

DOI:10.19651/j.cnki.emt.2517968

# 基于原型的正负学习用于远程监督关系抽取

徐国梁<sup>1,2</sup> 陈祺东<sup>2</sup> 徐宇璇<sup>3</sup>

(1.南京信息工程大学计算机学院、网络空间安全学院 南京 210044;2.无锡学院物联网工程学院 无锡 214105;  
3.江南大学人工智能与计算机学院 无锡 214122)

**摘要:** 基于多示例学习框架的远程监督关系抽取方法大多依赖于启发式生成的污染标签,侧重于在句袋级别进行关系预测。然而,这些方法在句子级别的预测中表现不佳,而句子级别的预测更适用于理解性任务,如问答和知识图谱补全。为解决上述问题,本文提出了一种新型的远程监督关系抽取方法,该方法通过正负学习在句子级别训练模型,以区分噪声数据并加速收敛。同时,构建了一个约束图,用于编码关系与实体类型之间的约束,并通过辅助损失向关系原型优化,从而实现不同关系之间的信息传播,使得模型能够学到更为本质且可解释的句子表示。本文方法不仅能够识别噪声数据,还可以通过迭代修正其标签,以改进远程数据的质量,进一步提高模型性能。本方法在 NYT 数据集的句子级关系抽取任务中表现出色,精确度达 77.69%,较当前最优基准模型提升 6.47%,在噪声标注测试集上的 F1 分数高达 85.88%,验证了其卓越的去噪能力。消融实验结果表明,约束图对关系原型优化的贡献为 11.02%。实验结果表明,该方法在句子级别的关系抽取任务中显著优于现有方法,不仅有效减少了噪声影响,还显著提升了模型性能,为远程监督关系抽取任务提供了一个高效的解决方案。

**关键词:** 远程监督关系提取;正负学习;原型表示;句子级别关系抽取;噪声过滤

中图分类号: TP391.1; TN911.4 文献标识码: A 国家标准学科分类代码: 520.2020

## Positive and negative learning with prototype for distant supervision relation extraction

Xu Guoliang<sup>1,2</sup> Chen Qidong<sup>2</sup> Xu Yuxuan<sup>3</sup>

(1. School of Computer Science & School of Cyber-space Security, Nanjing University of Information Science & Technology, Nanjing 210044, China; 2. School of IoT Engineering, Wuxi University, Wuxi 214105, China;  
3. School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China)

**Abstract:** Distant supervision relation extraction methods based on the multi-instance learning framework mostly rely on contaminated labels that are heuristically generated, and focus on predicting relations at bag-level. However, they show unsatisfactory performance on sentence-level prediction which is more friendly with comprehend sentence tasks, like question answering and knowledge graph completion. To solve the above problems, a novel distant supervision relation extraction method is proposed in this paper, in which we train the model at sentence-level via positive learning and negative learning to separate noisy data and enable faster convergence. Meanwhile, a constraint graph is constructed to encode the re-strictions between relations and entity types and is optimized by an auxiliary loss towards relation prototype, which allows information propagation among different relations that makes the model can learn essential and interpretable sentence representation. We not only identify noisy data but also revise the labels of them iteratively to refine the quality of distant data and further enhance model performance. This method performs well in the sentence-level relation extraction task of the NYT dataset, with an accuracy of 77.69%, which is 6.47% higher than the current optimal baseline model. The F1 score on the noisy annotated test set is as high as 85.88%, verifying its excellent denoising ability. The ablation experiment results show that the contribution of the constraint graph to the optimization of the relation prototype is 11.02%. The experimental results show that this method significantly outperforms the existing methods in the sentence-level relation extraction task, not only effectively reducing the impact of noise, but also significantly improving the model performance, providing an efficient solution for the remote supervision relation extraction task.

**Keywords:** distant supervision relation extraction; positive and negative learning; prototypical representation; sentence-level relation extraction; noise filtering

## 0 引言

关系提取(relation extraction, RE)旨在预测非结构化

文本中标注实体对之间的关系。为了克服监督学习方法对标注训练数据的依赖,避免其耗时且费力的标注过程,远程监督(distant supervision, DS)方法被提出。远程监督关系

抽取(distant supervision relation extraction, DSRE)通过将知识库中的实体与语料库中对应的实体提及对齐,自动生成大规模标注数据集。

然而,DS 生成的训练数据通常包含大量噪声句子,这主要源于其假设:如果句子中的两个实体在知识库中存在关系,那么该句子就可能表达了这一关系。这种较强的假设常常导致错误标注,例如,某些句子被错误地标注为某个关系,而实际语境中并不存在这种关系。由于语料库与知识库的不完全性,另一个噪声来源是句子被错误地标注为“NA”(表示没有关系)。这两种类型的错误导致 DS 数据集中存在大量噪声,进而影响了模型的性能。因此,噪声问题已经成为研究者关注的重点。

为了缓解错误标注问题,许多基于多示例学习(multi-instance learning, MIL)框架的方法应运而生。这些方法将关系标签分配给包含共同实体对的句子集合(袋),并在袋级别进行分类。为了减少噪声,Zeng 等<sup>[1]</sup>仅选择概率最高的实例作为袋的表示,而 Lin 等<sup>[2]</sup>则通过注意机制为袋中的每个实例分配不同的权重。此外,Ye 等<sup>[3]</sup>采用了袋内和袋间的注意力机制,分别在句子级别和袋级别处理噪声。由于其出色的性能,MIL 受到了广泛关注,并成为 DSRE 领域的主流方法。

然而,基于 MIL 框架的模型仅在袋级别进行操作,无法识别句子与关系之间的一对一映射。换言之,这些模型不适用于许多下游任务中需要精确标注句子的句子级别关系抽取。为了降低句子级别预测中的噪声,研究者利用强化学习和对抗训练来选择可信数据<sup>[4-8]</sup>。与这些方法不同,Jia 等<sup>[9]</sup>致力于通过引导模型集中学习关系标签的模式,并通过自举学习迭代地选择可信实例。然而,他们忽略了噪声标注问题的根本原因,即缺乏正确的关系标签。

针对上述问题,本文提出了一种新颖的方法,称为基于原型表示的正负联合学习用于句子级远程监督关系抽取(jointly positive and negative learning with prototypical representation, PNPRL)。该方法通过识别噪声实例并校正其标签,来提高数据利用率并减少噪声。具体而言,模型通过正负联合学习(jointly positive and negative learning, JPNL)进行训练,能够在保持收敛性的同时,区分有效实例和噪声数据。然后,设计了噪声过滤策略和重新标注策略,用于筛选噪声实例,并将其转换为带有新标签的有效数据,从而迭代地提升数据集的质量。此外,实体类型与关系标签之间的约束被构建为约束图(constraint graph, CG),该图通过辅助损失向关系原型优化,并进一步编码以解决 DSRE 中的长尾问题。一方面,它促使头关系和样本较少的尾关系之间的信息传递;另一方面,它提高了模型的可解释性。

本文的主要贡献总结如下:

1) 本文通过 JPNL 训练模型,该方法能够保护模型免受噪声数据的干扰,并加速模型在预期清洁数据上的收敛。

此外,本文还设计了噪声过滤策略和重新标注策略,以精细化远程监督数据,并提高模型性能。

2) 本文在实体类型和关系标签的约束上应用图卷积网络(graph convolutional network, GCN)<sup>[10]</sup>,以促进来自数据丰富的头部关系向数据稀缺的长尾关系传递信息,这有助于解决长尾问题。

3) 提出了一种面向关系原型的辅助损失,用于从远程监督数据中学习本质且可解释的句子表示。

4) 本文在一个广泛使用的基准数据集上,开展了句子级别的评估、去噪能力的对比实验、消融实验以及案例研究。结果表明,本方法在性能上显著超越了现有的先进模型,为远程监督关系抽取任务提供了一个高效的解决方案。

## 1 相关工作

### 1.1 DSRE

研究者提出了远程监督方法,以减轻监督关系抽取模型所需训练数据收集的高昂成本,该方法从知识库中获取关系事实,并将这些关系事实与语料库中的句子对齐,以生成大规模训练数据。然而,由于远程假设导致的大量噪声数据严重影响了模型的性能。作为对原始远程监督假设的改进,At-Least-One 假设放宽了约束条件,认为包含相同实体对的句子集合(袋)中至少存在一个句子表达了知识库中对应关系。他们引入了多示例学习方法用于 DSRE,并在袋级别进行分类。基于该假设,Zeng 等<sup>[1]</sup>首次通过分段卷积神经网络(piecewise convolutional neural networks, PCNN)提取句子特征,并仅使用概率最高的句子进行学习。与 Zeng 等<sup>[11]</sup>采用的硬方法不同,Lin 等<sup>[2]</sup>采用了软方法,通过注意力机制为所有实例分配不同的权重。作为对数据的进一步利用,Chen 等<sup>[12]</sup>提出了一种对比实例学习框架,以无监督的方式利用每个实例,以提高模型的性能。Shang 等<sup>[13]</sup>根据注意力得分从句子集合中过滤噪声实例,然后利用深度聚类算法对其进行修正。蔡伟龙等<sup>[14]</sup>利用 SENet 和多头注意力进行关系抽取后,用句子级注意力机制来解决远程监督的噪声。

除了基于 MIL 框架的众多研究外,还有其他方法被探索用于解决噪声问题。Feng 等<sup>[4]</sup>通过强化学习训练实例选择器,为关系分类器提供可信的句子。Jia 等<sup>[9]</sup>提出了一种基于注意力正则化的去噪框架,以训练一个更具可解释性的模型来捕捉关系模式。赵红燕等<sup>[15]</sup>在词、句、包进行多层级降噪处理,最大限度的降低噪声对模型的影响。

### 1.2 负面学习

受到模型可能记住污染标签风险的启发,负学习(negative learning, NL)作为一种间接训练卷积神经网络的方法被提出。NL 通过随机选择互补标签来训练模型,假设“输入实例不属于这些互补标签”,从而减少了因高概

率未选择真实标签作为互补标签而提供错误信息的风险。NL 不仅能够避免噪声数据的过拟合,还能在训练过程中区分噪声数据和干净数据。Ma 等<sup>[16]</sup>首次将 NL 引入 DSRE 任务,并验证了其有效性。

### 1.3 原型表示学习

在 DSRE 中,原型是对给定关系的不同表述进行抽象的嵌入,代表了该关系最本质的语义,作为数据表示簇的中心,并被表达相同关系的句子所包围。Ding 等<sup>[17]</sup>在大规模远程标注数据的帮助下,通过原型级分类学习了稳健的关系表示。与原型计算为所有表述的平均嵌入的原型网络<sup>[18]</sup>不同,他们通过 3 种目标训练句子编码器,包括句子间的对比目标函数、原型与句子间的对比目标函数以及原型级分类目标函数,从而生成具有意义且可解释的关系表示。

### 1.4 DSRE 中的长尾问题

DSRE 中的另一个挑战是长尾问题,然而,与噪声问题相比,它往往被忽视。早期的研究通过基于规则的方法来丰富尾部关系,但这些方法需要大量的时间成本。Han

等<sup>[19]</sup>提出了关系层次树,通过考虑关系标签之间的语义关系来解决该问题。Zhang 等<sup>[20]</sup>在关系层次树上应用了 GCN,通过邻域集成机制促进信息传播。Liang 等<sup>[21]</sup>认为,关系层次树不应是解决长尾问题的唯一选择,他们通过约束图将信息转移到数据稀缺的关系上。张万里等<sup>[22]</sup>也使用约束图来解决长尾问题,他们创新性的使用约束图来补充实体类型与关系的约束信息。

## 2 方 法

本文提出了一种基于远程标签的句子级关系抽取框架,如图 1 所示,该框架由实例的句子编码器、约束图的图编码器和通过 JPNL 及原型引导学习(prototype-towards learning, PTTL)优化的关系分类器组成。此外,框架还配备了噪声过滤策略和重新标注策略,以在每次迭代中生成更清洁的数据集。在每次迭代中,首先初始化包含句子编码器、图编码器和关系分类器的模型,然后通过 JPNL 和原型导向损失函数进行训练,接着采用噪声过滤策略和重新标注策略来修正原始数据集。

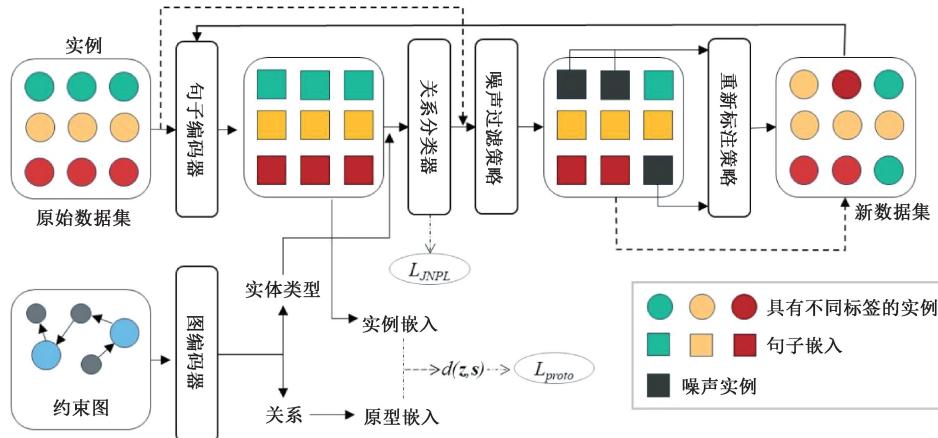


图 1 PNPRE 概述

Fig. 1 Overview of PNPRE

### 2.1 任务定义和符号

句子级关系抽取被定义为识别包含标注的头实体  $e_1$  和尾实体  $e_2$  的词序列  $s_i = \{w_1, w_2, \dots, w_n\}$  所表示的语义关系  $y_i \in R$ , 其中  $R = \{r_1, r_2, \dots, r_k\}$  是预定义的关系标签集合。本文将其视为一个分类任务。

### 2.2 句子编码器

句子编码器用于将句子转换为嵌入表示,进而表示输入上下文的特征。本文选择 PCNN 来编码输入嵌入,该嵌入通过将词嵌入和位置嵌入进行拼接获得,与 Zeng 等<sup>[1]</sup>中的方法相同。句子  $s_i$  的最终特征向量  $x_i \in R^{dh}$  由句子编码器  $SE_\varphi(\cdot)$  转换,其中  $\varphi$  是编码器的参数。

$$x_i = SE_\varphi(s_i) \quad (1)$$

值得注意的是,任何架构都可以作为本文框架中的句子编码器,因为它是与模型无关的。本文采用了最典型的

架构——PCNN 进行实现,该架构在参数较少的情况下能够达到令人满意的准确度。

### 2.3 约束图编码器

对于实体类型与关系标签之间的约束,例如,关系“/business/person/company”将其头实体限制为“人”类型,尾实体限制为“组织”类型,本文采用约束图来建模关系标签之间的语义依赖关系,如图 2 所示头实体和尾实体的类型分别是对应关系的前驱和后继。约束图由实体类型集合  $T = \{t_1, t_2, \dots, t_l\}$ 、关系集合  $R$  和约束集合  $C$  组成,其中约束被形式化为  $(t^{e_1}, r, t^{e_2})$ , 表示关系  $r$  的头实体类型和尾实体类型,其中  $t^{e_1}, t^{e_2} \in T$ 。约束图可以通过知识库提供的约束信息以及在训练实例中的统计的约束来构建。

给定约束图,构建一个节点集  $V = T \cup R$ , 将每个

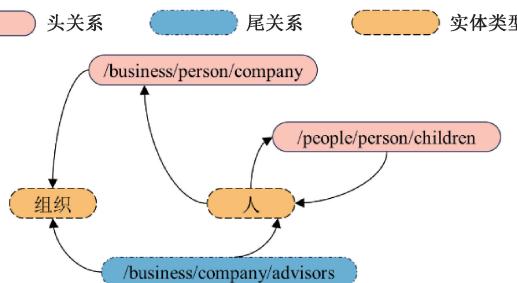


图 2 约束图的一部分

Fig. 2 A part of the constraint graph

$(t^{e1}, r, t^{e2}) \in C$  中的  $(t^{e1}, r)$  和  $(r, t^{e2})$  添加到边集  $E$  中。为了使用约束图中的信息,首先将其转换为向量表示。具体而言,第  $i$  个节点  $v_i \in V$  被表示为一个维度为  $d_v$  的嵌入向量,边集则表示为一个邻接矩阵  $A \in \mathbb{R}^{m \times m}$  (其中,  $m = |V|, |V| = l+k$ ):

$$A_{ij} = \begin{cases} 1, & (v_i, v_j) \in E \\ 0, & \text{其他} \end{cases} \quad (2)$$

至此,约束图可以通过嵌入矩阵  $V = \{v_1, v_2, \dots, v_m\}$  和邻接矩阵  $A$  表示。然后,通过两层 GCN 获得关系表示  $R = [r_1, r_2, \dots, r_k]$  和实体类型表示  $T = [t_1, t_2, \dots, t_k]$ , 其中  $R \in \mathbb{R}^{k \times dg}, T \in \mathbb{R}^{l \times dg}$ :

$$T \cup R = V = CGE_\sigma(V, E) \quad (3)$$

其中,  $\sigma$  是约束图编码器  $CGE_\sigma(\cdot)$  的一组参数。

## 2.4 分类器

正如文献[23-24]所讨论的,实体类型信息在关系提取中起着重要作用。因此,将  $\hat{t}$  与句子编码器的相应句子嵌入连接,作为分类的输入,其中,  $\hat{t}$  通过线性层转换得到,线性层的输入为实体类型  $t$ 。通过一个两层全连接分类器预测句子的最终概率分布:

$$P(r_i | s_i) = f_\theta(x_i, \hat{t}^{e1}; \hat{t}^{e2}), j = 1, 2, \dots, k \quad (4)$$

其中,  $x_i$  为句子嵌入,  $\hat{t}^{e1}$  和  $\hat{t}^{e2}$  分别为头实体类型和尾实体类型的嵌入, 符号“;”表示连接操作。

## 2.5 正负联合学习

正向学习(positive learning, PL): 将实例的远程标签表示为  $y$ , 其中  $y \in \{0, 1\}^k$  是一个  $k$  维的 one-hot 向量。PL 被广泛用于训练模型,使其能够预测给定的标签,如“输入实例属于该关系标签”:

$$L_{PL}(f, y) = - \sum_{i=1}^k y_i \log p_i \quad (5)$$

其中,  $p_i$  是关系  $r_i$  的对数。

负向学习: 与 PL 不同, NL 通过让模型从互补标签  $\bar{y} \in R - \{y\}$  中学习来进行训练。需要注意的是,互补标签意味着输入实例不属于该标签。通过实施 NL 方法,获得的模型可以在噪声环境下表现更好。其交叉熵损失函数为:

$$L_{NL}(f, y) = - \sum_{i=1}^k \bar{y}_i \log (1 - p_i) \quad (6)$$

正负联合学习: 本文通过 PL 和 NL 联合训练模型,结合了 PL 快速收敛的优势和 NL 在噪声环境下优异的表现。使用  $L_{NL}$  来计算每个实例的 NL 损失,但只选择预期可靠的实例作为计算 PL 损失的有效实例,从而考虑到噪声训练实例对模型的巨大影响。选择作为有效实例的标准包括两个方面: 1) 最大概率  $p_{max} > 0.5$ , 其中  $p_{max} = \max(p_1, p_2, \dots, p_k)$ ; 2) 其他类别的概率  $p_{max} < 1/k$ , 其中  $k$  是关系标签的数量。之所以不直接采用远程标签的最大概率作为标准,是因为该标签很可能是噪声。总的来说, JPNL 的损失函数可以描述为:

$$L_{JPNL} = \lambda_1 \cdot L_{PL} + L_{NL} \quad (7)$$

其中,  $\lambda_1$  是用于缩放 PL 整体幅度的系数,以确保其不会压倒 NL 的幅度。在实验中设置  $\lambda_1 = 0.1$ 。

## 2.6 原型导向学习

根据 Ding 等<sup>[17]</sup>的研究, 关系原型  $z \in Z$  抽象了关系  $r$  的最本质语义,且关系原型  $z$  的嵌入  $z$  与表达该关系的句子的嵌入处于相同的语义空间中。关系原型  $z \in \mathbb{R}^{d_p}$  的嵌入通过对由  $CGE_\sigma(\cdot)$  编码的关系表示进行线性映射获得:

$$z = MAP_1(r) \quad (8)$$

同样,来自句子编码器的句子嵌入  $x$  也通过线性映射到与关系原型  $z$  相同的语义空间中:

$$s = MAP_2(x) \quad (9)$$

本文认为,表达关系  $r$  的句子被称为  $r$  的陈述,并且每个陈述都应尽可能接近关系  $r$  的原型,同时应远离其他关系原型。在训练过程中,在式(7)中添加一个辅助损失,以训练一个原型导向模型:

$$L_{proto} = L_z + L_{z'} + \beta \cdot L_{z\_cls} \quad (10)$$

模型的优化目标是最大化句子嵌入和其对应的关系原型的相似度,同时最小化和其他关系原型的相似度。使用  $L_z$  和  $L_{z'}$  反映陈述和原型之间的归纳偏差,可以通过以下方式计算:

$$L_z = -\log d(z, s) \quad (11)$$

$$L_{z'} = -\sum_{z' \neq z, z' \in Z} (1 - \log d(z', s)) \quad (12)$$

式(11)鼓励句子嵌入与正确关系原型之间的更高相似度。式(12)通过降低句子嵌入与其他关系原型的相似度,进一步强化正确原型的区分性。式(11)中的相似度度量定义为:

$$d(z, s) = \frac{1}{1 + \exp(\frac{z}{\|z\|} \cdot \frac{s}{\|s\|})} \quad (13)$$

其中,  $z$  和  $s$  分别表示经过线性映射后的关系原型嵌入和句子嵌入。 $z$  和  $s$  是它们的欧几里得范数。

为进一步规范化原型的语义,在原型级别使用一个额外的分类器,其参数  $\gamma$  通过以下方式进行优化:

$$L_{z\_cls} = - \sum_{i=1}^k \log P(r_i | z_i) = - \sum_{i=1}^k \log g_\gamma(z_i) \quad (14)$$

综上所述,最终的损失函数由两部分组成:

$$L = L_{JPNL} + \lambda_2 \cdot L_{proto} \quad (15)$$

其中,  $\lambda_2$  决定了  $L_{proto}$  的权重, 在实验中将其设置为 0.1, 可以控制  $L_{proto}$  的梯度贡献在整个优化过程中处于适当水平, 既不会干扰 JPNL 的优化动态, 也能保证原型的优化效果。

值得一提的是, 本文仅在有效实例上执行 PTTL, 这与 PL 相同, 因为原型对噪声较为敏感。

## 2.7 噪声过滤与重新标注策略

**噪声过滤策略:** 经过多个训练周期后, 噪声实例通常会被分配到远程标签的较低概率值, 而干净的数据则通常被预测为中等或更高的概率。本文并未设置一个固定的阈值来区分噪声和训练数据, 而是基于一种名为 OTSU 的二值化算法来计算每个关系  $r$  在训练样本中的区分阈值, 该算法通过最大化类间方差来确定阈值。通过这种方式, 噪声过滤阈值不仅适应了每个类别的收敛程度, 还会自动和动态变化, 避免了选择合适值的麻烦, 并且对长尾数据更加友好。

**重新标注策略:** 噪声标注问题的根本原因并非缺乏有用信息, 而是缺失的关系标签。因此, 对通过之前噪声过滤策略筛选出的噪声实例进行重新标注。如果句子被预测为关系  $r^*$  的概率为  $P(r^* | s)$ , 且该概率大于  $r^*$  的重新标注阈值  $Th_{r^*}$ , 则将其重新标注为  $r^*$ , 否则为“NA”。注意, 在式(16)中的  $Th_{r^*}$  等于  $r^*$  噪声过滤阈值。

$$r^* = \begin{cases} r^*, P(r^* | s) > Th_{r^*} \\ NA, \text{ 其他} \end{cases} \quad (16)$$

## 2.8 模型训练算法

为了最大程度地利用每个部分的优势, 通过迭代训练模型, 逐步生成干净的数据集, 最终在验证集上以最佳的 F1 分数停止训练。模型训练算法的过程如算法 1 所示。

### Algorithm 1: Iterative Training Algorithm

**Input:** DS dataset  $D_0$ , RE model  $M$  parameterized by  $\varphi, \sigma, \theta, \gamma$ , the noise-filtering strategy, the re-labeling strategy

**Output:** A sentence-level RE model

```

1   for iteration  $i = 1, 2, \dots$  do
2     Initialize the model and train  $M_i$  on train set
       $D_{i-1}$  with Eq. (15);
3     Predict on  $M_i$  with  $M_i$ ;
4     for each  $r \in R$  do
5       Compute  $Th_r$  with OTSU;
6     end for
7     for each subset  $I_r$  with label  $r$  of train set do
8       Filtering noisy instances by  $Th_r$  and generate
       new label by the re-labeling strategy;
```

```

9   end for
10  Revised train set  $D_i$  ;
11   $i++$  ;
12 end for
```

## 3 实验

### 3.1 数据集和评估

本文在广泛使用的 DSRE 数据集上评估了所提出的框架。该数据集是基于主流远程监督方法构建的 NYT 数据集, 已被广泛应用于关系抽取任务的研究。在句子级关系抽取训练阶段, 遵循 Jia 等<sup>[9]</sup>的设置, 他们发布了一个手动标注的句子级测试集和一个用于评估算法去噪能力的测试集。数据集摘要如表 1 所示。该数据集包含 371 461 个训练实例, 其中正例数量为 110 518, 占总实例的 29.75%, 验证集和测试集中正负比例相对平衡, 分别包含 337 和 323 个实例。数据集呈现长尾特性, 各个关系分布如表 2 所示。

表 1 数据集统计数据<sup>[9]</sup>

Table 1 Statistics of dataset<sup>[9]</sup>

NYT-10	Train	Dev	Test
# Instances	371 461	2 379	2 164
# Positive	110 518	337	323

注: “Positive”表示未标有“NA”的正向实例。

表 2 数据集关系类型统计数据

Table 2 Dataset relationship type statistics

关系	Train	Dev	Test
# /location/location/contains	53 504	173	144
# /people/person/nationality	8 349	24	42
# /location/country/capital	7 959	8	5
# /people/person/place lived	7 438	75	73
# /business/person/company	5 788	38	46
# /location/nei.../neighborhood of	5 737	0	1
# /people/person/place of birth	3 279	6	8
# /people/person/place of death	2 002	5	4
# /business/company/founders	827	5	6
# /people/person/children	523	5	3

例如最常见的关系/location/location/contains 在训练集中有 53 504 个实例, 而关系/people/person/children 在训练集中仅有 523 个实例, 说明数据集存在较为显著的不均匀分布。数据集实体类型分布如表 3 所示, LOCATION 实体在训练集中的数量最多, 共计 402 511 个, 占所有实体的比例最大; 其次是 PERSON 实体, 为 302 915 个, 而 ORGANIZATION 实体的数量相对较少, 为 93 225 个。

**表 3 数据集实体类型统计数据**  
**Table 3 Dataset entity type statistics**

实体	Train	Dev	Test
LOCATION	402 511	1 064	1 046
PERSON	302 915	470	490
ORGANIZATION	93 225	170	149

根据先前的文献,直接在数据集中的所有单个实例上计算精确度(precision, P)、召回率(recall, R)和 F1 分数,用于句子级评估。

### 3.2 实验环境与配置

本研究所有实验均在台式计算机环境下完成。硬件配置方面,实验平台采用了 Intel Core i7-13700K 处理器以及 NVIDIA GeForce RTX 3090 显卡。软件环境方面,选择 Python 3.8 作为开发语言,并基于 PyTorch 1.8.0 框架构建模型,同时使用 CUDA 11.1 以充分发挥 GPU 加速性能。在 NYT-10 数据集上的训练过程中,整个模型的训练时间大约为 4 h。

### 3.3 基线

本文将所提出的框架与多个 DSRE 强基线模型进行了比较,包括:

CNN<sup>[1]</sup>是一种典型的 RE 架构,引入了位置嵌入以揭示单词与实体对之间的相对位置。

PCNN<sup>[2]</sup>是对 CNN 的改进,广泛应用于 RE 任务。它使用分段最大池化来提取更多的关系特征。

PCNN+SelATT<sup>[2]</sup>是一个袋级 RE 模型,它采用注意力机制对袋中的实例进行加权,以减少噪声实例的权重。

PCNN+ATT\_RA+BAG\_ATT<sup>[3]</sup>是一个袋级 RE 模型,利用袋内和袋间注意力机制处理噪声。

CNN+RL1<sup>[8]</sup>是一种基于强化学习的袋级方法。识别假阳性实例并将其重新分配为负样本。

CNN+RL2<sup>[4]</sup>是一个句子级 RE 模型,通过联合训练 CNN 分类器和实例选择器以去除无效样本。

ARNOR<sup>[9]</sup>是一个句子级 DSRE 模型,其中设计了注意力正则化机制,关注关系模式,并根据注意力得分选择可信实例。

SENT<sup>[6]</sup>是句子级模型,基于 NL 迭代执行噪声过滤和重新标注。

为了与已有工作进行客观对比,本文在句子级评估中直接引用了文献[16]中所报告的 CNN、PCNN、PCNN+SelATT 等基线模型的实验结果。文献[16]在数据规模、实验场景与评价指标方面与本研究保持一致,因而其结果在相同评测环境下具备良好的可比性和参考价值。

此外,为帮助读者更好地理解这些基线模型在远程监督关系抽取任务中的常见实现方式,本文在表 4 中列出了领域内较为通用的超参数配置,其中,对于 SGD 优化器,初始学习率一般设定在 0.01~0.1 之间;而对于 Adam 优化

器,初始学习率则设定在  $1 \times 10^{-5}$  至  $1 \times 10^{-3}$  区间内。表 4 一方面能够在一定程度上展示 CNN、PCNN、PCNN+SelATT 等模型的典型设置,另一方面也便于后续研究在相似条件下复现或比较。

**表 4 基线模型通用的超参数设置**

**Table 4 Common hyperparameter settings for baseline models**

参数	参数值
window size	3
word dimension	50
filter number	230
dropout rate	0.5
optimizer	SGD/Adam

### 3.4 实现细节

由于 PNPRE 是一个与模型无关的框架,考虑到训练时间和计算成本的情况下,使用 PCNN 实现了本文的模型。随机初始化 50 维的词嵌入和 50 维的位置嵌入,与 ARNOR 中的设置一致。PCNN 中的过滤器数量设置为 230,窗口大小设置为 3,和之前的工作相同。为了编码 CG,使用了一个输入嵌入维度为 50 的两层 GCN。对于 JPNL,设置  $\lambda_1 = 0.1$ ,并随机选择 10 个互补标签,与 SENT 中的设置一致。对于 PTTL,使用预期可靠实例在一个批次中占所有实例的比例来缩放原型级损失。在训练的过程中,使用 Adam 优化器,并在分类器之前应用 dropout 以防止过拟合。表 5 显示了实验中使用的所有超参数。

**表 5 超参数设置**

**Table 5 Hyper-parameters settings**

组件	参数	参数值
句子编码器	word dimension	50
	position dimension	50
	max length	256
	filter number	230
图编码器	window size	3
	input dimension	50
	layer1 dimension	128
	output dimension	128
JPNL	negative samples	10
	$\lambda_1$	0.1
	$d_p$	128
	$\lambda_2$	0.1
PTTL	batch size	256
	learning rate	$5 \times 10^{-4}$
	dropout rate	0.5
	epoch number	15

### 3.5 主要结果

将 PNPRE 的结果与其他基准模型进行了句子级评估比较。如表 6 所示,本文的方法在精确度和 F1 分数上显著优于基准模型。从结果中可以观察到:1)所有基准模型在句子级评估中都表现不佳,表明噪声数据对句子级训练造成了巨大的影响。2)包括 PCNN+SelATT、PCNN+ATT\_RA+BAG\_ATT 和 CNN+RL1 在内的袋级模型在句子级关系提取任务上表现不佳,表明袋级去噪方法不适合句子级评估。3)提出的 PNPRE 方法在精确度上取得了显著的提升,同时保持了可比的召回率,导致 F1 分数高于之前的去噪方法,表明本文的方法能够有效减少噪声数据的影响。

为了进一步证明 PNPRE 的去噪能力,本文在噪声标注测试集上进行了实验。如表 7 所示,本文所提的方法在 F1 分数上的表现优于其他测试方法,并且在保持与 SENT 相似的精确度的同时,召回率取得了显著的提高。值得注意的是,虽然 SENT 中的噪声过滤策略可以适应不同的关系,但仍然需要设置全局的数据过滤阈值和重新标注阈值,这一过程较为繁琐。相比之下,本文将过滤阈值和重新标注阈值合并为一个,并通过 OTSU 算法自动生成,保持其自适应性。此外,ARNOR 在 3 种方法中表现最差。这可能是因为 ARNOR 中的初始模式集可能阻碍了模型对多样化正确数据的泛化。

表 6 所有比较模型在句子级评估中的表现

Table 6 Performance of all compared models on sentence-level evaluation

方法	验证集			测试集		
	Prec.	Rec.	F1	Prec.	Rec.	F1
CNN	38.32	<b>65.22</b>	48.28	35.75	64.54	46.01
PCNN	36.09	63.66	46.07	36.06	<b>64.86</b>	46.35
PCNN+SelATT	46.01	30.43	36.64	45.41	30.03	36.15
PCNN+ATT_RA+BAG_ATT	49.84	46.90	48.33	56.76	50.60	53.50
CNN+RL <sub>1</sub>	37.71	52.66	43.95	39.41	61.61	48.07
CNN+RL <sub>2</sub>	40.00	59.17	47.73	40.23	63.78	49.34
ARNOR	62.45	58.51	60.36	65.23	56.79	60.90
SENT (BiLSTM)	66.71	57.27	61.63	71.22	59.75	64.99
PNPRE (PCNN)	<b>74.28</b>	60.83	<b>66.88</b>	<b>77.69</b>	60.37	<b>67.94</b>

注:除 PNPRE 外,先前模型的结果均引自文献[16]。

表 7 在 NYT-10 噪声标注测试集上的去噪能力比较

Table 7 Comparison of de-noising ability on a noise-annotated test set of NYT-10

去噪	Prec.	Rec.	F1
ARNOR	76.37	68.13	72.02
SENT (BiLSTM)	<b>80.00</b>	88.46	84.02
PNPRE (PCNN)	79.35	<b>93.59</b>	<b>85.88</b>

### 3.6 消融研究

本文同时进行了消融研究,以展示 PNPRE 中不同组件的有效性:1)在“-CGE”中,CGE 被移除,并用随机初始化的实体类型和关系的嵌入替代。2)在“-L<sub>proto</sub>”中,L<sub>proto</sub> 被移除,即在训练阶段不使用 PTTL。3)在“-CGE-L<sub>proto</sub>”中,表示既没有使用 CGE,也没有使用 L<sub>proto</sub>。4)在“-PL”中,移除了 JPNL 中的 PL,仅使用 NL 来优化关系分类器,与 SENT 相同。5)在“-OTSU”中,用 SENT 的设置替换了本文中的噪声过滤和重新标注策略。具体而言,噪声过滤的全局阈值设置为 0.25,重新标注阈值为 0.7。

对 NYT-10 测试集的消融研究结果如表 8 所示。显

然,移除 CGE 和 L<sub>proto</sub> 会损害最终性能,因为信息通过约束图中的节点和边有效传播,而 L<sub>proto</sub> 有助于模型获取关系的本质语义。然而,当同时移除 CGE 和 L<sub>proto</sub> 时,模型的表现要优于单独移除它们的情况。基于这一结果,可以推测,若缺乏 L<sub>proto</sub>,CGE 可能无法得到充分优化,而如果没有 CGE,关系原型嵌入可能会偏向包含噪声的句子嵌入。

表 8 对 NYT-10 测试集的消融研究

Table 8 Ablation study on NYT-10 test set

模型	Prec.	Rec.	F1
PNPRE (PCNN)	<b>77.69</b>	<b>60.37</b>	<b>67.94</b>
-CGE	66.67	59.44	62.85
-L <sub>proto</sub>	66.07	57.28	61.36
-CGE-L <sub>proto</sub>	68.59	58.82	63.33
-PL	66.29	54.18	59.63
-OTSU threshold	67.06	52.94	59.17

删除 PL 后,模型的性能主要取决于 SENT 中使用的互补标签,因此性能大幅下降。在图 3 中,分别使用 JPNL 和 NL 对训练数据进行训练,并绘制了交叉熵损失的变化趋势。可以观察到,去除 PL 后,方法的收敛性表现更差,

验证了 JPNL 的优越性。

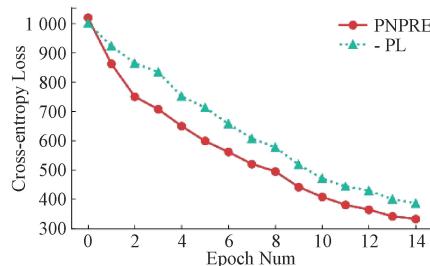


图 3 一次迭代中训练集的交叉熵损失曲线

Fig. 3 Cross-entropy loss curve on train set in one iteration

当把由 OTSU 确定的噪声过滤阈值替换为 SENT 方法时,模型的最终性能在分类和去噪方面都大大降低(如表 8、表 9 所示)。此外,图 4 展示了在两种不同阈值下,方法在长尾关系和头部关系中的效果对比,两种阈值分别为 SENT 设置的阈值和本文方法得到的阈值。可以明显看出,本文的过滤阈值比 SENT 更为严格,这也解释了为什么 PNPRE 在召回率上表现更好,因为它更倾向于尽可能过滤掉噪声数据,而不是错误地将干净数据误判为噪声。经过多次过滤和重新标注,最终的模型远超未使用 OTSU 的版本。

表 9 噪声标注测试集的消融研究

Table 9 Ablation study on noise-annotated test set

模型	去噪		
	P.	R.	F1
PNPRE (PCNN)	<b>79.35</b>	<b>93.59</b>	<b>85.88</b>
-OTSU threshold	77.88	82.37	80.06

### 3.7 案例研究

为了展示 PNPRE 在噪声检测和重新标注方面的能力,通过一个案例研究展示了模型从训练集中筛选出的部分句子(如表 10 所示)。如表中前两行所示,因实体对未参

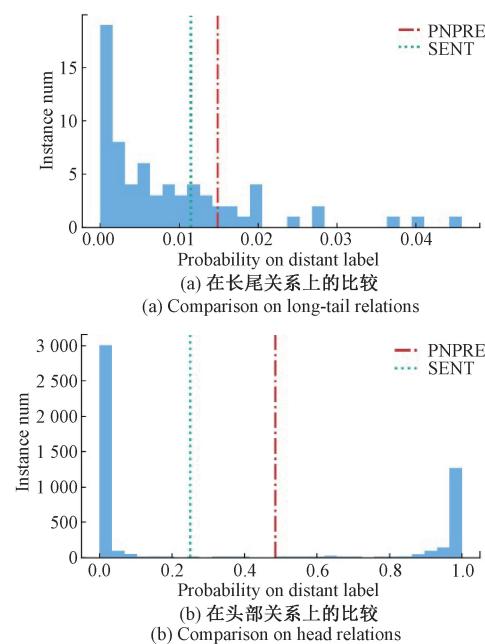


图 4 本文方法与 SENT 的噪声过滤阈值比较

Fig. 4 Comparison of noise-filtering thresholds between the proposed method and SENT

与知识库中的关系三元组,每个句子均被远程监督标注为“NA”,但实际上它们表达的是正向标签。PNPRE 识别了这些噪声句子,并进行了修正(将第 1 句重新标注为“/location/location/contains”,第 2 句重新标注为“/people/person/place\_lived”)。接下来的两句为正向实例,但被错误地标注为其他相似的正向关系标签,模型能够对其进行修正(将第 3 句重新标注为“/location/location/contains”,第 4 句重新标注为“/people/person/nationality”)。最后两句虽然被标注为正向远程标签,但实际上没有表达任何关系,表明了模型不仅能够处理假阴性问题,还能修正假阳性实例。

表 10 从训练集中筛选出的部分噪声句子

Table 10 Some noisy sentences filtered from train set

句子	远程标签	重新标注
His mother is an assistant principal of mathematics at Middle School 72 in <u>Jamaica</u> , <u>Queens</u> .	NA	/location/location /contains
The police said Mr. Buthorn pulled over the van's driver, identified as <u>Richard Williams</u> of <u>Brooklyn</u> , who was arrested and charged with leaving the scene of an accident and driving with a suspended license.	NA	/people/person /place_lived
He and Ms. Smith married in 1985, and the two trained in <u>Helsinki</u> , <u>Finland</u> , on a Fulbright grant for a year before returning to New York.	/location/country /capital	/location/location /contains
The English and Chinese-language media here have been full of unfavorable <u>Singapore</u> 's prime minister, <u>Lee Hsien Loong</u> , said Friday that Iran's role in the Shanghai organization was a way for Russia and China to demonstrate their influence.	/people/person /place_of_birth	/people/person /nationality

表 10 (续)  
Table 10 (continued)

句子	远程标签	重新标注
The English and Chinese-language media here have been full of unfavorable coverage of the park's modest size, and the financial secretary of <b>Hong Kong</b> , <u>Henry Tang</u> , has suggested that Disney reconsider how many people the park can really hold.	/people/person /place_of_birth	NA
Although <u>Iran</u> had no immediate response to the proposal, Western diplomats have said they expect <b>Tehran</b> initially to reject the proposal, which would take years to negotiate and carry out.	/location/country /capital	NA

注:加粗和下划线文本分别代表头实体和尾实体。

## 4 结 论

本文提出了PNPRE,一种用于DSRE的句子级框架。一方面,采用JPNL训练模型,不仅能够缓解噪声问题并区分噪声与干净数据,还能提高模型的收敛性和训练速度,从而提高模型性能。另一方面,引入约束图,通过实体类型和关系标签的约束,建模关系之间的依赖性,促进不同关系之间信息的传播。此外,PTTL被设计作为辅助模块,用于学习句子的最本质的语义。通过迭代进行噪声过滤和修正,PNPRE能够提高远程标签数据的质量并优化模型性能。在NYT-10数据集上的实验结果表明,PNPRE在句子级RE和噪声减少方面优于现有方法。

## 参考文献

- [1] ZENG D, LIU K, CHEN Y, et al. Distant supervision for relation extraction via piecewise convolutional neural networks[C]. 2015 Conference on Empirical Methods in Natural Language Processing, 2015: 1753-1762.
- [2] LIN Y K, SHEN SH Q, LIU ZH Y, et al. Neural relation extraction with selective attention over instances[C]. 54th Annual Meeting of the Association for Computational Linguistics, 2016: 2124-2133.
- [3] YE ZH X, LING ZH H. Distant supervision relation extraction with intra-bag and inter-bag attentions[J]. ArXiv preprint arXiv:1904.00143, 2019.
- [4] FENG J, HUANG M L, ZHAO L, et al. Reinforcement learning for relation classification from noisy data [C]. AAAI Conference on Artificial Intelligence, 2018.
- [5] QIN P D, XU W R, WANG W Y. DSGAN: Generative adversarial training for distant supervision relation extraction [J]. ArXiv preprint arXiv: 1805.09929, 2018.
- [6] HAN X, LIU ZH Y, SUN M S. Denoising distant supervision for relation extraction via instance-level adversarial training [J]. ArXiv preprint arXiv: 1805.10959, 2018.
- [7] ZENG X R, HE SH ZH, LIU K, et al. Large scaled relation extraction with reinforcement learning[C]. AAAI Conference on Artificial Intelligence, 2018.
- [8] QIN P D, XU W R, WANG W Y. Robust distant supervision relation extraction via deep reinforcement learning[J]. ArXiv preprint arXiv:1805.09927, 2018.
- [9] JIA W, DAI D, XIAO X Y, et al. ARNOR: Attention regularization based noise reduction for distant supervision relation classification [C]. 57th Annual Meeting of the Association for Computational Linguistics, 2019: 1399-1408.
- [10] 李文悦,何怡刚,邢致恺,等.基于双输入残差图卷积网络的电力变压器健康状态评估方法[J].电子测量与仪器学报,2024,38(11):15-24.  
LI W Y, HE Y G, XING ZH K, et al. Health evaluation of power transformer based on double input residual graph convolutional network[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(11):15-24.
- [11] ZENG D J, LIU K, LAI S W, et al. Relation classification via convolutional deep neural network[C]. COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, 2014: 2335-2344.
- [12] CHEN T, SHI H ZH, TANG S L, et al. CIL: Contrastive instance learning framework for distantly supervised relation extraction [J]. ArXiv preprint arXiv:2106.10855, 2021.
- [13] SHANG Y M, HUANG H Y, MAO X L, et al. Are noisy sentences useless for distant supervised relation extraction? [C]. AAAI Conference on Artificial Intelligence, 2020, 34(5): 8799-8806.
- [14] 蔡伟龙,毛建华.基于多头自注意力和SENet的远程监督关系抽取[J].电子测量技术,2020,43(21):132-136.  
CAI W L, MAO J H. Distant supervision relationship extraction based on multi-head self-attention and SENet [J]. Electronic Measurement Technology, 2020,43(21): 132-136.
- [15] 赵红燕,张莹刚,谢斌红.基于多层次注意力机制和动态阈值的远程监督关系抽取[J].计算机应用研究,2024,41(11):3288-3294.  
ZHAO H Y, ZHANG Y G, XIE B H. Distant supervision relation extraction based on multi-level

- attention mechanism and dynamic threshold [J]. Application Research of Computers, 2024, 41(11): 3288-3294.
- [16] MA R T, GUI T, LI L Y, et al. SENT: Sentence-level distant relation extraction via negative training[J]. ArXiv preprint arXiv:2106.11566, 2021.
- [17] DING N, WANG X B, FU Y, et al. Prototypical representation learning for relation extraction [J]. ArXiv preprint arXiv:2103.11647, 2021.
- [18] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning [C]. Advances in Neural Information Processing Systems, 2017.
- [19] HAN X, YU P F, LIU ZH Y, et al. Hierarchical relation extraction with coarse-to-fine grained attention[C]. 2018 Conference on Empirical Methods in Natural Language Processing, 2018: 2236-2245.
- [20] ZHANG N Y, DENG SH M, SUN ZH L, et al. Long-tail relation extraction via knowledge graph embeddings and graph convolution networks [J]. ArXiv preprint arXiv:1903.01306, 2019.
- [21] LIANG T M, LIU Y, LIU X Y, et al. Distantly-supervised long-tailed relation extraction using constraint graphs[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(7): 6852-6865.
- [22] 张万里, 佟安, 李文桥. 基于约束图的远程监督长尾关系抽取方法[J]. 现代电子技术, 2024, 47(21): 91-96.
- ZHANG W L, TONG AN, LI W Q. Distantly-supervised long-tailed relation extraction based on constraint graph[J]. Modern Electronics Technique, 2024, 47(21): 91-96.
- [23] LYU S, CHEN H H. Relation classification with entity type restriction [J]. ArXiv preprint arXiv: 2105.08393, 2021.
- [24] ZHOU W X, CHEN M H. An improved baseline for sentence-level relation extraction[J]. ArXiv preprint arXiv:2102.01373, 2021.

### 作者简介

徐国梁,硕士研究生,主要研究方向为深度学习、关系抽取。

陈祺东(通信作者),博士,讲师,主要研究方向为自然语言处理、远程监督关系抽取。

E-mail:cqd\_jnu@hotmail.com

徐宇璇,硕士,高级工程师,主要研究方向为信息抽取、问答系统。