

文章编号: 1000-5641(2020)05-0113-18

## 基于远程监督的关系抽取技术

王嘉宁<sup>1</sup>, 何 怡<sup>2</sup>, 朱仁煜<sup>1</sup>, 刘婷婷<sup>1</sup>, 高 明<sup>1</sup>

(1. 华东师范大学 数据科学与工程学院, 上海 200062; 2. 上海市大数据中心, 上海 200072)

**摘要:** 关系抽取作为一种经典的自然语言处理任务, 广泛应用于知识图谱的构建与补全、知识库问答和文本摘要等领域, 旨在抽取目标实体对之间的语义关系. 为了能够高效地构建大规模监督语料, 基于远程监督的关系抽取方法被提出, 通过将文本与现有知识库进行对齐来实现自动标注. 然而由于过强的假设使得其面临诸多挑战, 从而吸引了研究者的关注. 本文首先介绍远程监督关系抽取的概念和形式化描述, 其次从噪声、信息匮乏以及非均衡 3 个方面对比分析相关方法及其优缺点, 接着对评估数据集以及评测指标进行了解释和对比分析, 最后探讨了远程监督关系抽取面对的新的挑战以及未来发展趋势, 并在最后做出总结.

**关键词:** 关系抽取; 远程监督; 自然语言处理; 知识图谱; 噪声处理

**中图分类号:** TP311 **文献标志码:** A **DOI:** 10.3969/j.issn.1000-5641.202091006

## Relation extraction via distant supervision technology

WANG Jianing<sup>1</sup>, HE Yi<sup>2</sup>, ZHU Renyu<sup>1</sup>, LIU Tingting<sup>1</sup>, GAO Ming<sup>1</sup>

(1. School of Data Science and Engineering, East China Normal University, Shanghai 200062, China;

2. Shanghai Municipal Big Data Center, Shanghai 200072, China)

**Abstract:** Relation extraction is one of the classic natural language processing tasks that has been widely used in knowledge graph construction and completion, knowledge base question answering, and text summarization. It aims to extract the semantic relation from a target entity pair. In order to construct a large-scale supervised corpus efficiently, a distant supervision method was proposed to realize automatic annotation by aligning the text with the existing knowledge base. However, it highlights a series of challenges as a result of over-strong assumptions and, accordingly, has attracted the attention of researchers. Firstly, this paper introduces the theories of distant supervision relation extraction and the corresponding formal descriptions. Secondly, we systematically analyze related methods and their respective pros and cons from three perspectives: noisy data, insufficient information, and data imbalance. Next, we explain and compare some benchmark corpus and evaluation metrics. Lastly, we highlight new subsequent challenges for distant supervision relation extraction and discuss trends and directions of future research before concluding.

**Keywords:** relation extraction; distant supervision; natural language processing; knowledge graph; noise processing

收稿日期: 2020-08-07

基金项目: 国家重点研发计划 (2016YFB1000905); 国家自然科学基金 (U1911203, U1811264, 61877018, 61672234, 61672384); 中央高校基本科研业务费专项资金; 上海市科技兴农推广项目 (T20170303); 上海市核心数学与实践重点实验室资助项目 (18dz2271000)

通信作者: 何 怡, 女, 工程师, 研究方向为数据运营、数据分析、用户画像及社会网络挖掘.

E-mail: yhe01@shanghai.gov.cn

## 0 引言

随着计算机技术的发展和 Web 数据的日益激增, 如何从这些海量数据中获取有效的信息成为当前的研究热点. 信息抽取作为一种经典的自然语言处理任务, 其主要目标是从半结构化或非结构化的数据中提取结构化的语义信息, 是构建丰富的结构化知识库的核心部分<sup>[1]</sup>. 关系抽取是信息抽取的一个关键步骤, 其建立在给定的已知实体对和文本描述基础上, 挖掘实体对之间的语义关系, 形成由实体、关系、实体组成的三元组, 因此关系抽取通常应用在知识图谱构建与补全<sup>[2]</sup>(Knowledge Graph Construction and Completion)、知识库问答<sup>[3]</sup>(Knowledge Base Question Answering) 和文本摘要<sup>[4]</sup>(Text Summarization) 等领域.

关系抽取主要分为两大类, 一种是以无监督为主的自动抽取 (Auto Extraction), 另一种是以有监督或半监督为主的关系分类 (Relation Classification). 前者通常在没有确定关系标签的条件下, 根据语料的句法或语义结构自动地从文本中提取可以描述对应关系的词汇或短语. 例如, 我们可以从句子“南京是江苏省的省会城市”中根据句法结构自动提取出实体“南京”和“江苏省”的关系是“省会城市”. Hasegawa<sup>[5]</sup> 和 Oren<sup>[6]</sup> 等人提出了一些在无监督条件下自动从文本中抽取语义关系的方法. 后者通常将关系抽取视为一种分类任务, 即事先预定义有限个关系标签, 并对语料进行人工标注. 有监督的关系分类主要方法是对实体和文本通过特征工程或表征学习来提取句法或语义信息, 并通过分类器完成对关系的分类<sup>[7-8]</sup>. 由于有监督关系分类依赖于标注语料的质量和数量, 而人工标注语料的成本高、效率低, 半监督或弱监督的关系抽取成为研究的热点. 如 Kumlien<sup>[9]</sup> 提出了一种弱监督学习方法来从文本中提取结构化的语义关系.

基于远程监督的关系抽取方法 (Distant Supervision Relation Extraction, DSRE) 是由 Mintz<sup>[10]</sup> 提出的一种弱监督学习方法, 其主要思想是通过将大量的无标注语料与现有的知识库进行对齐, 从而快速地构建大规模监督语料. 图 1 给出一个知识库对齐的示例<sup>[11]</sup>, 如知识库中存在两个实体 Steve Jobs 和 Apple 时, 如果语料中存在这两个实体, 则这些语料是与这两个实体对齐的. 远程监督方法因其构建语料效率高、成本低的原因吸引了国内外研究者的关注. 例如, 杨<sup>[12]</sup> 等人利用远程监督方法自动构建初等数学语料并实现知识点的关系提取; Riedel<sup>[13]</sup> 通过获取纽约时报 (New York Times, NYT) 新闻语料并与知识库 Freebase<sup>[14]</sup> 对齐形成 NYT 数据集; Jat 等<sup>[15]</sup> 通过与谷歌 Web 数据库对齐制作了 GDS(Google-IISc Distant Supervision) 语料; Han 等通过远程监督方法构建了用于验证少样本学习的 FewRel 数据集<sup>[16]</sup>.

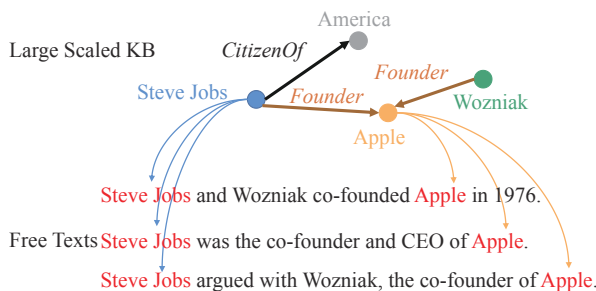


图 1 知识库与语料对齐的示例<sup>[11]</sup>

Fig. 1 An example of alignment between the knowledge base and corpus<sup>[11]</sup>

虽然远程监督可以解决语料的数量问题, 但不能保证语料的质量, 因为远程监督方法基于一个假设: 如果两个实体在知识库中具有某种关系, 任意一个包含这两个实体的文本都描述了这种关系. 而这一假设往往是不成立的, 从而导致产生的语料中存在大量的错误标签. 因此, 为了避免这一假设

对关系抽取性能的影响,如今有诸多工作尝试解决如何在含有大量噪声的语料中去学习真实的实体对关系<sup>[17-19]</sup>.然而在实际训练过程中,除了噪声对模型会产生影响外,研究者们还发现由于知识库不充分,部分实体对得不到充分的训练<sup>[20-22]</sup>,且每个标签对应的样本严重不均衡<sup>[23]</sup>,这些客观存在的问题无不影响远程监督关系抽取模型的性能,使得如今模型的测试效果非常低.因此结合近年来的研究工作,本文认为远程监督关系抽取面临的3个最大挑战是:

(1) 噪声.主要体现在被标注的样本标签与实际文本语义关系不匹配,导致产生大量的错误正样本(False Positive).

(2) 信息匮乏.由于知识库是不完善的,远程监督语料缺乏充分的语义信息来有效提升关系抽取的效果.与此同时还导致大量的样本被标注为无关(False Negative).

(3) 非均衡.非均衡是指关系标签对应的样本数量不均衡问题,主要体现在少部分的标签对应的样本数量多,而绝大多数的标签对应的样本数量很少,因此也称作二八定律、马太效应或长尾问题.

不同于现有一些调研工作仅从方法角度进行分析<sup>[24-25]</sup>,本文将以上述提到的3个研究挑战为出发点,重点讨论当前基于远程监督的关系抽取方法是如何应对这些挑战的.

本文第1节将介绍远程监督关系抽取的相关定义和问题描述;第2节将从应对噪声问题、缓解信息匮乏问题和解决非均衡问题3个方向介绍相关研究工作,并引出规则统计、多示例学习、对抗与强化学习、辅助信息增强、联合学习和少样本学习等技术;第3节将介绍远程监督关系抽取的评估数据集和评测指标;第4节将介绍一些新的研究挑战和未来发展趋势;第5节进行总结.

## 1 基本概念与任务描述

### 1.1 概念定义

关系抽取是非常重要的自然语言处理任务之一,与命名实体识别、实体属性抽取、知识推理等构成知识图谱构建的核心步骤<sup>[1]</sup>.关系抽取旨在从包含两个目标实体的非结构化文本中预测它们之间的语义关系<sup>[25]</sup>,因此关系抽取任务可以建模成多分类问题.为避免标注训练数据,远程监督基于给定的知识库实现目标实体对与非结构化文本的对齐,从而构建多分类模型的训练数据集.其中一些基本概念的定义如下:

(1) 目标实体对(Target Entity Pair) 实体是对抽象世界的描述,它是描述客观事物的名词或短语,对应于知识库中则是相互独立的节点.目标实体对则是由两个不同的实体组成的元组,通常这两个实体是有顺序的,第一个实体被称为头实体(Head Entity),第二个实体被称为尾实体(Tail Entity),分别可用 $h, t$ 表示.

(2) 关系(Relation) 关系是头实体与尾实体间语义联系的描述,在知识库中通常以边的形式存在,可以用 $r$ 表示.关系可以分为有向关系和无向关系.有向关系是描述头实体到尾实体的方位关系,如上下位、IsA、PlaceOfBirth等;如果是无向关系,则头尾实体通常是等价的,如Equal、Friend.

(3) 三元组(Triple) 是构成结构化知识的最小的单位,也被称为事实(Fact),通常用 $\mathcal{F} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ 表示,其中 $\mathcal{E}$ 表示实体集合, $\mathcal{R}$ 表示关系集合.如果满足 $(h, r, t) \in \mathcal{F}$ ,则表明头尾实体之间存在一个关系 $r$ .

(4) 知识库(Knowledge Base) 知识库是由结构化的知识组成的集合,现有的知识库包括FreeBase<sup>[14]</sup>、YaGo<sup>[26]</sup>等.通常表示为图数据结构 $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{F})$ .

### 1.2 任务描述

下面给出远程监督语料构建方法的形式化的描述.

**描述一** 给定一个知识库 $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{F})$ ,其中 $\mathcal{E}$ 表示实体集合, $\mathcal{R}$ 表示实体间关系集合, $\mathcal{F}$ 表示

所有三元组. 给定一个文本  $S$  和两个目标实体  $h, t$ . 如果同时满足条件  $h \in S$  且  $h \in \mathcal{E}$ ,  $t \in S$  且  $t \in \mathcal{E}$ ,  $(h, r, t) \in \mathcal{F}$ , 其中  $r \in \mathcal{R}$ , 则认为文本  $S$  是可以描述关系  $r$  的, 即  $(h, r, t)$  与  $S$  是对齐的. 为了方便描述, 本文记作  $(h, r, t) \sim S$ .

通常关系抽取被视为句子级别分类<sup>[27]</sup>任务, 然而因为远程监督方法是以实体对为基础与文本进行对齐的, 因此也有人认为远程监督关系抽取应作为一种多示例学习任务<sup>[17,28-29]</sup>. 本文以多示例学习为例给出如下基于远程监督的关系抽取的形式化描述.

**描述二** 假设根据远程监督方法(描述一)获得了一组语料  $B = \{(B_1, r_1), (B_2, r_2), \dots, (B_N, r_N)\}$ , 其中  $B_i = \{S_{i1}, S_{i2}, \dots, S_{i|B_i|}\}$ , 称之为包(bag),  $(h_i, r_i, t_i) \sim B_i$ ,  $|B_i|$  为包内句子的数量. 因此当给定一个包  $B_i$ , 预测为关系  $r_j$  的概率可表示为  $Pr(r = r_j | B_i, h_i, t_i)$ .

## 2 远程监督关系抽取研究问题与相关方法

现阶段基于远程监督的关系抽取工作从问题解决角度可分为 3 类: ①应对错误标注所产生的噪声问题; ②缓解因知识库不全和数据缺乏导致的信息匮乏问题; ③解决语料不均衡带来的长尾问题. 其中第 1 类是自远程监督方法提出以来一直被关注的问题, 形成了基于规则统计方法、多示例学习、对抗与强化学习等技术. 后两类是最近被关注的新问题, 主要对应于辅助信息增强、联合学习和少样本学习等技术. 本节分别介绍这 3 个方面的研究技术现状. 表 1 整理了针对 3 类挑战的相关研究方法.

表 1 远程监督关系抽取研究问题及相关方法

Tab. 1 Research problems and related methods for distant supervision relation extraction

研究挑战	技术类别	代表性方法	描述
噪声	规则统计	核方法与依存关系 <sup>[18]</sup> 、概率图 <sup>[13,30]</sup> 、矩阵补全 <sup>[31-32]</sup>	利用实体关系规则判断句子与标签是否匹配
	多示例学习	PCNN <sup>[17]</sup> 、多标记 <sup>[28-29]</sup> 、EM算法 <sup>[33-35]</sup> 、注意力机制 <sup>[36-41]</sup> 、正则化 <sup>[42]</sup> 、语言模型 <sup>[43]</sup>	使用包(bag)作为分类的单位, 并通过多示例学习来降低噪声对分类的影响
	对抗与强化学习	对抗样本训练 <sup>[19,44]</sup> 、生成对抗网络 <sup>[45-47]</sup> 、策略梯度 <sup>[11,48-50]</sup> 、Q学习 <sup>[49]</sup>	自动地从语料中过滤噪声, 将高质量的语料用于训练, 提升分类效果
信息匮乏	辅助信息增强	实体关系信息 <sup>[20-22,51]</sup> 、知识表示 <sup>[52-53]</sup> 、条件约束 <sup>[54-55]</sup>	引入额外的知识进行增强, 弥补由于知识库不充分导致的信息匮乏
	联合学习	监督与半监督联合学习 <sup>[56]</sup> 、实体关系联合抽取 <sup>[57-59]</sup>	结合其他任务进行端到端学习
非均衡	少样本学习	多任务学习 <sup>[60]</sup> 、语法规则 <sup>[23,61]</sup> 、关系层次表征 <sup>[62-63]</sup>	捕捉丰富的头尾数据的相关性, 缓解长尾关系预测不准确问题

### 2.1 应对噪声问题

在机器学习与自然语言处理领域中, 噪声是指真实标签与被标注标签不一致的样本. 在基于远程监督的关系抽取任务中, 因为过强的假设使得许多包含在训练数据中的句子并不能真实描述目标实体对间的关系, 导致产生大量的假正例(False Positive)和假负例(False Negative). 例如, 知识库中存在一个三元组(*Obama*, *PlaceOfBirth*, *US*), 那么与文本对齐后的结果如表 2 所示, 可知只有第一个句子能够表达 *PlaceOfBirth* 的关系, 而另外 4 个句子都属于噪声. 由于对齐的语料没有标注是否为噪声, 因此如何在关系抽取的过程中识别出噪声成为一个研究难题. 如今降噪的技术主要有基于规则统计、基于多示例学习和基于对抗与强化学习等几大类.

#### 2.1.1 规则统计方法

规则统计方法旨在通过挖掘实体对与关系之间的规则联系, 代表方法有核方法与依存关系、概率图模型、矩阵补全等.



表 2 远程监督语料噪声的示例

Tab. 2 Some examples of distant supervision noisy data

示例	对齐标签	正确标签
... [Obama] was born in [US]...	PlaceOfBirth	PlaceOfBirth
... [Obama] have said he loved [US]...	PlaceOfBirth	NA
... [Obama] was lived in [US] last year...	PlaceOfBirth	PlaceOfLived
... [Obama] was the president of [US] during 2008 and 2016...	PlaceOfBirth	President
... [Obama] will leave [US] for China...	PlaceOfBirth	NA

Zelenko<sup>[18]</sup> 提出一种基于核方法和依存关系的关系抽取. 其通过使用依存关系方法将文本解析为语法树, 并定义了一种基于语法树的节点相似匹配核函数, 使用支持向量机 (Support Vector Machine, SVM) 算法进行多类分类. 该方法在提出时达到了最优效果, 但并未考虑如何解决噪声问题.

Riedel<sup>[13]</sup> 提出一种基于概率图的方法, 为每个句子分配一个隐变量表示该句子是否是噪声, 其将关系抽取视为一种对隐变量的赋值任务. Takamatsu<sup>[30]</sup> 在此基础上进行了改进, 提出一种概率生成模型来模拟远程监督方法的启发式标记过程. 该模型也通过隐变量来预测标签是否正确.

Fan 等<sup>[31]</sup> 人则提出一种基于矩阵补全的方法, 通过将训练集的特征、标签和测试集的特征按照特征组合成分块矩阵, 剩余的一块则是测试集未知的标签矩阵, 因此其提出一种基于矩阵的秩最小化和核范数最小化的矩阵补全方法来预测未知标签. Zhang<sup>[32]</sup> 则对 Fan<sup>[3]</sup> 提出的方法做出了改进, 使用基于无参贝叶斯进行训练, 避免了大量调参. 虽然矩阵补全可以有效地在含有噪声的语料中预测关系, 但不适用于大规模数据, 且当新来数据时必须重构矩阵.

基于规则统计的方法可以挖掘出实体间显式的结构信息, 同时对关系的提取具有可解释性. 又因为该类大部分方法通常不需要训练的过程, 所以对时间和空间复杂度要求低. 但其过度依赖于大量的特征工程使得分类效果差, 泛化能力低.

### 2.1.2 多示例学习

多示例学习 (Multi-Instance Learning, MIL) 方法是目前远程监督关系抽取的主流方法之一, 其基本思想是将相同的实体对对齐的文本组成一个包 (或称为句袋), 并以包为单位进行预测. 按照预测方法, 可分为错误标签预测、至少一个假设和注意力机制.

(1) 错误标签预测 (Wrong Label Prediction) 旨在通过捕捉包及包内句子与标签的语义相关性来直接预测是否是噪声.

Hoffmann 提出一种经典的多示例学习方法——MultiR<sup>[28]</sup>, 其为每个句子分配与关系数量相等的隐变量, 并分别进行错误标签预测. MultiR 方法是通过对包内句子预测结果取并集来表示每个包的预测结果, 其采用类似于感知机的模型来学习参数. Surdeanu 则在 MultiR 方法的基础上做出改进, 提出一种新的多示例多标记 MIML 方法<sup>[29]</sup>, 其基本思路是通过模型捕捉包与不同标签的相关性进行预测, 不同于 MultiR 方法, 其依然采用基于概率统计的贝叶斯模型学习参数, 在准确率与时空复杂度上都优于 MultiR 方法.

Min<sup>[33]</sup> 和 Xu<sup>[34]</sup> 等人发现由于知识库不全以及噪声原因产生了大量的假负例 (False Negative) 数据. 其通过概率图模型, 使用 EM 算法预测每个句子是否为真. Ritter<sup>[35]</sup> 则同时兼顾 False Positive 和 False Negative 并使用一种软性约束机制缓解句子与包标签不匹配问题.

(2) 至少一个假设 (At-Least-One Assumption) 其建立在一个新的假设基础上, 即对齐的文本中至少存在一个句子可以描述实体对关系, 因此每次只选择包内一个句子进行预测.

Mintz 认为与知识库对齐的文本中至少存在一个文本可以描述实体对关系, 并基于此提出了 At-

Least-One 假设<sup>[10]</sup>, 在一个包内只选择概率最大的句子作为实体对关系的分类样本. 该假设在一定程度上可以保证正确的样本对分类的正面作用.

Zeng 基于这个假设提出一种分段的卷积神经网络<sup>[17]</sup>(Piecewise Convolutional Neural Networks, PCNN) 自动捕捉上下文信息. 首先使用预训练的词向量 (Word Embedding)<sup>[64-65]</sup> 将单词映射到低维度向量空间, 然后使用卷积神经网络对文本进行特征提取. 不同于传统文本分类的 CNN, 其根据两个实体将文本划分为三段, 并为每一段进行最大池化, 该方法被证明可以有效地处理长文本表征能力, 并降低时间复杂度<sup>[17]</sup>. 另外 Zeng 还使用位置表示信息 (Position Embedding), 使得模型可以更好地学习到实体在文本中的结构化信息, 进一步提升泛化能力.

(3) 注意力机制 (Attention Mechanism) 是目前关系抽取任务中的热门方法之一, 目标是为同一个包内的每个句子分配权重. 如果句子是噪声, 则分配较低的权重以降低其对包分类的影响; 如果句子不是噪声, 则会获得较高的权重以凸显其重要性. 换句话说, 注意力机制通过对每个句子进行加权求和来强化真实标签数据、弱化错误标签数据.

Lin<sup>[36]</sup> 认为基于 At-Least-One Assumption 分类时只选择一个句子, 会损失大量的语义信息, 因此提出一种句子级别的注意力机制 (Sentence-Level Attention). 首先使用 PCNN 对包内的句子进行卷积核最大池化. 然后对于每个句子与关系向量计算相似度并使用 softmax 归一化为权重, 以此对包内句子进行加权求和获得包表征 (Bag Embedding). 最后直接应用一层神经网络对这个包进行分类. 基于句子级别注意力的降噪方法也成为多示例学习的主流方法. Ji<sup>[37]</sup> 改进了 Lin 的权重分配方法<sup>[36]</sup>, 提出一种结合实体描述信息的句子级别注意力机制. 其改进之处是先对头尾实体进行表征, 并在模型训练过程中进行微调, 然后用尾实体向量与头实体向量的差来近似代替关系向量并将其与每个句子计算权重.

Jat<sup>[38]</sup> 认为每个单词对关系预测的贡献是不一样的, 因此提出一种基于单词和实体级别的注意力, 每个单词分配不同的权重用于表示句子向量, 实体则用于对不同的句子分配权重. 同时将单词和句子进行加权求和可以提取更关键的语义成分来提升预测能力.

Wu<sup>[39]</sup> 则在 PCNN 的基础上添加了神经噪声转换器 (Neural Noise Converter) 学习结构化的转移矩阵来获得含有噪声的数据集, 并使用条件最优选择器 (Conditional Optimal Selector) 从噪声中选择权重最大的句子用于分类.

Ye 认为除了包内句子与句子之间存在相关性, 包与包之间也存在相关性, 因此提出一种包内 (Intra-Bag) 和包间 (Inter-Bag) 注意力<sup>[40]</sup>. 对于包内的句子则是对所有句子进行加权求和; 对于包与包之间, Ye 提出一种包组 (Bag Group) 的概念, 将具有相同关系标签的包组在一起, 为每一个包学习一个注意力权重, 并进行加权求和形成包组向量. 通过结合句子级别注意力和包级别注意力可以更好地降噪. Yuan<sup>[41]</sup> 也提出了包级别注意力, 但不同于 Ye<sup>[40]</sup>, 其首先使用 PCNN 和句子级别注意力获得包表征, 然后提出一种交叉关系注意力 (Cross-Relation Attention) 将包与每个关系标签结合起来计算权重, 同时通过贝叶斯公式求得后验概率, 并使用 softmax 归一化后的权重对不同的关系向量进行加权求和以获得融合关系信息的包向量, 再通过交叉包注意力 (Cross-Bag Attention) 将所有包与同一个关系融合的向量加权求和组合成超包 (SuperBag), 最后通过超包学习不同实体对与关系之间的语义联系并进行预测.

Jia<sup>[42]</sup> 从另一个角度使用注意力对噪声进行处理, 他认为句子中的模式 (Pattern) 能够对降噪起到很好的作用, 因此提出一种注意力正则化 (ARNOR) 方法, 通过关注句子中的不同模式来对句子是否是噪声进行判断. Alt 等人则结合了语言模型 (Language Model)<sup>[43]</sup>, 利用 BERT<sup>[66]</sup> 等预训练语言模型获得文本的语义表达, 再通过一个分类器学习实体对与关系标签的映射. 预训练模型通常包含大量的先验知识, 将其迁移到关系抽取上可以进一步缓解噪声的影响.

随着深度学习的发展,如今基于多示例方法通常以神经网络模型为主,相比基于规则统计的传统方法,避免了特征工程带来的误差传播风险,使得模型更具有泛化能力,但与此同时,时间和空间复杂度要求要高于传统的方法.

### 2.1.3 对抗与强化学习

虽然多示例学习可以很好地平滑噪声的影响,但是在训练过程中依然会存在两个问题:

(1) 无法处理所有包内句子都是噪声的情况. 如果某一个实体对对齐的文本全部都是噪声,对于多示例学习方法, At-Least-One Assumption 和注意力机制都会至少为一个句子分配一个较大的权重,因此无法处理所有句子都是噪声的情况.

(2) 基于包的预测并不能很好地处理句子与关系标签的映射,另外,其容忍了一部分噪声使得模型学习到了一些错误的映射关系,从而可能影响分类效果.

随着对抗学习<sup>[67-68]</sup>与强化学习<sup>[69]</sup>的发展,以及它们在自然语言处理中的广泛应用,基于远程监督的关系抽取中噪声处理问题形成了一个新的思路——学习一种采样策略,使得在训练模型之前先过滤错误标注的句子,通过提升语料的质量来提升关系抽取的效果.

Shi<sup>[19]</sup>和 Wu<sup>[44]</sup>均提出一种远程监督关系抽取对抗训练的思路. 通过在对句子使用 RNN 进行表征的时候,在每一个时刻将模拟为噪声的对抗样本与真实样本一同输入模型中进行训练,试图让模型能够识别出真样本和噪声样本,与此同时提升模型在含有噪声的语料中训练的鲁棒性.

Qin<sup>[45]</sup>则使用生成对抗网络<sup>[67]</sup>(Generative Adversarial Nets, GAN)来模拟采样过程. 其参考多示例学习方法将语料划分为包,并在每个包内通过采样器(Sampler)根据学习的概率分布采样出真实样本企图欺骗判别器(Discriminator),剩余的则是采样器认为的噪声. 判别器则将采样器采样的真实样本认为是噪声并给予低分,而将剩余的噪声认为是真实样本给予高分. 因为 GAN 的训练比较困难<sup>[67]</sup>,因此 Qin 首先对采样器和判别器进行预训练,其次两者进行联合训练. 对 GAN 的训练是一个极小极大问题,最终可以获得一个较优的采样器使得其采样的样本都可以被认为是正确的.

不同于 Qin<sup>[45]</sup>, Li<sup>[46]</sup>认为大量的 False Negative 样本影响了关系抽取效果,因此首先通过对维基百科中的实体描述进行统计构建出小规模专家语料,认为如果某个实体的描述信息中出现了另一个实体,则这两个实体间是有关系的. 其次在对抗训练过程中,生成器分别从原始语料和专家语料中采样,判别器则分别给予低分和高分. Han<sup>[47]</sup>等人提出另一种降噪方法,首先通过预训练的分类器将远程监督语料划分为2个集合,分别是确信集(Confident Part)和不确信集(Unconfident Part). 在对抗训练阶段,通过采样器从确信集中采样企图欺骗判别器,并不断将被打高分的样本更新到确信集中,直到2个集合不再变化为止.

Feng<sup>[48]</sup>从强化学习的角度提出一种远程监督语料的降噪方法,在基于马尔可夫决策过程条件下,定义了状态、动作、奖励等基本要素. 基本思路是利用策略梯度 REINFORCE 算法<sup>[48]</sup>学习一个选择器(Selector)在包中对每个句子执行一个动作,即选择或不选择,并在已选择的句子集合中使用句子级别的 PCNN 分类器(Classifier)进行分类,并通过挑选的句子 loss 的均值作为奖励. 实验表明这种方法可以有效提升关系抽取的效果,尤其是对于全部都是噪声的包能够起到过滤作用. He<sup>[49]</sup>改进了 Feng 的方法,主要体现在:①依然使用基于包级别的预测,认为只考虑句子级别的分类会丧失句子之间的语义关联性;②使用 Q 学习来训练选择器;③同时考虑了 True Positive 和 Unlabeled 所产生的损失,实验表明识别 True Positive 的精确度得到了提升.

Qin<sup>[50]</sup>则站在挑选噪声的角度,将每相邻的两次迭代所挑选噪声的交集作为计算奖励的依据. 不同于 Feng 的是, Qin 认为使用 F1 值的变化作为奖励可以提高策略的学习能力. Zeng<sup>[11]</sup>和 Sun<sup>[70]</sup>则训练一个智能体进行关系预测,动作空间的大小是关系标签的数量,而奖励则为对包分类的预测准确度.



通过策略梯度法不断调整智能体的策略使得能够给出正确的关系预测。

基于对抗与强化学习的降噪方法进一步地提升了语料的质量,让模型可以学习到更加准确的实体对关系,进一步提升训练模型的泛化能力和鲁棒性。但是相比于前两种方法,对抗学习与强化学习都需要训练两个模型,训练难度大,通常需要先对各个模块进行预训练来避免在联合训练过程中的不稳定问题<sup>[45,48]</sup>,因此其时间和空间复杂度要求进一步提高,在实际工业界应用场景中难以落地。

## 2.2 缓解信息匮乏问题

由于现阶段知识库和训练语料大量缺失,使得部分实体对对齐的语料数量过少,从而造成这些实体对的关系抽取训练不充分,最终导致信息匮乏<sup>[20,51]</sup>。如图2所示,Kuang分别对NYT和GDS数据集的每个实体对在数据集中的共现次数进行了统计<sup>[22]</sup>,其中横坐标表示实体对共现次数,纵坐标表示对应的实体对数量。发现绝大多数的实体对仅存在少量的样本,在本身还有大量噪声的情况下,真实样本含有的语义信息十分稀少,不利于训练。如今解决远程监督关系抽取中信息匮乏问题的主要分为辅助信息增强和联合学习。

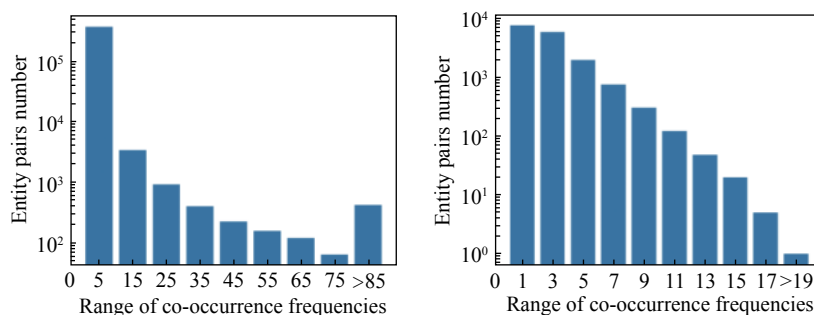


图2 NYT(左)和GDS(右)中实体对共现次数统计<sup>[22]</sup>

Fig. 2 Statistics of co-occurrence frequencies of entity pairs in the NYT (left) and GDS (right) datasets<sup>[22]</sup>

### 2.2.1 辅助信息增强

辅助信息增强是一种最直接地缓解训练不充分的问题的方法,其基本思路是通过引入额外知识信息来提升对关系的预测,如实体关系信息、条件约束、知识表示等。

Vashishth<sup>[20]</sup>认为知识库中包含一些辅助信息(Side Information)以及实体类型信息(Entity Type)可以辅助增强关系抽取:在计算句子权重之前将辅助信息与句子向量结合起来;在形成包向量之后直接与实体类型对应的向量进行拼接。相比之下,Li<sup>[21]</sup>则是使用自注意力机制(Self-attention)结合实体信息来实现对语义信息的增强,其验证了自注意力机制可以有效地帮助模型关注更重要的语义成分以弥补数据不充分的缺陷。

Kuang<sup>[22]</sup>认为实体对在大量维基百科语料中的共现次数可以间接地描述它们之间的潜在关系,从而弥补训练集中部分实体对不充分的问题。其构建了一个实体邻接图(Entity Proximity Graph),其中节点表示各个实体,边上的权值代表对应相连的两个实体在维基百科语料中的共现次数。随后使用LINE<sup>[71]</sup>图算法从低阶和高阶两个方面捕捉实体对之间的隐式相互关系(Implicit Mutual Relation),并结合实体类型信息进行增强。与此同时,Su<sup>[51]</sup>是同时在文本和知识库两个方面对实体对共现次数进行统计,并提出一种全局关系表征(Global Relation Embedding)用于提升关系预测能力。

Xu<sup>[52-53]</sup>则结合了知识表示(Knowledge Base Embedding, KBE)和文本句子表征,提出基于异构表征方法来增强远程监督关系抽取。其思路是使用双向长短期记忆神经网络(Bi-directional Long Short-Term Memory, BiLSTM)<sup>[72]</sup>对句子进行表征,并基于TransE模型<sup>[73]</sup>学习知识库内的实体和关系向量。文本句子表征与知识表征分别对应使用交叉信息熵的局部损失和基于相似度排序的全局损失,在实



验中二者结合起来联合训练.

Liu 则考虑到了实体类型会对关系预测起到约束作用,如果两个实体类型分别是“Person”和“Film”,则预测的关系应当更可能是“DirectorOf”,而不可能是“PlaceOfBirth”.因此 Liu 提出一种多粒度的实体类型约束方法<sup>[54]</sup>,并集成到现有的远程监督关系抽取模型中. Ye 也考虑到了这类约束,并分别从连贯性 (Coherent) 和语义性 (Semantic) 两个层面上提出两种不同的约束损失函数计算方法<sup>[55]</sup>,该工作不依赖于关系抽取模型,因此具有即插即用 (Plug-and-Play) 的优势.

### 2.2.2 联合学习

联合学习则是另一类缓解信息匮乏问题的方法. Beltag<sup>[56]</sup>认为远程监督关系抽取中的语料是不充分的,但可以通过其他有监督语料来辅助增强,因此提出一种联合监督与远程监督的训练方法.不过使用监督语料并不能从根本上解决训练不充分问题.

Wei<sup>[57]</sup>提出一种联合命名实体识别 (Named Entity Recognition, NER) 和关系抽取的端到端方法,其基本思路是先对文本中的实体进行标注,获得了实体层面的语义信息后映射到对应的关系语义表征上.该工作在远程监督数据集 NYT 上验证了同时结合两个任务学习模型可以弥补只进行关系抽取任务可能出现的学习不充分的问题. Ren 提出一种联合抽取模型 COTYPE<sup>[58]</sup>,主要解决实体标注与关系分类分离导致的错误传播问题,他认为联合两个任务学习还可以缓解噪声问题.

Takanobu<sup>[59]</sup>的出发点与 Wei<sup>[57]</sup>相同,但实现策略完全相反.其是先挖掘文本中的关系,再根据关系到文本中寻找两个实体.为了能够很好地提取关系和实体, Takanobu 使用一种分层的强化学习方法,先后交替训练关系抽取和实体识别,并在 NYT 数据集上验证了该方法既可以通过联合两个任务缓解含有噪声的语料中的训练不充分的问题,也能够处理关系重叠问题 (Overlapping).

## 2.3 解决非均衡问题

非均衡是远程监督语料面临的新问题,近几年开始被关注.非均衡问题主要体现在各个关系标签对应的实体对数量或对齐的语料数量不均匀.通常只有少部分标签对应的实体对或语料数量较多,而大部分标签则很少,这种现象被称为二八定律、马太效应或长尾问题.例如, Zhang 对 NYT 语料进行了统计<sup>[62]</sup>,如图 3 所示,横轴表示关系标签的编号,按照对应样本数降序排序,纵轴表示各个关系标签对应样本的数量. NYT 一共包括 53 个关系标签,居于头部的关系标签语料充足,而近 40 个居于尾部的关系标签对应的语料严重缺乏,使得居于长尾的实体对不能够得到充分的学习,致使训练得到有偏的模型.

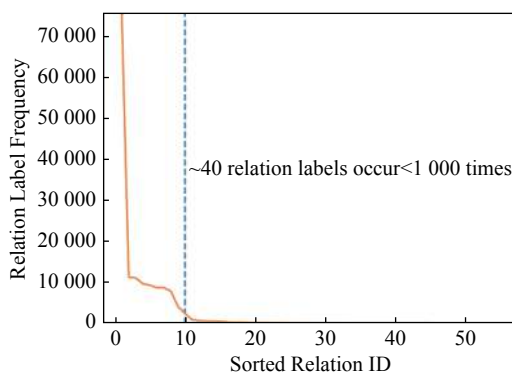


图3 NYT数据集各个关系标签频数统计<sup>[62]</sup>

Fig. 3 Statistics of relation label frequency of the NYT corpus<sup>[62]</sup>

为了解决非均衡的问题, Ye<sup>[61]</sup>等人提出一种多任务学习 (Multi-Task) 框架,认为绝大多数的实体对关系是 NA(无关),只有少数的实体对具有关系.因此在 CNN 的输出层部分添加两个分类器,分别

为 Relation Identification 和 Relation Classification. 前者使用交叉信息熵进行关系识别, 即判断当前实体对是否存在关系, 后者则使用排序损失方法预测具体的关系. 但这种方法只能够缓解 NA 标签对分类的影响.

Krause<sup>[23]</sup> 提出一种基于规则的关系抽取系统解决长尾问题. 其从远程监督语料中提取了大量的语法规则, 每个关系标签平均有 40k 个规则模板. 然而大量的规则极大地消耗了人力物力, 且局限于特定领域的关系抽取. 为了解决规则带来的问题, Gui<sup>[61]</sup> 等人进行了改进, 提出一种基于可解释学习 (Explanation-Based Learning, EBL)<sup>[61]</sup> 的方法, 其可以在少量规则的引导下提取出实体关系信息, 同时对于长尾的关系, 还可以通过领域知识进行纠正.

解决长尾问题的代表性方法还有少样本学习 (Few-shot Learning). Han<sup>[63]</sup> 则利用了深度学习技术来解决长尾问题. Han 发现在 FreeBase<sup>[14]</sup> 等知识库中关系标签是按照层次结构组织的, 如关系 “/people/person/place\_of\_birth” 可以分为 “/people” “/people/person” 和 “/people/person/place\_of\_birth” 3 层, 对于长尾关系通常表现在第 3 层, 但是可以通过前两层使得长尾关系的样本容量变大. 基于此, 其提出一种分层的关系抽取方法, 根据关系的层次结构由粗到细地 (Coarse-to-Fine) 进行预测, 进一步缓解长尾问题.

Zhang<sup>[62]</sup> 认为基于 Few-shot 的训练方法具有挑战性, 因此在 Han<sup>[63]</sup> 基础上进行改进. 其先后使用 TransE<sup>[73]</sup> 和图卷积网络 (Graph Convolutional Network, GCN) 模型<sup>[74]</sup> 对知识库的图结构以及层次关系的树结构进行表征, 分别获得隐式和显式的关系信息, 进而通过由粗到细粒度的注意力机制将不同层次的信息进行融合. 该方法可以很好地借助关系的层次结构以及知识库信息来缓解长尾问题.

### 3 评估数据集与评测指标

#### 3.1 评估数据集

在国际评测中, 监督类实体关系抽取的公开数据集主要有 SemEval 2010 Task8<sup>[75]</sup>、ACE2004/2005. 前者是国际语义评测会议 SemEval 于 2009 年构建的语义关系分类任务, 其包含 9 种关系共 10 717 个示例, 语料主要来源于词网 (WordNet)、维基百科 (Wikipedia) 等. 后者是 ACE 会议公开使用的数据集, 其中包括实体识别和关系抽取两个子任务, 语料来源于语言数据联盟 (Linguistic Data Consortium, LDC), 通常可以被用于联合抽取任务. 另外还有 MUC、MPQA2.0 等语料. 虽然这些有监督的关系抽取数据集被广泛用于评测模型, 但它们共同的缺点在于数据集的语料数量过少、大量人工标注成本过高、因局限于特定领域而不能得到广泛的应用. 因此有一系列相关工作致力于远程监督方法构建数据集, 表 3 分别列出了相关数据集的统计信息.

表 3 评测数据集统计信息  
Tab. 3 Statistics of the evaluate dataset

数据集	训练集示例	训练集实体对	测试集示例	测试集实体对	关系标签
NYT <sup>[13]</sup>	522 611	281 270	172 448	96 678	53
NYT11 <sup>[59]</sup>	62 648	74 312	369	370	12
GDS <sup>[15]</sup>	13 161	7 580	5 663	3 247	5
KBP <sup>[76]</sup>	9 153 100	183 062	166 700	3 334	41
FewRel <sup>[16]</sup>	56 000	56 000	14 000	14 000	100

(1) NYT 数据集 Riedel 通过将纽约时报与 FreeBase<sup>[14]</sup> 对齐产生大量的语料并简称 NYT 数据集<sup>[13]</sup>, 其中训练集由 2005 年和 2006 年新闻组成, 测试集由 2007 年新闻组成. NYT 数据集包括 52 种语义关

系和 NA 标签,训练集共包括 522611 个示例、281270 个实体对,测试集包括 172448 个示例、96678 个实体对. NYT 也成为目前远程监督关系抽取使用最为广泛的评估数据集. 后来有相关工作将 NYT 原始数据集进行了简约处理,构造了 NYT11<sup>[59]</sup> 等版本语料,如表 3 所示,实体对数量之所以比示例数量多,是因为存在一些句子包含多个实体对或相同的实体对有多个关系.

(2) GDS 数据集 Jat 等人认为 NYT 数据集中包含大量的 NA 标签致使模型不易学习到正确的映射关系,因此用远程监督方法额外构建了 GDS (Google-IISc Distant Supervision) 数据集<sup>[15]</sup>, 其将谷歌关系抽取语料库与 Web 进行对齐. 其中训练集有 13161 个示例、7580 个实体对,测试集包括 5663 个示例、3247 个实体对,一共有 4 种语义关系以及 NA 标签. GDS 相比 NYT 数据集来说降低了数据的规模,且使得 5 种标签对应的示例数量相对平衡.

(3) KBP 数据集 是由 Surdeanu<sup>[76]</sup> 通过 KBP2010<sup>[77]</sup> 和 KBP2011<sup>[78]</sup> 评测任务上获得的语料,该语料是将 150 万个文档与知识库对齐产生的. KBP 数据集中包含 183062 个训练实体对,3334 个测试实体对和 41 种关系标签,且平均每个实体对对应 50 个句子<sup>[76]</sup>. 因此该语料具有规模大、数据真实等特点<sup>[33]</sup>. 不同于前面的数据集, Surdeanu 设计了 200 个基于实体名称的头实体查询,并根据头实体获得对应的所有三元组及其对齐的语料<sup>[76]</sup>. 例如, Surdeanu<sup>[29]</sup> 和 Min<sup>[33]</sup> 随机挑选 40 个查询作为训练,剩余的 160 个查询作为测试.

(4) FewRel 数据集 该数据集的主要目标是验证模型在少样本训练后在未知的实体对关系下是否具有泛化能力,其是由清华大学团队提出的关系抽取数据集<sup>[16]</sup>,该数据集使用基于远程监督的思想,通过与维基百科进行对齐. 不同于 NYT 的是,其包含自动对齐和人工标注两个步骤. 在自动对齐阶段,其首先过滤掉样本数少于 1000 的标签,避免长尾问题出现,其次去掉实体对重复的示例,使得每个实体对仅出现一次;在人工标注阶段,主要由标注人员进行筛选,保证语料的质量,最终保留了 100 个关系,每个关系对应 700 个示例. 由于每一个示例对应唯一的实体对,因此,总共有 70000 个实体对. FewRel 定义了 80 个关系为训练集 (其中 16 个关系为验证集), 20 个关系为测试集. 随后 FewRel 2.0<sup>[79]</sup> 版本则对第一版本进行了改进,使用 N-Way、K-Shot 设定,其表示给定模型  $N$  个未知的类型,每个类型包含  $K$  个示例,以此验证模型是否可以有效地分类. FewRel 已被多个工作作为验证模型的评测数据集,也是远程监督关系抽取的新的评测任务.

另外,国内也有相关研究团队构建了大规模中文关系抽取语料. 例如, Xu<sup>[80]</sup> 等人构建了可同时完成实体识别和关系分类的中文语料库;清华大学自然语言处理实验室开源了神经关系抽取工具包项目 OpenNRE<sup>[81]</sup>;国内最大的开源知识图谱社区 (OpenKG) 则收纳了超过 130 个开放中文知识图谱语料.

### 3.2 评测指标

在对远程监督关系抽取模型的评估方面, Mintz<sup>[10]</sup> 提出两种评估方法,分别是基于自动评估 (Held-out Evaluation) 和基于人工评估 (Manual Evaluation).

自动评估是指通过模型预测的结果自动与样本标签进行比较,通常选择的评测指标有准确率 (Precision)、召回率 (Recall)、 $F_\beta$  值和 P-R 曲线 (AUC 值) 等. 然而,由于远程监督方法构建的语料中,测试集中也普遍存在大量噪声,可能受到 False Positive 和 False Negative 的影响<sup>[17,37]</sup>,即在测试阶段给定的实体对被预测正确的关系,但由于错误的标注或知识库不充分等原因导致被判为预测错误,因此有时需要进行人工评估. 如今诸多工作同时兼顾两种评估方法. 表 4 对这些评测指标进行了整理.

#### 3.2.1 自动评估 (Held-out Evaluation)

在自动评估中,远程监督关系抽取属于多类分类任务,常用混淆矩阵 (Confusion Matrix) 来表示每个类预测的情况. 假设给定的数据集预定义的  $n$  个关系标签组成的集合为  $r = \{r_1, r_2, \dots, r_n\}$ , 混淆矩阵可以表示为  $n \times n$  的方阵,本文记作  $confmatrix \in \mathbb{R}^{n \times n}$ , 其中每一行表示样本的实际标签,每一

列表示模型预测的标签, 矩阵的元素  $confmatrix(i, j)$  则表示实际关系标签  $r_i$  对应的样本被预测为  $r_j$  的个数. 当且仅当  $i = j$  时表示预测正确. 可以得到相应的评测指标分别如下.

表 4 远程监督关系抽取评测指标

Tab. 4 Evaluation metrics of distant supervision relation extraction

评测指标	描述	功能
准确率(Precision)	指在测试集某个关系类上所有样本被预测正确的占比, 通常分为微平均和宏平均	评价关系预测的准确程度
召回率(Recall)	指在测试集上预测为某个关系类中正确的占比, 通常分为微平均和宏平均	评价关系预测的查全程度
$F_\beta$ 值	指准确率和召回率的综合评价, 公式为 $F_\beta = (1 + \beta^2) \frac{\text{Precision} \times \text{Recall}}{(\beta^2 \times \text{Precision}) + \text{Recall}}$	综合评价关系抽取在查准率和查全率方面的效果
P-R曲线	指以Recall为横轴、以Precision为纵轴的曲线	评价分类器的优劣性能
AUC值	ROC曲线与坐标轴包围部分的面积, $0 \leq AUC \leq 1$	评价分类器的优劣性能
$P@N$	通常表示按照准确率降序排序时第N(或N%)个值	避免False Negative对关系预测错误评估的影响
Hits@K	表示预测结果的前K个关系中如果存在真实标签则记为1, 否则记为0	评估在关系抽取基于相似度排序问题上的准确效果
MRR	值为所有排序位置对应倒数的和	评估在关系抽取基于相似度排序问题上的准确效果
Recall@K	表示在测试集上Hits@K指标的期望, 公式为 $\text{Recall}@K = \frac{1}{N} \sum_{i=1}^N \text{Hits}_i@K$	评估在关系抽取基于相似度排序问题上的查全效果

(1) 微平均准确率 (Micro-Avg Precision) 表示所有样本中被预测正确的占比. 表示为

$$\text{Precision (Micro - Avg)} = \frac{\sum_{i=1}^n confmatrix(i, i)}{\sum_{i=1}^n \sum_{j=1}^n confmatrix(j, i)}.$$

(2) 微平均召回率 (Micro-Avg Recall) 表示被预测为真的样本中正确的占比. 表示为

$$\text{Recall (Micro - Avg)} = \frac{\sum_{i=1}^n confmatrix(i, i)}{\sum_{i=1}^n \sum_{j=1}^n confmatrix(i, j)}.$$

(3) 宏平均准确率 (Macro-Avg Precision) 表示在每个关系标签上被预测正确的样本占所有样本比值的期望值, 其中每个类的准确率则为对角元素与其所在列之和的比值. 表示为

$$\text{Precision (Macro - Avg)} = \frac{1}{n} \sum_{i=1}^n \frac{confmatrix(i, i)}{\sum_{j=1}^n confmatrix(j, i)}.$$

(4) 宏平均召回率 (Macro-Avg Recall) 表示在每个关系标签上被预测正确的样本占真实样本比值的期望值, 其中每个类的召回率则为对角元素与其所在行之和的比值. 表示为

$$\text{Recall (Macro - Avg)} = \frac{1}{n} \sum_{i=1}^n \frac{confmatrix(i, i)}{\sum_{j=1}^n confmatrix(i, j)}.$$

因此可以得到相应的  $F_\beta$  值, 其是用来衡量二者之间重要性性能的指标, 公式为

$$F_\beta = (1 + \beta^2) \frac{\text{Precision} \times \text{Recall}}{(\beta^2 \times \text{Precision}) + \text{Recall}}.$$

其中  $\beta = 1$  时为  $F_1$  值:



$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

在远程监督关系抽取自动评估中, 研究者们通常取宏平均的准确率和召回率作为评价指标, 且不计入标签 NA 的结果.

为了能够对模型进行综合评测, 还引入了 Precision-Recall(P-R) 曲线和 AUC 值作为模型比对的指标. 其中 P-R 曲线用来衡量查准率和查全率的综合效果, 在多元分类中, 通过获得每个样本在各个类预测的概率分布及其对应独热编码, 并按照概率进行降序排序形成一个向量, 若对每个元素作为分界线, 则可以依次得到相应的查准率和查全率, 以此可以获得相应的曲线. P-R 曲线目前是远程监督关系抽取性能评估的主要指标. 另外, ROC 曲线是反映在关系抽取时敏感性 (FPR) 和精确性 (TPR) 的趋势, AUC 值 ( $0 \leq AUC \leq 1$ ) 表示 ROC 曲线与两坐标轴 (TPR 和 FPR) 包围的面积, AUC 值越大则说明模型的性能越好.

### 3.2.2 人工评估 (Manual Evaluation)

人工评估法是为了弥补由于知识库语料不充分所导致的一些 False Negative 样本对实验结果的影响, 因此人工评估方法也被广泛作为远程监督的评价依据. 目前人工评估的方法主要有如下几类.

(1) 直接重新从测试集中随机挑选一部分数据并进行人工标注, 确保每个示例都是正确无误的, 然后在此基础上验证预测的效果. 例如, Feng<sup>[48]</sup> 通过从测试集中随机选择了 300 个样本并人工标注了每个示例是否是噪声, 来验证他们的方法是否可以有效过滤噪声.

(2)  $P@N$  指标. 由于 P-R 曲线中随着 Recall 值的不断上升, 可能包含的 False Negative 数量会变大, 使得模型预测的正确的示例会被错误评判, 因此通常取一小部分示例用于测试. 一种方法是通过人工挑选  $N$  (或  $N\%$ ) 个示例; 另一种是直接按照 Precision 值从大到小排列并选择第  $N$  (或  $N\%$ ) 个值作为评估结果, 这种情况下,  $P@N$  指标相当于对模型预测的准确率最高的前  $N$  个 (或前  $N\%$ ) 示例中模型预测的准确率, 而可能含有大量的 False Negative 样本便被排除在外, 可以有效避免错误评判带来的问题. 通常大多数工作<sup>[17,36-37,53]</sup> 取值为  $N \in \{100, 200, 300, 500\}$  或  $N \in \{10\%, 20\%, 30\%, 50\%\}$ , 后来也有工作<sup>[22,82]</sup> 为了避免人工评估方法带来的时间消耗, 将  $P@N$  指标应用到自动评估方法中.

(3) One-Two-All. 远程监督语料中大多数的实体对共现次数仅为 1<sup>[36]</sup>, 在基于多示例学习方法中, 为了验证多示例学习方法在面对包内句子数量少的情况下是否能够有效处理噪声, 通常做法是在测试集中为每一个包只随机挑选 1 (One) 个或 2 (Two) 个句子, 为了做对比, 也将整个包 (All) 进行评估<sup>[36,82]</sup>. 该指标常与  $P@N$  结合使用.

### 3.2.3 其他指标

上述几种指标常被广泛用于评估模型在噪声语料中的关系抽取能力, 除此之外也有一些新的指标被用于具体评估模型在关系抽取中的某些性能.

(1) Hits@ $K$  通常被用于排序类问题, 如知识补全<sup>[73]</sup>、知识库问答<sup>[3]</sup>等. 其基本思路是通过一个得分函数对所有候选的结果进行排序, 并根据阈值选择得分最高的若干个结果, 为了评判预测结果的效果, 当且仅当正确结果出现在预测结果的 Top $K$  中才被认为预测正确. 在远程监督关系抽取任务中, 如果将关系抽取视为示例与关系标签的相似度匹配和排序问题, 则可以使用 Hits@ $K$  指标. 例如, Zhang<sup>[62]</sup> 和 Han<sup>[63]</sup> 通过利用 Hits@ $K$  ( $K$  的取值范围为  $\{10, 15, 20\}$ ) 指标来验证模型可以在长尾关系预测上得到不错的效果, Xu<sup>[52]</sup> 则利用 Hits@1 验证结合知识表示信息可以提升关系抽取性能.

(2) MRR 是另一种排序指标, 通常被用于搜索类问题. 在关系抽取中, 如果预测的关系排在第  $k$  位, 则其得分为  $1/k$ . 目前该指标在关系抽取中也有被应用<sup>[52]</sup> 但并不广泛.

(3) Recall@ $K$ . Hits@ $K$  指标通常只能表示准确程度, 但不能描述查全程度, Recall@ $K$  则可以被用

于评价查全率. 其计算公式如下<sup>[54]</sup>:

$$\text{Recall}@K = \frac{1}{N} \sum_{i=1}^N \text{Hits}_i@K.$$

其中  $N$  表示测试集大小. 例如, Liu<sup>[54]</sup> 等人则结合 Hits@ $K$  和 Recall@ $K$  来验证在实体类型约束下关系抽取的效果.

#### 4 研究挑战与未来趋势

基于远程监督的关系抽取作为一种可以快速生成大规模监督语料的方法一直以来备受关注, 而如今也形成了许多解决方法, 但是如今知识库语料不充分的原因使得噪声依然是“瓶颈”问题, 在今后的研究工作中还有很长的路要走. 基于如今的一些技术, 远程监督依然存在一些新的挑战.

(1) Overlapping 问题. 通常表现为两种, 一种是相同的实体对有多个关系, 如实体“Obama”与“US.”可能存在“President”“PlaceOfBirth”“PlaceOfLived”等关系; 另一种是相同的头实体(尾实体)和关系对应不同的尾实体(头实体), 即 one-to-many 或 many-to-many 的情况, 如“China”“Japan”“Thailand”均与实体“Asia”具有相同的关系“Located”. 而如今一些主流方法均只考虑了单一的关系, 即一个包内所有示例均只与某一个关系进行比较, 因此如何处理 Overlapping 是接下来关系抽取的一个目标.

(2) False Negative 问题. 目前大多数的降噪工作均是处理 False Positive, 然而由于知识库不全, 许多包含正确的语义关系的实体对却被错误地标注为 NA, 这在训练集和测试集中普遍存在. 而如今一些工作通过人工评估方法可以避免 False Negative 对评估的影响, 但事实上在训练集中却无法有效避免, 不论是以多示例学习还是以对抗与强化学习的方法, 均会导致模型学习到错误的映射关系, 从而影响对噪声的权重分配或选择过滤. 因此远程监督关系抽取面临的新的挑战则是如何直接处理 False Negative 示例.

(3) 长尾问题. 目前解决长尾问题的工作并不多, 解决的方案也局限于对长尾关系的信息补偿, 并没有直接改变数据集长尾分布. 而如何让长尾关系得到充分的学习, 一方面可以通过在特征提取后的语义层面上进行增强, 如迁移学习<sup>[82]</sup>等; 另一方面则是如何在数据集根源上缓解或改变不平衡的问题. 另外对于一些直接过滤噪声的方法<sup>[19,45,47]</sup>, 虽然进一步提高了语料的纯度, 但可能导致一些关系标签对应的语料数量大大降低, 从而加剧长尾问题. 因此如何兼顾考虑噪声和长尾两个问题也是一大挑战.

本文认为基于远程监督的关系抽取还有一些新的未来发展趋势.

(1) 端到端学习. 端到端学习在机器翻译(Machine Translation)、图像描述(Image Caption)等任务<sup>[83]</sup>中均有成功应用. 端到端学习旨在将 pipeline 任务合并为一个任务, 从而可以有效缓解不同任务之间存在的误差传播问题. 实体和关系进行联合抽取的工作相对不多, 因此如何在含噪的语料中结合实体识别任务直接提取出所有的三元组是接下来研究的趋势.

(2) 领域关系抽取. 目前基于远程监督的关系抽取的评测任务仍以开放领域为主, 而不同的领域内容会受到自身的业务场景影响, 使得关系抽取任务会变得复杂. 例如, 杨<sup>[12]</sup>事先构建小型的领域知识库及少量标注语料, 其次利用远程监督的思想对初等数学知识点关系进行标注. 由于知识点的结构复杂、粒度细、数量多等原因, 还需要专门设计必要的关系抽取系统实现具体的业务需求. 因此如何将远程监督关系抽取应用到医学、教育、金融等特定领域成为未来研究的重心.

(3) 对噪声样本的纠正. 关于如何降低噪声对训练的影响, 最近有一种新的思路, 即对噪声进行纠正. 其主要分为两个子步骤: 需要正确识别出噪声, 重新为这些样本分配标签. 对噪声的纠正既可以不

损失原有数据的规模,又可以提升语料的质量让模型学习到更准确的映射关系.例如,Shang等<sup>[84]</sup>人最近通过使用简单的阈值来过滤掉权重较低的句子作为噪声,并提出一种深度聚类算法对这些噪声重新分配标签,并设计损失函数用于同时训练分类器和聚类器.如何为噪声准确地分配新的标签目前还处在探索阶段,是将来研究的新目标.

(4) 多模态学习.对于如今丰富的数据类型,不再单独以自然语言呈现,更多的则是结合图像、语音等更复杂的数据结构.多模态学习目前已经成功地在情感分析、推荐系统、模式识别等领域<sup>[85]</sup>内发挥作用,因此如何在异构信息的指导下辅助增强实体关系抽取和降噪也是未来的研究趋势.

## 5 总 结

基于远程监督的关系抽取作为一项重要的自然语言处理任务,已经广泛应用到知识图谱构建与补全、知识库问答和文本摘要等领域.自被提出以来,因其具有可以快速高效地获得大规模监督语料的优势,一直以来吸引了众多研究者的注意.通过对近年来主要工作的调研,本文根据远程监督关系抽取的研究动机和难点将其划分为噪声、信息匮乏和非均衡三大挑战,并分别讨论了基于规则统计、多示例学习、对抗与强化学习、辅助信息增强、联合学习和少样本学习等技术路线.另外,本文介绍并详细对比了一些公开数据集和语料,同时系统梳理了目前常用的评估方法和评测指标.最后指出远程监督关系抽取任务未来要面对的新的挑战和发展趋势.

## [参 考 文 献]

- [1] 刘峤,李杨,段宏,等.知识图谱构建技术综述[J].计算机研究与发展,2016,53(3):582-600.
- [2] KEJRIWAL M, SEQUEDA J, LOPEZ V, et al. Knowledge graphs: Construction, management and querying: Editorial [J]. Social Work, 2019, 10(6): 961-962.
- [3] YU M, YIN W, HASAN K S, et al. Improved neural relation detection for knowledge base question answering [C]// Meeting of the Association for Computational Linguistics. 2017: 571-581.
- [4] ALLAHYARI M, POURIYEH S, ASSEFI M, et al. Text summarization techniques: A brief survey [J]. International Journal of Advanced Computer Science and Applications, 2017, 8(10): 397-405.
- [5] HASEGAWA T, SEKINE S, GRISHMAN R, et al. Discovering relations among named entities from large corpora [C]// Meeting of the Association for Computational Linguistics. 2004: 415-422.
- [6] ETZIONI O, BANKO M, SODERLAND S, et al. Open information extraction from the web [J]. Communications of the ACM, 2008, 51(12): 68-74.
- [7] LI F, ZHANG M, FU G, et al. A Bi-LSTM-RNN model for relation classification using low-cost sequence features[J]. ArXiv: Computation and Language, 2016.
- [8] 姚春华,刘潇,高弘毅,等.基于句法语义特征的实体关系抽取技术[J].通信技术,2018,51(8):1828-1835.
- [9] KUMLIEN M C J. Constructing biological knowledge bases by extraction information from text sources [C]// Proc Int Conf Intell Syst Mol Biol. 1999: 77-86.
- [10] MINTZ M, BILLS S, SNOW R, et al. Distant supervision for relation extraction without labeled data [C]// International Joint Conference on Natural Language Processing. 2009: 1003-1011.
- [11] ZENG X, HE S, LIU K, et al. Large scaled relation extraction with reinforcement learning [C]// National Conference on Artificial Intelligence. 2018: 5658-5665.
- [12] 杨东明,杨大为,顾航,等.面向初等数学的知识点关系提取研究[J].华东师范大学学报(自然科学版),2019(5):53-65.
- [13] RIEDEL S, YAO L, MCCALLUM A, et al. Modeling relations and their mentions without labeled text [C]// European Conference on Machine Learning. 2010: 148-163.
- [14] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: A collaboratively created graph database for structuring human knowledge [C]// International Conference on Management of Data. 2008: 1247-1250.
- [15] JAT S, KHANDELWAL S, TALUKDAR P P, et al. Improving distantly supervised relation extraction using word and entity based attention [J]. ArXiv: Computation and Language, 2018.
- [16] HAN X, ZHU H, YU P, et al. FewRel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation [C]// Empirical Methods in Natural Language Processing. 2018: 4803-4809.

- [17] ZENG D, LIU K, CHEN Y, et al. Distant supervision for relation extraction via piecewise convolutional neural networks [C]// Empirical Methods in Natural Language Processing. 2015: 1753-1762.
- [18] ZELENKO D, AONE C, RICARDELLA A, et al. Kernel methods for relation extraction [J]. Journal of Machine Learning Research, 2003, 3(6): 1083-1106.
- [19] SHI G, FENG C, HUANG L, et al. Genre separation network with adversarial training for cross-genre relation extraction [C]// Empirical Methods in Natural Language Processing. 2018: 1018-1023.
- [20] VASHISHTH S, JOSHI R, PRAYAGA S S, et al. RESIDE: Improving distantly-supervised neural relation extraction using side information [C]// Empirical Methods in Natural Language Processing. 2018: 1257-1266.
- [21] LI Y, LONG G, SHEN T, et al. Self-attention enhanced selective gate with entity-aware embedding for distantly supervised relation extraction [C]// National Conference on Artificial Intelligence. 2020.
- [22] KUANG J, CAO Y, ZHENG J, et al. Improving neural relation extraction with implicit mutual relations [C]// International Conference on Data Engineering. 2020.
- [23] KRAUSE S, LI H, USZKOREIT H, et al. Large-scale learning of relation-extraction rules with distant supervision from the web [C]// International Semantic Web Conference. 2012: 263-278.
- [24] 白龙, 靳小龙, 席鹏弼, 等. 基于远程监督的关系抽取研究综述 [J]. 中文信息学报, 2019, 33(10): 10-17.
- [25] 鄂海红, 张文静, 肖思琪, 等. 深度学习实体关系抽取研究综述 [J]. 软件学报, 2019, 30(6): 1793-1818.
- [26] SUCHANEK F M, KASNECI G, WEIKUM G, et al. Yago: A core of semantic knowledge [C]// The Web Conference. 2007: 697-706.
- [27] ZHOU P, SHI W, TIAN J, et al. Attention-based bidirectional long short-term memory networks for relation classification [C]// Meeting of the Association for Computational Linguistics. 2016: 207-212.
- [28] HOFFMANN R, ZHANG C, LING X, et al. Knowledge-based weak supervision for information extraction of overlapping relations [C]// Meeting of the Association for Computational Linguistics. 2011: 541-550.
- [29] SURDEANU M, TIBSHIRANI J, NALLAPATI R, et al. Multi-instance multi-label learning for relation extraction [C]// Empirical Methods in Natural Language Processing. 2012: 455-465.
- [30] TAKAMATSU S, SATO I, NAKAGAWA H, et al. Reducing wrong labels in distant supervision for relation extraction [C]// Meeting of the Association for Computational Linguistics. 2012: 721-729.
- [31] FAN M, ZHAO D, ZHOU Q, et al. Distant supervision for relation extraction with matrix completion [C]// Meeting of the Association for Computational Linguistics. 2014: 839-849.
- [32] ZHANG Q, WANG H. Noise-clustered distant supervision for relation extraction: A nonparametric bayesian perspective [C]// Empirical Methods in Natural Language Processing. 2017: 1808-1813.
- [33] MIN B, GRISHMAN R, WAN L, et al. Distant supervision for relation extraction with an incomplete knowledge base [C]// North American Chapter of the Association for Computational Linguistics. 2013: 777-782.
- [34] XU W, HOFFMANN R, ZHAO L, et al. Filling knowledge base gaps for distant supervision of relation extraction [C]// Meeting of the Association for Computational Linguistics. 2013: 665-670.
- [35] RITTER A, ZETTLEMOYER L, ETZIONI O, et al. Modeling missing data in distant supervision for information extraction [C]// Transactions of the Association for Computational Linguistics. 2013: 367-378.
- [36] LIN Y, SHEN S, LIU Z, et al. Neural relation extraction with selective attention over instances [C]// Meeting of the Association for Computational Linguistics. 2016: 2124-2133.
- [37] JI G, LIU K, HE S, et al. Distant supervision for relation extraction with sentence-level attention and entity descriptions [C]// National Conference on Artificial Intelligence. 2017: 3060-3066.
- [38] JAT S, KHANDELWAL S, TALUKDAR P P, et al. Improving distantly supervised relation extraction using word and entity based attention [J]. ArXiv: Computation and Language, 2018.
- [39] WU S, FAN K, ZHANG Q, et al. Improving distantly supervised relation extraction with neural noise converter and conditional optimal selector [J]. National Conference on Artificial Intelligence, 2019, 33(1): 7273-7280.
- [40] YE Z, LING Z. Distant supervision relation extraction with intra-bag and inter-bag attentions [C]// North American Chapter of the Association for Computational Linguistics. 2019: 2810-2819.
- [41] YUAN Y, LIU L, TANG S, et al. Cross-relation cross-bag attention for distantly-supervised relation extraction [J]. National Conference on Artificial Intelligence, 2019, 33(1): 419-426.
- [42] JIA W, DAI D, XIAO X, et al. ARNOR: Attention regularization based noise reduction for distant supervision relation classification [C]// Meeting of the Association for Computational Linguistics. 2019: 1399-1408.
- [43] ALT C, HUBNER M, HENNIG L, et al. Fine-tuning pre-trained transformer language models to distantly supervised relation extraction [C]// Meeting of the Association for Computational Linguistics. 2019: 1388-1398.
- [44] WU Y, BAMMAN D, RUSSELL S, et al. Adversarial training for relation extraction [C]// Empirical Methods in Natural Language Processing. 2017: 1778-1783.
- [45] QIN P, WEIRAN X U, WANG W Y, et al. DSGAN: Generative adversarial training for robust distant supervision relation extraction [C]// Meeting of the Association for Computational Linguistics. 2018: 496-505.



- [46] LI P, ZHANG X, JIA W, et al. GAN driven semi-distant supervision for relation extraction [C]// North American Chapter of the Association for Computational Linguistics. 2019: 3026-3035.
- [47] HAN X, LIU Z, SUN M, et al. Denoising distant supervision for relation extraction via instance-level adversarial training [J]. ArXiv: Computation and Language, 2018.
- [48] FENG J, HUANG M, ZHAO L, et al. Reinforcement learning for relation classification from noisy data [C]// National Conference on Artificial Intelligence. 2018: 5779-5786.
- [49] HE Z, CHEN W, WANG Y, et al. Improving neural relation extraction with positive and unlabeled learning [C]// National Conference on Artificial Intelligence. 2020.
- [50] QIN P, XU W, WANG W Y, et al. Robust distant supervision relation extraction via deep reinforcement learning [C]// Meeting of the Association for Computational Linguistics. 2018: 2137-2147.
- [51] SU Y, LIU H, YAVUZ S, et al. Global relation embedding for relation extraction [C]// North American Chapter of the Association for Computational Linguistics. 2018: 820-830.
- [52] XU P, BARBOSA D. Investigations on knowledge base embedding for relation prediction and extraction [J]. ArXiv: Computation and Language, 2018.
- [53] XU P, BARBOSA D. Connecting language and knowledge with heterogeneous representations for neural relation extraction [C]// North American Chapter of the Association for Computational Linguistics. 2019: 3201-3206.
- [54] LIU Y, LIU K, XU L, et al. Exploring fine-grained entity type constraints for distantly supervised relation extraction [C]// International Conference on Computational Linguistics. 2014: 2107-2116.
- [55] YE Y, FENG Y, LUO B, et al. Integrating relation constraints with neural relation extractors [C]// National Conference on Artificial Intelligence. 2020.
- [56] BELTAGY I, LO K, AMMAR W, et al. Combining distant and direct supervision for neural relation extraction [C]// North American Chapter of the Association for Computational Linguistics. 2019: 1858-1867.
- [57] WEI Z, SU J, WANG Y, et al. A novel hierarchical binary tagging framework for joint extraction of entities and relations [J]. ArXiv: Computation and Language, 2019.
- [58] REN X, WU Z, HE W, et al. CoType: Joint extraction of typed entities and relations with knowledge bases [C]// The Web Conference. 2017: 1015-1024.
- [59] TAKANOBU R, ZHANG T, LIU J, et al. A hierarchical framework for relation extraction with reinforcement learning [J]. National Conference on Artificial Intelligence, 2019, 33(1): 7072-7079.
- [60] YE W, LI B, XIE R, et al. Exploiting entity BIO tag embeddings and multi-task learning for relation extraction with imbalanced data [C]// Meeting of the Association for Computational Linguistics. 2019: 1351-1360.
- [61] GUI Y, LIU Q, ZHU M, et al. Exploring long tail data in distantly supervised relation extraction [C]// LIN C Y, XUE N, ZHAO D, et al. Natural Language Understanding and Intelligent Applications. ICCPOL 2016, NLPCC 2016. Lecture Notes in Computer Science, 2016.
- [62] ZHANG N, DENG S, SUN Z, et al. Long-tail relation extraction via knowledge graph embeddings and graph convolution networks [C]// North American Chapter of the Association for Computational Linguistics. 2019: 3016-3025.
- [63] HAN X, YU P, LIU Z, et al. Hierarchical relation extraction with coarse-to-fine grained attention [C]// Empirical Methods in Natural Language Processing. 2018: 2236-2245.
- [64] MIKOLOV T, CHEN K, CORRADO G S, et al. Efficient estimation of word representations in vector space [C]// International Conference on Learning Representations. 2013.
- [65] PENNINGTON J, SOCHER R, MANNING C D, et al. Glove: Global vectors for word representation [C]// Empirical Methods in Natural Language Processing. 2014: 1532-1543.
- [66] DEVLIN J, CHANG M, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C]// North American Chapter of the Association for Computational Linguistics. 2019: 4171-4186.
- [67] GOODFELLOW I, POUGETABADIE J, MIRZA M, et al. Generative adversarial nets [C]// Neural Information Processing Systems. 2014: 2672-2680.
- [68] SALVARIS M, DEAN D, TOK W H, et al. Generative adversarial networks [J]. ArXiv: Machine Learning, 2018: 187-208.
- [69] ANDREW A M. Reinforcement learning: An introduction [J]. Kybernetes, 1998, 27(9): 1093-1096.
- [70] SUN T, ZHANG C, JI Y, et al. Reinforcement learning for distantly supervised relation extraction [J]. IEEE Access, 2019(7): 98023-98033.
- [71] TANG J, QU M, WANG M, et al. LINE: Large-scale information network embedding [C]// The Web Conference. 2015: 1067-1077.
- [72] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [73] BORDES A, USUNIER N, GARCIADURAN A, et al. Translating embeddings for modeling multi-relational data [C]// Neural Information Processing Systems. 2013: 2787-2795.
- [74] KIPF T, WELLMING M. Semi-supervised classification with graph convolutional networks [C]// International Conference on Learning Representations. 2017.

- [75] HENDRICKX I, KIM S N, KOZAREVA Z, et al. SemEval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals [C]// North American Chapter of the Association for Computational Linguistics. 2009: 94-99.
- [76] SURDEANU M, GUPTA S, BAUER J, et al. Stanford's distantly-supervised slot-filling system [R]. Stanford, CA: Stanford University, 2011.
- [77] JI, HENG, GRISHMAN, RALPH, et al. Overview of the TAC 2010 knowledge base population track [C]// Text Analysis Conference. 2009.
- [78] JI H, GRISHMAN R, DANG H. Overview of the TAC2011 knowledge base population track [C]// Text Analysis Conference. 2011.
- [79] GAO T, HAN X, ZHU H, et al. FewRel 2.0: Towards more challenging few-shot relation classification [C]// International Joint Conference on Natural Language Processing. 2019: 6249-6254.
- [80] XU J, WEN J, SUN X, et al. A discourse-level named entity recognition and relation extraction dataset for Chinese literature text [J]. ArXiv: Computation and Language, 2017.
- [81] HAN X, GAO T, YAO Y, et al. OpenNRE: An open and extensible toolkit for neural relation extraction [C]// International Joint Conference on Natural Language Processing. 2019: 169-174.
- [82] LIU T, ZHANG X, ZHOU W, et al. Neural relation extraction via inner-sentence noise reduction and transfer learning [C]// Empirical Methods in Natural Language Processing. 2018: 2195-2204.
- [83] REN Z, WANG X, ZHANG N, et al. Deep reinforcement learning-based image captioning with embedding reward [C]// Computer Vision and Pattern Recognition. 2017: 1151-1159.
- [84] SHANG Y M, HUANG H, MAO X, et al. Are noisy sentences useless for distant supervised relation extraction [C]// National Conference on Artificial Intelligence. 2020.
- [85] CAO Z, HIDALGO G, SIMON T, et al. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields [J]. ArXiv: Computer Vision and Pattern Recognition, 2018.

(责任编辑: 李万会)