其中，硬提示分别为：(1) pet1=ent1[MASK]一种 ent0；(2) pet2=ent1+[MASK]+ent0+['的下位词']；(3) pet3=ent1+ent0+[MASK]+['上下位关系']，其中 ent0 和 ent1 为头部实体和尾部实体，[MASK]位置会随术语长度的不同而发生改变。结果如表 2 所列，通过分析实验数据，发现软提示对模型性能的提升效果最好，而在硬提示中，硬提示模板的效果与软提示相差较大，可以看出，硬提示的模板选取对模型的性能影响很大，需要认真考虑和选择。因此，本文在提示学习模块选择软提示作为提示模板。

表 2 提示性能实验结果对比

Table 2 Comparison of prompt performance experiment results

(%)

Algorithm	*AUC*	*F*1
pet1	88.91	85.93
pet2	89.52	81.86
pet3	80.76	78.30
ptuning	94.94	89.33

4.7 消融实验

为了验证本文模型中各模块的有效性，本文进行了消融实验，其中，“w/o”表示删除该部分模块，实验结果如表 3 所列。

表 3 模型各组件的消融实验结果对比

Table 3 Comparison of ablation experimental results of each component of the proposed model

(%)

Algorithm	*AUC*	*F*1
w/o att	90.42	85.25
w/o prompt	91.72	86.68
w/o sdp	93.15	87.33
w/o gcn	94.21	86.67
Our Method	94.94	89.33

对比实验结果发现，不添加注意力机制、提示学习、图神经网络和最短依存路径后，所有评价指标都有不同程度的下降。虽然去除图神经网络对模型性能的影响较小，但 AUC 值和 F1 值都有所下降，说明专利中术语间的共现信息对模型还是有积极作用的。不添加注意力机制对实验结果影响最大，说明采用远程监督方法构建的数据集噪声较多，注意力机制可以减少噪声的干扰，当去除注意力机制时，模型可能无法有效地过滤噪声，导致性能下降。不添加提示学习模块的影响次之，说明提示学习模块提供了与语义关系相关的信息，这些信息可以帮助模型更好地理解术语之间的语义关系。去掉提示学习模块后，模型无法准确地捕捉到术语之间的语义关系，导致性能下降。

结束语 本文采用远程监督的思想构造中文专利领域的上下位关系数据集，使用提示学习和注意力机制联合训练的方法对上下位关系识别任务进行了改进，使用注意力机制使模型更关注数据集中的正确特征，并利用图神经网络融入了专利文本中大量存在的术语之间的共现信息，同时通过向输入拼接提示模板，并将最短依存路径融入提示模板中，充分利用了预训练语言模型的知识。实验结果表明，所提方法在上下位关系识别任务上 AUC 和 F1 值得到了明显提高，从而提高了关系识别模型的性能。

未来的工作将进一步优化上下位关系识别模型。首先，将继续探索如何降低噪声对模型性能的影响，以提高识别效果的稳定性和可靠性。其次，将探索如何更好地利用提示信息来提高模型的性能和泛化能力。

参 考 文 献

[1] CHEN Y, HUANG Y Y, FANG J G. Paten Infomation Collection and Analysis[M]. Beijing: Tsinghua University Press, 2006.

[2] MILLER G A. WordNet: A lexical database for English [J]. Communications of the ACM, 1995, 38(11): 39-41.

[3] ZHANG S J, CHEN J X, WU X Y. A Novel Distant supervision Relation Extraction Approach Based on Sentence Bag Attention [J]. Computer Application and Software, 2022, 39(8): 193-203.

[4] MINTZ M, BILLS S, SNOW R, et al. Distant supervision for relation extraction without labeled data[C]// Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP. 2009: 2-7.

[5] 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. 2015：1753-1762.

[6] JIANG X T，WANG Q，LI P，et al. Relation Extraction with Multi-instance Multi-label Convolutional Neural Networks [C]// Proceedings of the 26th International Conference on Computational Linguistics：Technical Papers. 2016：1471-1480.

[7] LIN Y K，SHEN S Q，LIU Z Y，et al. Neural Relation Extraction with Selective Attention over Instances[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics(Volume 1：Long Papers). 2016：2124-2133.

[8] QIN P D，XU W R，WANG W Y. DSGAN：Generative Adversarial Training for Distant Supervision Relation Extraction [C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics(Volume 1：Long Papers). 2018：496-505.

[9] YUAN Y，LIU L，TANG S，et al. Cross-Relation Cross-Bag Attention for Distantly-Supervised Relation Extraction[C]/Proceedings of the Conference on Artificial Intelligence. Association for the Advancement of Artificial Intelligence(AAAI). 2019.

[10] CHEN T，SHI H Z，TANG S L，et al. CIL：Contrastive instance learning framework for distantly supervised 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 1：Long Papers). 2021：6191-6200.

[11] LI D，ZHANG T，HU N，et al. HiCLRE：A Hierarchical Contrastive Learning Framework for Distantly Supervised Relation Extraction[C]// Proceedings of the Conference on the Association for Computational Linguistics. 2022：2567-2578.

[12] RATHORE V，BADOLA K，SINGLA P，et al. PARE：A Simple and Strong Baseline for Monolingual and Multilingual Distantly Supervised Relation Extraction [C] // Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. 2022：340-354.

BAI Yu，born in 1982，lecturer. His main research interests is natural language processing.

WANG Xinzhe，born in 1998，postgraduate. Her main research interests include natural language processing and relationship recognition.