面向远程监督数据联合抽取的不确定性感知自举学习

Yufei Li[1], **Xiao Yu**[2], **Yanchi Liu**[3], **HaifengChen**[3], **Cong Liu**[1]1加利福尼亚大学河滨分校 [2]Stellar Cyber [3]NEC Labs America[1]{yli927,cong}@ucr.edu, [2]xyu@stellarcyber.ai,[3]{yanchi,haifeng}@nec-labs.com

arXiv:2305.03827v2 [cs.CL] 8 Jun 2023

摘要

在处理带有歧义或噪声标签的远程监督数据时，同时抽取实体对及其关系颇具挑战。为减轻此类影响，我们提出不确定性感知自举学习，其基本直觉是：实例的不确定性越高，模型置信度就越可能与真实标签不一致。具体而言，我们首先研究实例级数据不确定性，以构建一组初始的高置信度示例。该子集可用于过滤噪声实例，并帮助模型在早期阶段快速收敛。在自举学习过程中，我们提出将自集成作为正则化项，以缓解噪声标签产生的模型间不确定性。此外，我们定义联合标注概率的概率方差，以估计模型内部的参数不确定性，并利用该不确定性选择和构建新的可靠训练实例，用于下一轮迭代。在两个大型数据集上的实验结果表明，我们的方法优于现有的强基线和相关方法。

1 引言

联合抽取是指使用单一模型抽取多种类型的实体及其相互之间的关系，这对于自动构建知识库至关重要（Yu 等人，2020）。一种低成本获取大量标注数据以训练联合抽取模型的方法是采用远程监督（DS）（Mintz 等人，2009年）。远程监督通过手工制定的规则或逻辑约束，将知识库（KB）与未标注语料库对齐。由于缺乏人工标注者，远程监督会产生大量噪声标签，例如某些情况下噪声实例占比超过30%（Mintz 等人，2009年），导致模型无法学习到有用的特征。这些噪声可能是由上述基于规则的匹配假设造成的错误关系，也可能是由于开放领域知识库对实体的覆盖有限而产生的错误实体标签。

现有的远程监督方法主要通过启发式方法（例如强化学习（RL）（Nooralahzadeh 等人，2019；Hu 等人，2021年）和对抗学习（Chen 等人，2021年），或基于模式的方法（Jia 等人，2019；Shang 等人，2022）来筛选可信实例，从而对噪声建模。然而，这些方法需要设计启发式规则或手工编写模式，这可能会促使模型利用虚假特征，而不考虑其预测的置信度或不确定性。

针对这些问题，我们提出了**UnBED——不确定性感知的**自举学习，用于远程监督数据上的联合**抽**取。UnBED 假设：1）在初始阶段，使用预训练语言模型（PLM）时，较低的数据不确定性表明实例具有较高的可靠性；2）经过训练后，模型应能够根据自身的不确定性识别可信的实体和关系标签。我们将不确定性作为自举学习的原则，以减轻噪声标签对模型学习的影响，并验证输入序列，从而控制每一步训练样本的数量。具体而言，我们根据实例的获胜分数（Hendrycks 和 Gimpel，2017年）和熵（Shannon，1948）来量化数据不确定性。我们定义了由联合预训练语言模型对序列中每个词元进行估计的平均最大概率，以使先前的技术适用于联合抽取方案。我们收集数据不确定性较低的实例，构建一个初始子集，并利用该子集调整联合预训练语言模型标注器，从而加快收敛速度。随后，我们从两个角度定义参数不确定性：模型间不确定性和模型内不确定性。前者通过自集成（Wang 和 Wang，2022）进行量化，并作为正则化项，以提高模型在训练过程中抵御噪声标签的鲁棒性。后者则通过蒙特卡洛 Dropout（Gal 和 Ghahramani，2016）中的概率方差进行捕获，用于选择新的高置信度实例，以供下一轮训练迭代使用。这两种

折叠模型不确定性相互强化，引导模型逐步提升其鲁棒性，并从可靠知识中学习。

2 相关工作

联合抽取方法 联合抽取使用单个模型检测实体及其关系，有效整合了两类信息，因此，与流水线方法相比，在两个子任务上都取得了更好的结果（Zheng等人，2017年）。例如，统一方法会同时标注实体和关系；比如（Zheng等人，2017年）提出了一种新颖的标注方案，将联合抽取转化为序列标注问题；（Dai等人，2019）引入查询位置和序贯标注来抽取重叠关系。与参数共享神经模型（Miwa和Bansal，2016；Gupta等人，2016）相比，此类方法避免产生冗余信息；同时，无需结构化系统中使用的手工设计特征（Yu等人，2020）。

为了解决从远程监督数据中学习的挑战，预训练变换器（例如 BERT、GPT-2）受到了广泛关注。它们通过多个注意力层为文本序列构建具有强大表达能力的上下文感知表示，并在各种自然语言处理任务中取得了最先进的性能（Radford 等人，2019；Devlin 等人，2019；Li 等人，2022）。它们可以通过低成本微调来解决包括命名实体识别和关系分类在内的不同下游任务。具体而言，BERT 使用掩码语言建模在大规模英语语料库上进行训练。多头注意力权重表示每对词语之间的交互关系，而其隐藏状态则整合了整个句子的语义信息，并用于解码不同的标注结果。

不确定性方法 不确定性通常来自两个来源——偶然不确定性和认知不确定性。前者也称为数据不确定性，用于描述数据生成过程中固有的噪声。缓解此类不确定性的方法包括数据插值（Dong 等人，2018）、获胜分数和温度缩放（Guo 等人，2017年）。后者也称为模型不确定性，用于描述结构选择和模型参数是否能最佳地描述数据分布。缓解模型不确定性的一种主要方法是贝叶斯神经网络（BNN）（Klein 等人，2017年），它为其权重设置先验分布。为了节省计算成本，蒙特卡罗

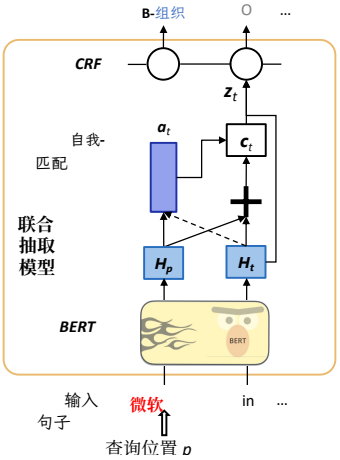

图1：作为词元分类任务的联合抽取。

dropout 被提出作为变分贝叶斯推断的近似（Gal 和 Ghahramani，2016），其通过使用丢弃层训练模型，并采用随机推断进行测试来量化概率方差。除了贝叶斯神经网络之外，集成学习（Wang 和 Wang，2022）通过测量具有相同架构的模型之间的输出方差，也已被证明能够有效降低模型间的参数不确定性。

3 联合抽取模型

标注方案 对于输入序列 $\mathcal{X}=\{x_1,...,x_n\}$，我们根据不同的查询位置 n 对序列进行标注，具体遵循（Dai等人，2019）的方法。如果 p 是实体（查询实体 e_1）的起始位置，则该序列是一个实例。在 p 处标注实体类型，与查询实体 e_2 存在关系的其他实体则标注为关系类型re。其余词元标注为“O”（外部），表示它们不对应查询实体。因此，我们将联合抽取转换为词元分类任务，并在每个实例中抽取关系三元组$\{e_1, re, e_2\}$，如图 1所示。

位置注意力编码器我们使用 BERT（Devlin 等人，2019）将句子编码为 $\mathcal{X}$ 词元级表示 $\boldsymbol{h}=\{\boldsymbol{h}_1,..,\boldsymbol{h}_n\}$，其中 $\boldsymbol{h}_i \in \mathbb{R}^d$ 是一个 d维向量，对应于i中第 $\mathcal{X}$个词元。对于每个查询 p，我们应用自匹配来计算位置注意力 $\boldsymbol{a}_t \in \mathbb{R}^T$，衡量位于 p 的词元与目标位置 t处各词元之间的关系。该方法将句子表示与自身进行比较，以收集上下文信息（Tan等人，2018）。生成的位置感知表示 $\boldsymbol{c}_t \in \mathbb{R}^{T\times d}$ 是经过注意力加权的句子向量 $\boldsymbol{c}_t = \boldsymbol{a}_t^\top \boldsymbol{h}$。最后，我们将$\boldsymbol{h}_t$ 和 $\boldsymbol{c}_t$ 拼接起来，生成位置感知和上下文

感知表示 $\boldsymbol{u}_t = [\boldsymbol{h}_t|\boldsymbol{c}_t]$。

CRF解码器（Lafferty等人，2001年）对于每个位置感知表示 $\boldsymbol{u}_t$，我们首先学习一个线性变换 $\boldsymbol{z}_t = \boldsymbol{W}\boldsymbol{u}_t \in \mathbb{R}^C$，以表示第 t个词元的标签分数。这里 C 是不同标签的数量。对于标签为 $\boldsymbol{y} = \{y_1, ..., y_n\}$的实例，解码分数 $s(\boldsymbol{z}, \boldsymbol{y})$是从标签 y_t 到标签y_{t+1} 的转移分数与输入分数 $z_t^{y_t}$之和。条件概率 $p(\boldsymbol{y}|\boldsymbol{z})$是针对所有可能标签序列 $\boldsymbol{y}'$，对 $s(\boldsymbol{z}, \boldsymbol{y})$进行Softmax得到的。训练过程中，我们最大化正确标签序列的对数似然$\mathcal{L}_c = \sum\log\ p(\boldsymbol{y}|\boldsymbol{z})$。

4 不确定性感知自举学习

研究动机 自举学习面临的主要挑战之一是评估已标注实例的“正确性”。我们从不确定性的角度考察这一问题，并假设不确定性较低的实例更有可能被正确标注。本节首先提出实例级数据不确定性，用于筛选噪声样本并构建初始子集。随后，我们介绍两类模型不确定性，它们有助于在自举学习过程中逐步减轻远程监督效应，并不断构建可信样本。

4.1 数据不确定性

在不同训练阶段按由易到难的顺序呈现示例有助于模型（Platanios等人，2019；Zhou等人，2020）。我们提出数据不确定性，用以量化实例的“难度”。为更好地估计数据不确定性，我们使用预训练语言模型（PLMs）为序列中的每个词元生成标签概率。我们的直觉是，不确定性更高的输入更难由预训练语言模型生成，因为它已具备语言方面的合理性依据。因此，我们提出两种数据不确定性，可单独使用，也可结合使用：

获胜分数（WS）最大Softmax概率反映了输入的数据不确定性（Hendrycksand Gimpel，2017年）。给定一个输入实例 $\mathcal{I} = \{x_1, ..., x_n\}$，我们将数据不确定性定义为 u^d（$\mathcal{I}$）词元分类获胜分数取负后的平均值：

$$u^d(\mathcal{I}) = -\frac{1}{n}\sum_{t=1}^{n} \max_{c\in[1,C]} P(y_t = c|x_t) \quad (1)$$

熵 香农熵（Shannon，1948）被广泛用于反映信息不确定性。我们提出数据不确定性 $u^d(\mathcal{I})$，将词元分类熵取平均：

$$u^d(\mathcal{I}) = \frac{1}{n}\sum_{t=1}^{n}\sum_{c=1}^{C} P(y_t = c|x_t)\log P(y_t = c|x_t) \quad (2)$$

我们过滤掉不确定性得分较高的样本，并用“简单”样本构建一个初始子集。在训练早期，模型尚不清楚合理的分布 $P(y|x)$ 应该是什么，因此，数据不确定性有助于模型通过在相对“简单”的子集上进行调优而快速收敛。

4.2 模型不确定性

在自举学习中，我们定义模型不确定性，即认知不确定性（Kendall和Gal，2017年），以衡量模型参数能否最好地描述数据分布，遵循（Zhou等人，2020）。较低的模型不确定性表明，模型确信当前训练数据已得到充分学习（Wang等人，2019）。我们采用蒙特卡洛Dropout（Gal和Ghahramani，2016）来近似贝叶斯推断，从而捕捉模型内部的参数不确定性。具体而言，我们对联合模型进行 K 次前向传播。在每次传播中，网络中的部分神经元会被随机禁用。最后，我们得到ˆ ˆK 个关于模型参数的样本 {θ_1，.....，θ_K}。我们使用所有标签上的平均词元分类**概率方差（PV）**（Shelmanov等人，2021年）作为实例 $\mathcal{I}$：

$$u^m(\theta) = \frac{1}{n}\sum_{t=1}^{n}\sum_{c=1}^{C} \mathrm{Var}\left[P(y_t = c|x_t, \hat{\theta}_k)\right]_{k=1}^{K} \quad (3)$$

其中 Var[.] 表示在以下分布上的方差 K 遍历次数，遵循（Dong 等人，2018；Xiao 和 Wang，2019 ）。因此，模型能够感知每个实例的置信度，以及标签带有噪声的可能性。

4.3 训练策略

不确定性感知损失 除了用于衡量模型内参数不确定性的蒙特卡洛 Dropout，我们还考虑减轻模型间的参数不确定性，以稳定训练过程中的权重。具体而言，我们采用自集成（He 等人，2020；Wang 和 Wang，2022）来计算相同模型之间的损失，从而提高模型鲁棒性，并减少标签噪声对模型性能的影响。

算法 1自举学习

输入： 原始数据集 $\mathcal{D} = \{(\mathcal{I}^n, y^n)\}_{n=1}^N$，两个联合模型 f_1, f_2，参数为 θ_1, θ_2；

1：计算每个样本的数据不确定性 $u^d(\mathcal{I})$，实例 $\mathcal{I}$ 位于 $\mathcal{D}$；

2：初始数据集 $\mathcal{C} \leftarrow$ 选择数据对 $(\mathcal{I}^n, y^n)$ 使得 $u^d(\mathcal{I}) < \tau^d$ 来自 $\mathcal{D}$；

3：**对**轮次 $e = 1$，..... **执行**。

4：　　训练 f_1，f_2 在 $\mathcal{C}$ 使用公式（5）时；

5：　　计算 $u^m(\theta_1)$ 上的模型不确定性 $\mathcal{D}$；

6：　　$\mathcal{C} \leftarrow$ 选择数据对 $(\mathcal{I}^n, y^n)$，使得 $u^m(\mathcal{I}; \theta_1) < \tau^m$ 来自 $\mathcal{D}$；

我们创建另一个具有相同框架（例如架构、损失函数和超参数）的联合模型，并计算自集成损失$\mathcal{L}_e$，以最小化两个模型针对相同输入所产生的输出之间的差异：

$$\mathcal{L}_e = \sum KL(f(\mathcal{I}; \theta_1), f(\mathcal{I}; \theta_2)) \quad (4)$$

其中，$KL(.)$表示两个概率分布之间的 Kullback-Leibler 散度，θ_1、θ_2 分别表示第一个模型和第二个模型的参数。我们将最终的感知不确定性目标函数 $\mathcal{L}$定义为条件随机场损失与自集成损失之和：

$$\mathcal{L} = \mathcal{L}_c + \alpha\mathcal{L}_e \quad (5)$$

其中，α 表示自集成的权重，$\mathcal{L}_c$ 表示词元分类损失。

自举学习过程 为减轻DS效应对模型性能的影响，我们提出了一种两折自举学习策略（见算法 1）。具体而言，我们首先利用数据不确定性筛除“较难”的示例，并重新分配可靠的初始训练数据 $\mathcal{M}$。随后，我们按照由易到难的顺序，将示例迭代输入模型。在每次训练迭代中，我们通过自集成损失对联合模型进行正则化，以减少噪声标签对模型参数的影响。然后，我们利用概率方差选择新的置信训练实例 $\mathcal{D}'$，这些实例可由模型解释，并作为下一轮训练输入。选取的确定性示例越多，模型就越有可能学到有益信息并更快收敛。我们重复上述过程，直到验证集上的F1分数收敛。

5 实验

5.1 设置

我们在两个数据集NYT和Wiki-KBP上评估UnBED的性能。NYT（Riedel等，2010年）数据集收集了《纽约时报》的新闻，其训练数据由远程监督自动标注。我们采用经修订的测试数据集（Jia等，2019），该数据集经过人工标注以确保质量。Wiki-KBP（Ling和Weld，2012）数据集收集了维基百科的文章。其训练数据由远程监督标注（Liu等，2017年），测试集则经过人工标注（Ellis等，2013）。

我们将 UnBED 与以下基线方法进行比较：**ARNOR** (Jia 等人，2019)，这是一种基于模式的方法，用于减少远程监督三元组抽取中的噪声。**PURE** (Zhong 和 Chen，2021年)，这是一种流水线方法，使用预训练的 BERT 实体模型首先识别实体，然后利用关系模型检测其潜在关系。**FAN** (Hao 等人，2021年)，这是一种对抗方法，包含一个 Transformer 编码器，用于减少远程监督三元组抽取中的噪声。

评估 我们根据精确率（Prec.）、召回率（Rec.）和 F1 值评估每个句子抽取出的三元组。如果关系类型 re 和两个实体 e_1、e_2 均正确，则三元组 $\{e_1, re, e_2\}$ 判定为正确。我们从测试集中随机抽取 10% 的句子，构建验证集。

实现细节我们使用 Hugging Face*bert-large-uncased* (Devlin等人，2019) 预训练模型作为骨干网络。对于 ARNOR，隐藏向量的维度设置为 300。在正则化训练中，我们发现两个数据集的最优参数 α 均为 1。我们使用 PyTorch 实现 UnBED 和所有基线模型，采用 Adam 优化器，初始学习率为 10^{-5}，dropout 率为 0.1，批量大小为 8。对于初始子集配置，我们选择 0.5 作为数据不确定性阈值。对于自举学习，结合最佳验证集 F1 分数，将模型不确定性阈值设为 0.6。

5.2 总体结果

如表1所示，UnBED 在精确率和 F1 指标上显著优于所有基线。具体而言，与去噪方法 ARNOR 和 FAN 相比，UnBED 在 NYT 上的 F1 提升了 8%（在 Wiki-KBP 上提升了 3%）。我们的方法也优于采用预训练 Transformer 的基线方法（PURE 和 FAN），表明不确定性感知自举学习能够有效降低噪声标签的影响。

方法	NYT			Wiki-KBP		
	精确率	Rec.	F1	精确率	Rec.	F1
ARNOR（Jia 等人，2019）	0.588	0.614	0.600	0.402	0.471	0.434
PURE（Zhong 和 Chen，2021年）	0.536	0.664	0.593	0.395	0.433	0.413
FAN（Hao 等人，2021年）	0.579	0.646	0.611	0.391	0.467	0.426
UnBED-WS	**0.662**	0.730	**0.694**	**0.429**	0.501	**0.462**
UnBED-Entropy	0.651	**0.741**	0.693	0.422	**0.509**	0.461

表 1：在 NYT 和 Wiki-KBP 数据集上的评估结果。**粗体**数字表示最佳指标。UnBED-WS 和 UnBED-Entropy 分别表示将获胜分数和熵作为数据不确定性的 UnBED。

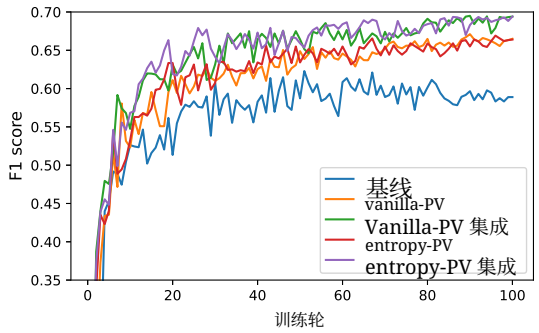

图 2：不同设置下 F1 分数与训练轮数的关系。Vanilla-PV 集成表示 UnBED-WS，entropy-PV 集成表示 UnBED-Entropy。

5.3 进一步分析

我们在图2中分析了不同组件的作用。我们发现，熵-PV 和 vanilla-PV 在 F1 方面均优于基线（直接在原始远程监督数据集上训练的联合模型）（5∼7% 的提升），这证明了过滤噪声标签并利用概率方差筛选可信实例的效果。此外，自集成进一步提升了训练后期的性能（2∼4F1 提升），证明缓解模型间不确定性有助于增强模型抵御噪声标签的鲁棒性。

6 结论

本文提出了一种面向远程监督联合抽取的新型不确定性感知自举学习框架。具体而言，我们在一般的词元分类中定义数据不确定性，以筛除错误率较高的实例并构建初始高置信度子集，用于调优联合抽取模型，从而实现快速收敛。随后，我们提出了一种两阶段自举学习过程，通过迭代减轻远程监督影响对模型鲁棒性的影响，并选择新的可信训练实例。在两个基准数据集上的实验结果表明，UnBED 显著优于其他去噪技术，实现了更好的效果。

局限性

本文提出了一种用于联合抽取的不确定性感知自举学习框架。尽管与其他去噪技术相比，该框架取得了最先进的性能，但考虑到集成损失是在两个大型预训练语言模型之间计算的，而概率方差是在预训练语言模型联合抽取模型上计算的，UnBED 需要大量训练资源。在未来的工作中，我们希望在训练过程中引入剪枝技术，以提高效率。我们还将考虑实体之间更复杂的关系，例如句子边界之外的关系，以适应现实世界的信息抽取场景。

致谢

本研究得到美国国家科学基金会 CNS 2135625、CPS 2038727、CNS Career 1750263 以及美国国防高级研究计划局壳牌公司的资助。

参考文献

Tao Chen、Haochen Shi、Liyuan Liu、Siliang Tang、Jian Shao、Zhigang Chen 和 YuetingZhuang。2021年。通过协同对抗训练增强远程监督关系抽取。载于第三十五届 *AAAI*人工智能大会（*AAAI 2021*）、第三十三届创新人工智能应用大会（*IAAI 2021*）、第十一届人工智能教育进展研讨会（*EAAI 2021*），线上活动，*2021*年*2*月*2*日至*9*日，第12675–12682页。AAAI出版社。

Dai Dai, Xinyan Xiao, Yajuan Lyu, Shan Dou, Qiao- qiao She, and Haifeng Wang。2019。使用位置感知序列标注联合抽取实体和重叠关系。载于第三十三届*AAAI*人工智能大会，*AAAI 2019*，第三十一届人工智能创新应用大会，*IAAI 2019*，第九届*AAAI*人工智能教育进展研讨会，*EAAI2019*，檀香山，夏威夷，

美国，*2019*年*1*月*27*日至*2*月*1*日，第6300–6308页。AAAI出版社。

Jacob Devlin、Ming-Wei Chang、Kenton Lee 和 Kristina Toutanova。2019 年。 BERT：用于语言理解的深层双向 Transformer 预训练。载于《*2019* 年计算语言学协会北美分会年会论文集：人类语言技术》，第 *1* 卷（长篇与短篇论文），第 4171–4186 页，明尼阿波利斯，明尼苏达州。计算语言学协会。

Li Dong、Chris Quirk 和 Mirella Lapata。2018 年。 神经语义解析的置信度建模。载于计算语言学协会第 *56* 届年会论文集（第 *1* 卷：长篇论文），第 743–753 页，澳大利亚墨尔本。计算语言学协会。

Joe Ellis、Jeremy Getman、Justin Mott、Xuansong Li、Kira Griffitt、Stephanie M. Strassel 和 Jonathan Wright。2013 年。 2013 年知识库构建评测的语言资源。载于 第六届文本分析会议论文集，*TAC 2013*，*2013* 年 *11* 月 *18–19* 日，美国马里兰州盖瑟斯堡。美国国家标准与技术研究院。

Yarin Gal 和 Zoubin Ghahramani。2016。Dropout 作为贝叶斯近似：表示深度学习中的模型不确定性。载于第*33*届国际机器学习会议论文集，*ICML 2016*，美国纽约市，*2016*年*6*月*19*日至*24*日，第48卷 *JMLR Workshop and ConferenceProceedings*，第1050–1059页。JMLR.org。

Chuan Guo、Geoff Pleiss、Yu Sun 和 Kilian Q. Weinberger。2017年。现代神经网络的校准。载于第*34*届国际机器学习会议论文集，*ICML 2017*，澳大利亚新南威尔士州悉尼，*2017*年*8*月*6*日至*11*日，第70卷 *Proceedings ofMachineLearning Research*，第1321–1330页。PMLR。

Pankaj Gupta、Hinrich Schütze 和 Bernt Andrassy。2016。用于联合实体与关系抽取的表填充式多任务循环神经网络。载于*COLING 2016* 论文集：第*26*届国际计算语言学会议技术论文，第2537–2547页，日本大阪。COLING 2016 组委会。

Kailong Hao、Botao Yu 和 Wei Hu。2021年。认识假阴性：一种用于远程监督关系抽取的对抗训练方法。载于*2021*年自然语言处理实证方法会议论文集，第9661–9672页，线上及多米尼加共和国蓬塔卡纳。计算语言学协会。

Jianfeng He、Xuchao Zhang、Shuo Lei、Zhiqian Chen、Fanglan Chen、Abdulaziz Alhamadani、Bei Xiao 和 ChangTien Lu。2020 年。提升文本分类中不确定性估计的准确性。载于*2020* 年实证方法会议论文集自然语言处理（*EMNLP*），第 8362–8372 页，在线出版。计算语言学协会。

Dan Hendrycks 和 Kevin Gimpel。2017年。用于检测神经网络中错误分类样本和分布外样本的基线。载于第*5*届国际学习表征会议（*ICLR 2017*）论文集，法国土伦，*2017*年*4*月*24*日至*26*日，会议论文集。OpenReview.net。

Xuming Hu、Chenwei Zhang、Yawen Yang、Xiaohe Li、Li Lin、Lijie Wen 和 Philip S. Yu。2021年。面向低资源关系抽取的梯度模仿强化学习。载于*2021*年自然语言处理实证方法会议论文集，*EMNLP 2021*，线上活动／多米尼加共和国蓬塔卡纳，*2021*年*11*月*7*日至*11*日，第2737–2746页。计算语言学协会。

Wei Jia、Dai Dai、Xinyan Xiao 和 Hua Wu。2019。ARNOR：基于注意力正则化的远程监督关系分类噪声降低。载于第*57*届计算语言学协会年会论文集，第1399–1408页，意大利佛罗伦萨。计算语言学协会。

Alex Kendall 和 Yarin Gal。2017年。计算机视觉的贝叶斯深度学习需要哪些不确定性？载于《神经信息处理系统进展》第*30*卷：*2017*年神经信息处理系统年会，*2017*年*12*月*4*日至*9*日，美国加利福尼亚州长滩，第5574–5584页。

Aaron Klein、Stefan Falkner、Jost Tobias Springenberg 和 Frank Hutter。2017年。使用贝叶斯神经网络进行学习曲线预测。收录于 第五届国际学习表征会议（*ICLR 2017*），*2017*年*4*月*24*日至*26*日，法国土伦，会议论文集。OpenReview.net。

John D. Lafferty、Andrew McCallum 和 Fernando C. N. Pereira。2001年。条件随机场：用于分割和标注序列数据的概率模型。载于第十八届机器学习国际会议论文集，ICML '01，第282–289页，美国加利福尼亚州旧金山。Morgan Kaufmann出版社。

Shuyang Li、Yufei Li、Jianmo Ni 和 Julian McAuley。2022。SHARE：一种用于分层辅助食谱编辑的系统。载于*2022*年自然语言处理实证方法会议论文集，第11077–11090页，阿布扎比，阿拉伯联合酋长国。计算语言学协会。

Xiao Ling 和 Daniel S. Weld。2012年。细粒度实体识别。载于*2012*年*7*月*22*日至*26*日在加拿大安大略省多伦多举行的第二十六届 *AAAI* 人工智能会议论文集。AAAI出版社。

Liyuan Liu、Xiang Ren、Qi Zhu、Shi Zhi、Huan Gui、Heng Ji 和 Jiawei Han。2017年。用于关系抽取的异构监督：一种表示学习方法。载于*2017*年自然语言处理实证方法会议论文集，第46–56页，丹麦哥本哈根。计算语言学协会。

Mike Mintz、Steven Bills、Rion Snow 和 Daniel Ju- rafsky。2009年。无需标注数据的关系抽取远程监督。载于第*47*届 *ACL* 年会暨第*4*届 *AFNLP* 自然语言处理国际联合会议论文集，第1003–1011页，新达城，新加坡。计算语言学协会。

Makoto Miwa 和 Mohit Bansal。2016。在序列和树结构上使用长短期记忆网络进行端到端关系抽取。载于计算语言学协会第*54*届年会论文集，*ACL 2016*，*2016*年*8*月*7*日至*12*日，德国柏林，第*1*卷：长篇论文。计算机语言学协会。

Farhad Nooralahzadeh、Jan Tore Lønning 和 Lilja Øvrelid。2019年。通过部分标注对远程监督命名实体识别进行基于强化学习的去噪。载于第二届低资源自然语言处理深度学习方法研讨会论文集，*DeepLo@EMNLP-IJCNLP2019*，中国香港，*2019*年*11*月*3*日，第225–233页。计算语言学协会。

EmmanouilAntonios Platanios、Otilia Stretcu、Graham Neubig、Barnabás Póczos 和 Tom M. Mitchell。2019年。用于神经机器翻译的基于能力的课程学习。载于*2019*年计算语言学协会北美分会会议论文集：人类语言技术，*NAACL-HLT 2019*，美国明尼苏达州明尼阿波利斯，*2019*年*6*月*2*日至*7*日，第*1*卷（长篇与短篇论文），第1162–1172页。计算语言学协会。

Alec Radford、Jeff Wu、Rewon Child、David Luan、Dario Amodei 和 Ilya Sutskever。2019。语言模型是无监督的多任务学习者。

Sebastian Riedel、Limin Yao 和 Andrew McCallum。2010年。在无标注文本的情况下对关系及其提及进行建模。载于数据库中的机器学习与知识发现，欧洲会议，*ECML PKDD 2010*，西班牙巴塞罗那，*2010*年*9*月*20*日至*24*日，论文集，第*III*部分，第6323卷，收录于计算机科学讲义，第148–163页。施普林格。

Yuming Shang、Heyan Huang、Xin Sun、Wei Wei 和 Xian-Ling Mao。2022。用于远程监督关系抽取的模式感知自注意力网络。信息科学，584：269–279。

克劳德·E·香农。1948。通信的数学理论。贝尔系统技术期刊，27(3):379–423。

Artem Shelmanov、Evgenii Tsymbalov、Dmitri Puzyrev、Kirill Fedyanin、Alexander Panchenko 和 Maxim Panov。2021年。你的 Transformer 有多确定？载于第*16*届欧洲计算语言学协会分会会议论文集：主卷，第1833–1840页，线上举行。计算语言学协会。

Zhixing Tan、Mingxuan Wang、Jun Xie、Yidong Chen 和 Xiaodong Shi。2018。基于自注意力的深度语义角色标注。载于第三十二届*AAAI*人工智能大会、第三十届人工智能创新应用大会和第八届*AAAI*人工智能教育进展研讨会论文集，AAAI’18/IAAI’18/EAAI’18。AAAI出版社。

Hongjun Wang 和 Yisen Wang。2022。通过自集成对抗训练提升鲁棒性。载于第十届国际学习表征会议，*ICLR 2022*，线上活动，*2022*年*4*月*25*日至*29*日。OpenReview.net。

Shuo Wang、Yang Liu、Chao Wang、Huanbo Luan 和 Maosong Sun。2019。通过基于不确定性的置信度估计改进回译。载于*2019*年自然语言处理实证方法会议和第九届国际自然语言处理联合会议论文集，*EMNLP-IJCNLP 2019*，中国香港，*2019*年*11*月*3*日至*7*日，第791–802页。计算语言学协会。

Yijun Xiao 和 William Yang Wang。2019。量化自然语言处理任务中的不确定性。载于第三十三届*AAAI*人工智能会议（*AAAI 2019*）、第三十一届创新人工智能应用会议（*IAAI 2019*）、第九届*AAAI*人工智能教育进展研讨会（*EAAI 2019*），美国夏威夷州檀香山，*2019*年*1*月*27*日至*2*月*1*日，第7322–7329页。AAAI出版社。

Bowen Yu、Zhenyu Zhang、Xiaobo Shu、Tingwen Liu、Yubin Wang、Bin Wang 和 Sujian Li。2020。基于新颖分解策略的实体与关系联合抽取。载于*ECAI 2020*——第*24*届欧洲人工智能会议，*2020*年*8*月*29*日至*9*月*8*日，西班牙圣地亚哥-德孔波斯特拉——包括第十届人工智能杰出应用会议（*PAIS 2020*），第325卷，人工智能与应用前沿，第2282–2289页。IOS出版社。

Suncong Zheng、Feng Wang、Hongyun Bao、Yuexing Hao、Peng Zhou 和 Bo Xu。2017年。基于新颖标注方案的实体与关系联合抽取。收录于计算语言学协会第*55*届年会论文集（第*1*卷：长篇论文），第1227–1236页，加拿大温哥华。计算语言学协会。

Zexuan Zhong 和 Danqi Chen。2021年。一种令人惊讶地简单的实体与关系抽取方法。收录于

《*2021*年北美计算语言学协会分会会议论文集：人类语言技术》，第 50–61 页，线上举行。计算语言学协会。

Yikai Zhou、Baosong Yang、Derek F. Wong、Yu Wan 和 Lidia S. Chao。2020 年。面向神经机器翻译的不确定性感知课程学习。载于第 *58* 届计算语言学协会年会论文集，第 6934–6944 页，线上举行。计算语言学协会。