Modeling and Simulation 建模与仿真, 2024, 13(3), 2579-2590
Published Online May 2024 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2024.133235

Hans汉斯

基于Span-BERT的实体关系联合抽取的研究

张　生，宋　琦，韩　韧

上海理工大学光电信息与计算机工程学院，上海

收稿日期：2024年4月19日；录用日期：2024年5月17日；发布日期：2024年5月23日

摘　要

实体抽取和关系分类是自然语言处理领域其他任务的基石，二者效果直接或间接影响其他任务的效果。近年来得益于预训练语言模型在自然语言处理应用的巨大成功，实体关系联合抽取发展迅速，而当下使用BERT进行预训练基于Span抽取方式的联合抽取模型在解决实体重叠等问题的同时，依旧存在面对长文本效果变差，模型泛化性较差等问题。本文提出一个基于预训练语言模型Span-BERT进行微调的联合抽取模型，以Span为单位实现实体关系的联合抽取，在抽取训练过程中引入负样本采样策略，并在Span-BERT中进行有效提取，以此来增强模型性能和鲁棒性。实验结果和消融实验表明了该方法的有效性，通过不同程度的噪音化数据集SciERC，证明了本模型具有良好的鲁棒性，同时在ADE、CoNLL2004和SciERC个基准数据集上取得了不错的结果。

关键词

联合抽取，深度学习，实体关系分类，负样本采样

Research on Entity Relation Joint Extraction Based on Span-BERT

Sheng Zhang, Qi Song, Ren Han

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 19th, 2024; accepted: May. 17th, 2024; published: May. 23rd, 2024

Abstract

Entity extraction and relation classification serve as the cornerstones of various tasks in the field of natural language processing, with their effectiveness directly or indirectly impacting the outcomes of other tasks. In recent years, owing to the remarkable success of pre-trained language models in natural language processing applications, joint extraction of entities and relations has

文章引用: 张生, 宋琦, 韩韧. 基于 Span-BERT 的实体关系联合抽取的研究[J]. 建模与仿真, 2024, 13(3): 2579-2590.
DOI: 10.12677/mos.2024.133235

rapidly evolved. However, current pre-training methods utilizing BERT based on Span extraction suffer from challenges such as diminished performance on long texts and poor model generalization, despite addressing issues like entity overlap. This paper proposes a joint extraction model fine-tuned on the pre-trained language model Span-BERT. It achieves joint extraction of entities and relations at the Span level, introducing a negative sample sampling strategy during the extraction training process. Effective extraction is carried out within Span-BERT to enhance model performance and robustness. Experimental results and ablation studies demonstrate the effectiveness of this approach. Evaluated on different levels of noisy datasets such as SciERC, the model exhibits robustness. Moreover, it achieves promising results on benchmark datasets, including ADE, CoNLL2004, and SciERC.

Keywords

Joint Extraction, Deep Learning, Entity Relationship Classification, Negative Sample Sampling

1. 引言

在自然语处理任务中，实体抽取(命名实体识别)和关系抽取(关系分类)是知识图谱构建、语义搜索等自然语言处理任务的基石。不同于先抽取实体再关系分类的传统流水线模型，实体关系联合抽取能有效减轻流水线模型中的暴露偏差等问题，因此实体关系联合抽取模式被越来越多的学者所关注。

通常联合抽取模型如 CasRel [1]通过基于序列标注的方法实现，其实现简单，准确率高，但无法解决重叠实体的问题。基于指针网络的模型[2]在保证准确率的同时能够解决重叠实体的问题，但容易遇到标签不平衡问题。基于 Span 的联合抽取模型把实体认定为一个 Span 片段，通过划分开始和结束位置定义抽取目标的范围，精确捕获需要抽取的信息。候选实体会被用于进行实体分类和关系分类，最终得到三元组。基于 Span 的抽取方式精度更高，能更好处理实体重叠的问题。

本文提出一种基于 Span 的实体关系联合抽取模型，以 Span 为单位进行实体关系联合抽取，通过 Span-BERT 预训练语言模型[3]进行微调。在使用 Span 抽取方式解决实体重叠问题的基础任务上，改善该抽取方式对标注数据过度依赖导致的模型鲁棒性差的现象。以 Span 为单位极大的提高后续实体关系分类时选取实体边界的准确性，识别效果得到了保障。我们在模型中引入负样本采样策略，用于降低这些稀疏的负样本对模型造成的影响，进一步提高模型的鲁棒性。我们还进行噪音添加，来直观的检验模型鲁棒性的提升。

根据上面所述，本文提出的模型主要贡献如下：

1. 提出基于 Span-BERT 预训练模型的实体关系联合方法。在遵循基于 Span 的方法进行实体的识别和过滤，有效解决实体重叠的问题的同时取得了不错的抽取效果，以及用一个无标记的上下文表示进行关系分类。

2. 本文采取一种负样本(Negative sample)标记方法。有效改善在面对文本长度增加时，无效的负样本数量也会增加的问题。该方法在训练时将候选 Span 中的非实体 Span 和无关实体标记为负样本，同时这些负样本往往是非常稀疏的，负样本采样能够提高模型面对长文本中大量冗余的负样本时判别的精度，降低这些负样本带来的噪声对模型影响的同时，提高了模型的效率和鲁棒性。

本文使用了 Span-BERT 实现了基于 Span 的实体关系联合抽取模型，并研究了上述所提到的问题，在 ADE、SciERC、CoNLL04 三个基准数据集上验证了该模型。实验结果表明，本文提出的模型相较于各 SOTA 模型，在实体关系联合抽取评估指标 F1 和模型的鲁棒性上取得了出色的效果。此外在 SciERC 数据集上进行了模型的泛化性和鲁棒性的测试，结果表明本模型在面对存在大量噪音的数据的同时，相较于其他模型鲁棒性和泛化性得到了明显的提高。

2. 相关工作

2.1. 联合抽取

实体抽取是关系抽取的前置任务，主要是识别命名实体的范围，将其分类为预定义的类别。关系抽取则是从文本中提取实体之间具有的语义关系。最初 RNN [4]、CNN [5]、LSTM [6]等网络模型都被用于抽取任务中，并取得了良好的效果。传统联合抽取采用流水线(pipeline)方式的模型忽略两个子任务的相关性，实体抽取的效果会直接影响关系抽取的性能，有暴露偏差和误差积累的问题，还存在信息冗余、实体重叠等问题。

联合学习模型很好的利用实体和关系的交互信息，解决了上述问题。抽取方式常见有基于序列标注，其将实体关系组合成 BIO/BILOU 的复合标签[7]，但这样导致了标签样本比较稀疏，难以收敛，且不能解决实体重叠问题；另一种通过预测多个序列的方式得到多个三元组，例如针对每种关系预测其头实体和尾实体[8]，或以每个 token 作为 head 来生成对应的关系和 head-tail [9]等，这样的方法相对更易收敛，也可以解决实体对重叠问题，但模型较复杂。

Span-Re [10]对文中所有 span 片段进行列举后，根据实体分类生成候选关系 Span，经过多头注意力机制去除重叠关系后完成分类。DYGIE [11]，构建图来实现 Span 表示的共享，之后基于动态 Span 图进行实体关系抽取。TPLinker [12]则是提出了基于标记嵌入矩阵(基于填表)的实体关系联合抽取方法，通过对所有的 token 对构造一个矩阵，对矩阵中的每一个 token 进行链接标注(类似于填表)，最后一步抽取出实体和实体对之间的关系，不存在任何相互依赖的步骤，避免了暴露偏差问题。Spert 模型[13]提出的一个经典的基于 Span 的实体关系抽取框架，它学习了一个 wideth-embedding 表征 Span 的长度，将 Span 范围内的 token 进行 max-pooling，与 BERT 输出的[CLS] token 向量以及长度向量进行拼接作为 Span 的向量表示，在关系分类时，将两个 Span 之间的 token 向量进行 max-pooling 获得了一个上下文表示，并将其与两个 Span 实体的向量进行拼接传入分类器进行关系分类的预测。

2.2. 预训练模型

近些年预训练语言模型在实体关系联合抽取等自然语言处理任务中表现出色。预训练语言模型可以通过大规模的自监督学习，从大量未标注的文本数据中自主学习语言模式和丰富的语义表示，然后根据少量标注样本进行微调，以适应具体下游任务的需要。在实体关系联合抽取任务中，如 BERT 通常被用来进行条件随机场或序列标注等传统方法所使用的下游任务微调。

联合抽取模型[7]通常使用 BERT 对句子进行编码，提取出句子中上下文信息，进而获得预训练模型学习到的丰富的特征，大幅提高了联合抽取任务的效果。此外上述实体联合抽取模型均是基于 BERT 对下游任务进行微调，取得了较好的效果。但这些基于 BERT 的实体关系联合抽取模型由于 BERT 对输入文本的长度有一定的限制，使得其难以处理长文本或者需要考虑全文情况的任务。此外 BERT 中的 NSP 任务还会学习一些与下游任务无关的信息，如句子之间的转移关系等，不仅影响下游任务的性能，还会使得模型过多的关注句子之间的关系，从而忽略句子的语义信息，影响模型效果。本文中使用 Span-BERT 对下游任务进行微调，它针对 BERT 在处理实体关系抽取等任务中存在的局限性进行了改进，并在多项

自然语言处理任务中都取得了优秀的表现。

2.3. 负样本采样策略

负样本通常指的是训练集中难以分类的、具有挑战性的负样本。这些样本往往容易被误分类为正样本，且与其他负样本相比具有更高的难度和噪声。通过引入负样本进行训练，可以帮助模型更好地理解类别之间的边界和区别，从而提高模型的鲁棒性和泛化性能。

2019 年 Jessa Bekke 等人在论文[14]中使用了 PU-Learning (Positive-Unlabeled Learning)方法，在训练模型时将标记正确的样本作为正例，从未标记的样本中随机采样一定数量的负例作为负样本。

2019 年 Zhi-Xiu Ye 等人该论文[15]中在进行 few-shot 关系分类任务时，通过随机采样一些与目标关系无关的负样例来增强模型的泛化能力。2021 年 Kenton Lee 等人该论文[16]中使用了一种基于随机扰动的负样本采样方法，通过对原始关系图进行多次随机扰动生成负样本，以增加训练数据量和增强模型的泛化能力。其中 spert 在第二个阶段中，模型使用了负样本采样策略来增强模型的泛化能力和鲁棒性。具体而言，该论文中使用了两种负样本采样策略：基于句子级别的随机负样本采样和基于实体级别的随机负样本采样。其中，句子级别的随机负样本采样是通过从与正样本不同的句子中随机采样一定数量的样本作为负样本；实体级别的随机负样本采样则是通过从与正样本不同的实体对中随机采样一定数量的样本作为负样本，结论认为来自同一句子的负样本产生的训练既高效又有效，足够数量的强负样本似乎是至关重要的，这对我们有启发式的思考。

3. 基于 Span-BERT 的联合抽取方法方法

我们的模型使用预训练好的 Span-BERT 作为模型为基础，如图 1 所示：一个句子被输入到 Span-BERT 中后，首先进行 tokenize，得到一串 n 个字节对编码(BPE)的 token，经过 Span-Masking 后得到的是基于 Span 被掩盖的 token，其中“some football games”(黑色)作为一个 Span 被掩盖，再送入 SBO 任务后，得

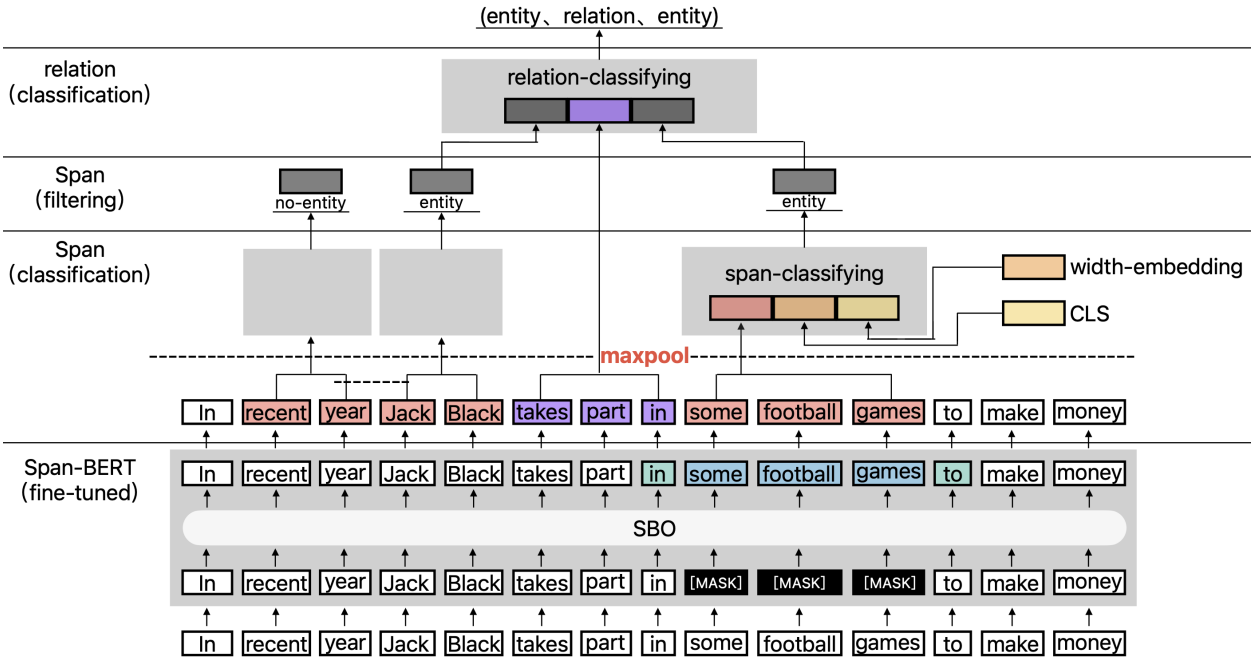

Figure 1. Entity relation extraction process
图 1. 实体关系抽取过程

到一个长度为 $n+1$ 的嵌入序列 $E:=(e_1,e_2,\cdots,e_n,C)$ 其引入边界词来预测所掩盖的分词内容，例如“in”和“to”(青色)作为“some football games”(红色)的边界词来预测被掩盖的内容。我们的方法是在所有 Span 子序列中检测实体，例如一个 Span 序列(some, football, games)映射到跨度(some)、(some, football)、(some, football, games)等检测实体，再拼接长度信息(棕色)和 CLS 语义信息(黄色)对该跨度进行实体分类得到实体。最后将过滤获取的有效实体对，与上下文文本内容(紫色)一起拼接进行关系分类，最终得到三元组。

3.1. 基于 Span-BERT 预训练

跨度选择模型(Lee 等人[17]通常使用其边界标记(开始和结束)创建跨度的固定长度表示，Span 末端的表示能够尽可能多地总结内部 Span 内容，它通过引入一个跨度边界目标来实现这一点，该目标涉及仅使用边界处观察到的标记的表示来预测一个被屏蔽跨度的每个标记，很好的支持了基于 Span 的模型。这里我们使用 Span-BERT 进行基于 Span 的转化，首先输入一个句子 $S:=(t_1,t_2,\cdots,t_n)$，之后 Span-BERT 将其按照 Span 进行随机分割为根据采样长度进行 Span-Masking，共有 15%的 token 会被以 Span 为单位被 mask，即 $S_m:=(t_1,\cdots,t_s,\cdots,t_e,\cdots,t_n)$，$s$ 到 e 位置的 token 代表被 mask 的一个 Span，其中使用几何分布来获得平均采样长度，分词长度平局 3.8。之后将通过 MLM 和 SBO 任务进行训练，其中 SBO 任务目的是让被 mask 的 Span 前后边界的两个词也能够学习到被 mask 到内容，做法是将前后边界的两个词向量加上被 mask 词的位置向量来预测原词，具体训练所输入的向量包含(t_{s-1}，t_{e+1}，p_{i-h+1})，i 是在被 mask 的 Span 中的序号为 i 的 token，$s-1$ 是被 mask 的 Span 的前面的边界，$e+1$ 是被 mask 的 Span 后面的边界。最终获得基于 Span-BERT 的到的嵌入序列 $E:=(e_1,e_2,\cdots,e_n,c)$，其在后续基于 Span 的任务获得了更好的表现，我们下面的任务就是在此基础上进行训练。

3.2. 基于跨度进行实体分类

将来自于 Span-BERT 的嵌入序列 E 按照任意的候选跨度作为输入，这里定义该跨度为 $s:=(e_i,e_{i+1},\cdots,e_{i+k})$，$k$ 表示其 Span 的长度。这里定义了一个预先定义好一个实体类别的集合 $Te\cup\{\text{none}\}$，例如人或地点等，其中还定义了 none 类型，即当最后抽取出的 Span 类型不属于任何一个实体类型，该 Span 将会被归类于 none 中，之后部分会作为负样本，其余的被过滤处理。

实体分类器的组成如图 1 所示，这里使用融合函数 max-pooling 通过最大池化操作组合嵌入序列 E，即 $f:=(e_i,e_{i+1},\cdots,e_{i+k})$。融合函数这里使用 max-pooling 进行最大池化操作，从而可以很好的提取出最有效的表征。

Span 宽度嵌入：在模型训练过程中为每个不同的 Span 宽度训练一个固定大小的宽度嵌入表示 $(1,2,\cdots,k+1)$ [18]，这里从特定的嵌入矩阵中为宽度嵌入是为 $k+1$ 的嵌入表示 W_{k+1}，它们通过反向传播进行学习，这里 $M(s)$代表二者拼接的跨度。

$$M(s)=f(e_i,e_{i+1},\cdots,e_{i+k})\circ W_{k+1}$$

后续我们添加分类器 token c，[CLS]符号是一个特殊的分类器标记，它代表整个语句的语义特征向量，其对应的输出作为文本的语义表示，用于捕捉整个句子的上下文。上下文语义是一个重要的消歧义来源，如关键字“职位”和“学生”是实体类“人”的强力指标。

下面用于输入实体分类的 Span 语义由三部分拼接而成

$$c(e)=(M(s),c)$$

进行实体分类时，我们使用了 T-softmax [19]，引入了温度超参数 T。因为本文中引入了负样本策略，但当原始的分布熵较小的时候，负标签的值很接近 0，这对于损失函数的贡献非常的小，违背了我们引

入负样本策略的初衷，当 T 值越高时，其输出概率分布就会越来越平滑，分布熵越大，负标签信息的作用会被放大，模型就能够更好的关注和充分利用负样本，进而能够有效地提升模型的性能，此外温差 softmax 还能够帮助模型学习到一定的隐藏信息，帮助模型提高准确率的同时在进行实体分类时避免陷入局部最优解。

$$q_i = \frac{\exp(Z_i/T)}{\sum_j \exp(Z_i/T)}$$

我们把上述输入送到温差 softmax 分类器中：

$$Y_s = \text{T-SoftMax}(c(e), T)$$

之后通过查看来自温差 softmax 分类器的得分，确定该 Span 所属于的实体类，即跨度分类器根据分数将输入的跨度 s 对应的类型输出为 Y_s。然后这里会将那些未被识别出来的 Span 归类为 none，并保留这些 noneEntity 用于构建关系分类中的负样本，最终获得了一个预测得到的实体集 S_e。

3.3. 关系分类

定义了关系类型的集合 $T_r \cup \{none\}$，同实体分类，加入预定义好的类型 *none*，用于归类那些没有匹配到合适类型的关系 Span。得到来自上游实体抽取阶段的实体集 S_e 后，从中抽取出每一对候选实体对(s_1, s_2)，为了表示候选实体对，这里也为它们拼接上宽度嵌入，同第二部分实体分类一样，这里拼接的跨度用 $M(s_1)$和 $M(s_2)$表示，并用该拼接跨度来判断 R 中的关系对于候选实体对是否成立，用于关系分类的关系元组定义如式所示

$$\langle s_1, s_2 \rangle \in \{S_e \otimes S_e\}$$

$$s_1 \neq s_2$$

但关系抽取阶段不同于实体抽取阶段，虽然上述所说语境中的职位和学生能够直接作为表示关系的重要指标 c，但它并不能直接表达多种关系的长句子，因此关系抽取阶段直接使用实体周围的 token 来抽取更本地化的上下文语义，即从第一个实体 s_1 的末尾到第二个实体店开始的跨度，并使用最大池化操作将其嵌入组合起来，获得了两个实体店上下文表示 $c(s_1, s_2)$。

下面用于实体分类的 Span 语义由三部分拼接而成

$$c(r) = (M(s_1), c(s_1, s_2), M(s_2))$$

进行关系分类时我们同上使用温差 softmax 进行分类，

$$Y_r = \text{T-SoftMax}(c(r), T)$$

3.4. polyloss 损失函数

本文联合抽取模型的损失函数定义如下：

$$\mathcal{L} = \mathcal{L}_e + \mathcal{L}_r$$

$\mathcal{L}_e$ 是实体分类所使用的基于交叉熵损失(cross-entropy loss)的 polyloss [20]损失函数。$\mathcal{L}_r$ 是关系分类所使用的二分类交叉熵损失函数-BCEWithLogitsLoss()。

其中，polyloss 函数可认为是一个用于理解和改进交叉熵损失的一个框架，该框架的改进源自于 CE-loss 函数中基于 $(1-P_t)^j$ 的泰勒展开式子，将其分解为一系列的加权多项式 $\sum_{j=1}^{\infty} \alpha_j (1-P_t)^j$，其中 P_t 是

目标类标签的预测概率，α_j 为多项式系数，用于为多项式基进行加权，当 $\alpha_j = 1/j$ 时，该式等效于 CE-loss，可以灵活地应用于不同的场景，但在联合抽取任务中，经常出现某个类别的训练数据量很少的情况，polyloss 则可以通过对样本进行加权，从而更好地处理这种类别不平衡的问题，此外得益于核函数的特性，它能够更好的处理训练数据和测试数据之间的偏差问题，提高神经网络对于噪声和异常数据的鲁棒性和泛化性，保证了模型的稳定性和可靠性，提高了性能和分类精度。

$$F_{Ploy\text{-}N} = -\log(P_t) + \sum_{j=1}^{N} \varepsilon_j (1-P_t)^j$$

BCEWithLogitsLoss 则较为常见，其将 sigmoid 层和 BCEloss 函数整合在一起，相比之前更加稳定。

4. 实验

4.1. 数据集

本模型在 ADE [21]、CoNLL04 [22]、SciERC [23]三个基准数据集上进行的实验，以下简称 ADE、CoNLL04 和 SciERC。

1. ADE (Adverse Drug Events Corpus)是一个在医学领域用于研究和发展药物不良反应的关系抽取的数据集。在 ADE 中每一篇文档都标注了药物与不良反应之间的关系。标注的关系包括药物名称、不良反应名称以及二者之间的关系类型，如“治疗–不良反应”、“预防–不良反应”等。这些标注信息使得研究人员可以在关系抽取任务中训练和评估模型。

2. CoNLL2004：CoNLL2004 是一个命名实体识别数据集。其目标是识别新闻文章中的实体(人物、组织、地点等)。CoNLL2004 数据集包含训练集、开发集、测试集，其中训练集使用 Wall Street Journal 语料库，开发集和测试集使用其他不同来源的语料库。CoNLL2004 的标注格式为 BIO (Beginning, Inside, Outside)，即将每个标记划分为 Begin、Inside 或 Outside 三种标记，简单且易于训练。

3. SciERC：SciERC 是一个面向科学领域的实体及关系标注数据集，与 ADE 和 CoNLL2004 相比，它针对于科学论文的实体和关系做了更为细致全面的标注。SciERC 数据集包含 500 篇计算机科学领域的论文摘要，并将这些文本据此标注为实体及其之间的关系。相应的实体分为四类：Task、Method、Metric 和 Material，每个实体类别使用了不同的标签。SciERC 数据集的目的是为科研社区提供一个更加精细全面的文本实体/关系标注数据集，从而推动科研领域的文本抽取研究。

4.2. 实验配置

本文使用 Span-BERT 作为嵌入生成器，将其预先从大量语料中获得的知识迁移到本模型中，本模型围绕它对其进行微调，从而进行下游实体关系联合抽取到任务。本文把模型训练的 batch 设置为 8，dropout 设置为 0.2，宽度嵌入的维度设置为 50，学习率设置为 5e − 5，weight decay 设置为 0.01，梯度剪裁阈值设置为 1，对于不同的数据集，本文的 epoch 设置不同，文中对于 Span 过滤的阈值均设置为 10，本文同时采用负采样策略来提高模型的鲁棒性和泛化性，实体和关系两者负样本采样的值均设置为句子中正样本的 30 倍。此外本模型中使用温差 softmax 时，其超参数温度 T 取值为 1 时，其效果相当于传统的 softmax，T 的取值小于 1 时，分布逐渐极端化，模型输出稍微有点起伏并开始收敛，此处认为模型学习到了知识，最终等价于 argmax，在 T 大于 1 时，分布逐渐趋于均匀分布，曲线比价平滑，loss 会变得很大，这样可以防止陷入局部最优解。

4.3. 基准模型

本文所选模型是目前在上述基准数据集上取得了优异效果的基于 Span 抽取方式的模型，或典型的基

于 Span 抽取方式的模型，与这些基准模型相比，抽取评估结果得到了不同程度的提升，模型的鲁棒性也能得到了一定程度的提升。

1. Spert 模型：

Spert 一种基于跨度的实体关系联合抽取模型，使用 Transformer 网络进行编码，并在跨度级别进行标签预测实体关系。在基于跨度的抽取方式中在 ADE 和 CoNLL04 数据集上取得了最优的结果。

2. CLDR + CLNER 模型[24]：

CLDR + CLNER 将对比学习应用于语言模型嵌入中，其通过对抗性学习实体关系语境生成对比样本，来指导语言模型生成更丰富和准确的嵌入表示，巧妙的提升模型的鲁棒性。

3. PL-Marker [25]模型：

PL-Marker 是一种经典的管道抽取模型，它把实体类型作为前后缀融入到关系分类到句子中，引入了实体信息，并且获得了显著的性能表现，在 SciERC 数据集上取得了 SOTA 级的表现。虽然管道模型仍然存在很多问题，但 PL-Marker 也证明了实体信息对于关系分类的重要性。

4. Multi-turn QA [26]模型：

Multi-turn QA 模型结合了阅读理解和对话建模的技术，能够理解和回答由多个问答对组成的对话。它能够利用对话历史和上下文信息来更好地理解问题，并给出准确的答案，是 CoNLL04 数据集上基于序列标注的较优模型。

4.4. 实验结果和分析

4.4.1. 对比实验

主要叙述在经典实体关系抽取的评价指标 F1 上取得的成功，然后在叙述本模型的准确率和召回率，以下基准模型均为基于 BERT 而来的实验结果。

Table 1. Test set result (ADE, CoNLL04, SciERC)
表 1. 测试集结果(ADE、CoNLL04、SciERC)

Dataset	Model	Entity			Relation		
		Precision	Recall	F1	Precision	Recall	F1
ADE	CLDR + CLNER	-	-	88.30	-	-	79.97
	Spert	88.69	89.20	88.94	77.77	79.96	79.24
	$SPAN_{SpanBert}$	88.73	90.05	90.20	77.85	78.99	79.53
CoNLL04	Multi-turn QA	89.00	86.60	87.80	69.20	68.20	68.90
	Spert	85.78	86.84	88.94	74.75	71.52	71.47
	$SPAN_{SpanBert}$	89.37	87.02	89.02	71.51	71.68	71.58
SciERC	Spert	68.53	66.73	67.62	49.79	43.53	46.44
	PL-Marker	-	-	69.90	-	-	53.20
	$SPAN_{SpanBert}$	69.49	67.31	68.74	49.58	44.14	49.37

本文模型和基准、SOTA 模型的对比结果如表 1 所示，本文模型表示为 $SPAN_{Span}$-BERT，表示是基于预训练模型 Span-BERT 进行实体关系联合抽取的模型。使用的评价指标是 NLP 领域常用的准确率、召回率和 F1-score。其中 F1 是统计学中用于衡量而分类模型精确度的一种指标，兼顾了精确率和召回率，是精确率和召回率调和平均数，效果好于算数平均数。其中 SciERC 数据集是专门用于实体关系联合抽

取的数据集，ADE 数据集的结果则已经将重叠实体的情况考虑进去。

从表 1 中可以发现，$SPAN_{Span}$-BERT 在这些基准数据集上与其他基准模型相比取得了不同的突破。具体来说，在 ADE 数据集上，$SPAN_{Span}$-BERT 在实体抽取方面 F1 值提高了 1.90 (CLDR + CLNER)和 1.26 (Spert)，在关系抽取方面 F1 值提高了 0.19 (Spert)。在 Conll04 数据集上，$SPAN_{Span}$-BERT 在实体抽取方面取得了 1.22 (multi-turn QA)和 0.08 (Spert)的 F1 值的提高，关系抽取上取得了 2.68 (Multi-turn QA)和 0.11 (spert)的 F1 值的提高。而在 SciERC 数据集上，$SPAN_{Span}$-BERT 的实体抽取和关系抽取方面 F1 提高并不明显，实体抽取方面 F1 值提高了 1.12 (spert)，关系抽取 F1 值提高了 2.93 (spert)，实体抽取和关系抽取相较于 PL-Marker 均并没有取得效果上的提升，是由于 PL-Marker 借鉴了 PURE 的思想，在关系抽取时引入实体类型等相关信息，关系抽取获得了优异的性能提升，实体信息对关系抽取有启发性的作用。但 PL-Marker 模型属于 pipeline 模型，虽然在抽取性能上提升巨大，不过对于 pipeline 模型传统的例如暴露偏差等弊端并未完全解决，但其在联合抽取中的关系分类这一环节中引入实体信息的思想是值得学习的。

Table 2. The effect of noise level on the model effect
表 2. 噪音程度对模型效果对影响

Dataset	Model	Entity			Relation		
		0%	20%	40%	0%	20%	40%
SciERC	Spert	67.62	60.47	39.67	46.44	38.37	21.65
	PL-Marker	69.9	59.15	29.37	53.20	43.88	17.63
	$SPAN_{SpanBert}$	68.74	65.82	56.43	49.37	45.06	36.72

4.4.2. 模型鲁棒性测试

之后我们测试了模型的鲁棒性和泛化性，我们在 SciERC 数据集中均匀的添加了不同程度的噪音，以此来判断模型受到不同程度噪音的影响是否依然鲁棒，添加噪声的方法是按照一个特定的噪声转移矩阵将一个类别样本的标签随机转换为一个特定类别的标签，来形成类别之间的混淆[27]。

由图 2 明显可观察，蓝色代表 $SPAN_{SpanBert}$，橙色代表 Spert，绿色代表 PL-Marker，其中 $SPAN_{SpanBert}$F1 值下降较为平稳，受噪音影响较小；而 Spert 和 PL-Marker 下降严重，受噪音影响较大。由具体数据表 2 可知，当噪音程度为 20%时，spert 和 PL-Marker 模型 F1 指标均受到影响，实体和关系 F1 指标下降在 10%~16%左右，而本模型下降指标仅为 5%左右。当噪音程度为 40%时，能够明显的发现 spert 和 PL-Marker 模型效果明显变差，F1 指标下降幅度增大，指标下降达到 26%~38%左右，而本模型仅下降 12%左右，证明了模型在面对越来越多的噪音的同时，能避免过多地受到噪音的影响，能够很好地保证鲁棒性，在面对不同领域的数据模型依旧能保证了不错的效果，泛化性得到了保障。可以观察到，PL_Marker 模型由于采用 pipeline 方式进行抽取，虽然其在实体关系联合抽取上表现十分优异，但当面对充斥着噪音的数据时，其表现下滑严重，尤其是关系抽取阶段，pipeline 模型中关系抽取直接收到上游任务实体抽取的影响，误差积累现象严重，关系抽取表现大幅下滑，在图 3 中，当噪音程度达到 40%时，模型对关系的识别遭到了毁灭性的打击，F1 值仅为 17.63。因此对于如何保证模型鲁棒性和泛化性的研究是具有意义的。

4.4.3. 消融实验

首先这里进行评估 Span-BERT 语言模型的预训练效果，本文主要的思想是 Span-BERT 对基于 Span 抽取方式的联合抽取的效果是否达到了预期效果，因此在本文中，选择使用在实体关系联合抽取任务中最为常见的预训练模型 BERT 进行对比，如表 3 所示，使用预训练模型 BERT 的效果在实体抽取和关系

抽取上均有下降，关系抽取效果下降大于实习抽取效果，说明关系抽取受到上游实体抽取任务的影响，该数据也证明了 Span-BERT 对于基于 Span 抽取方式的有效性。

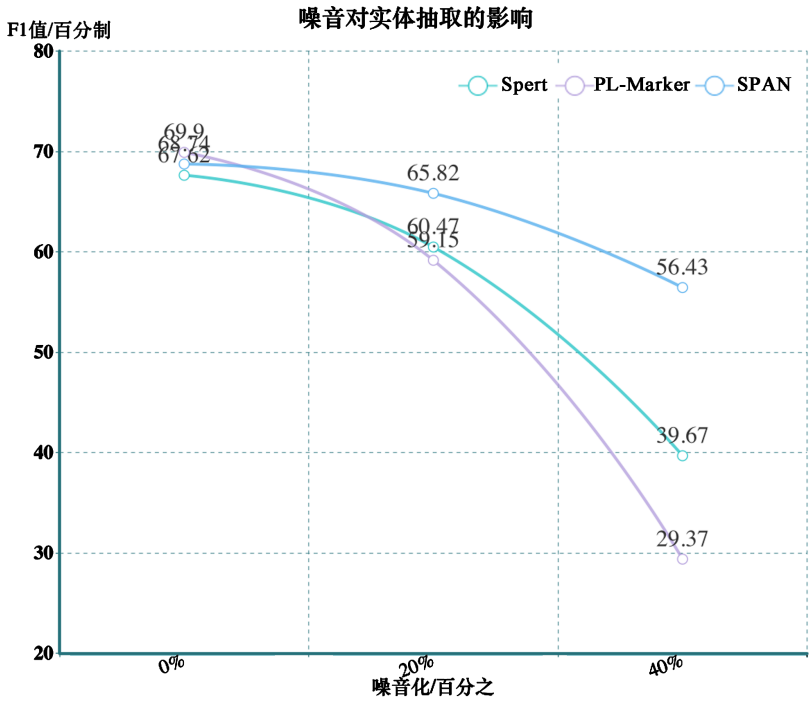

Figure 2. The effect of noise on model entity extraction
图 2. 噪音对模型实体抽取效果的影响

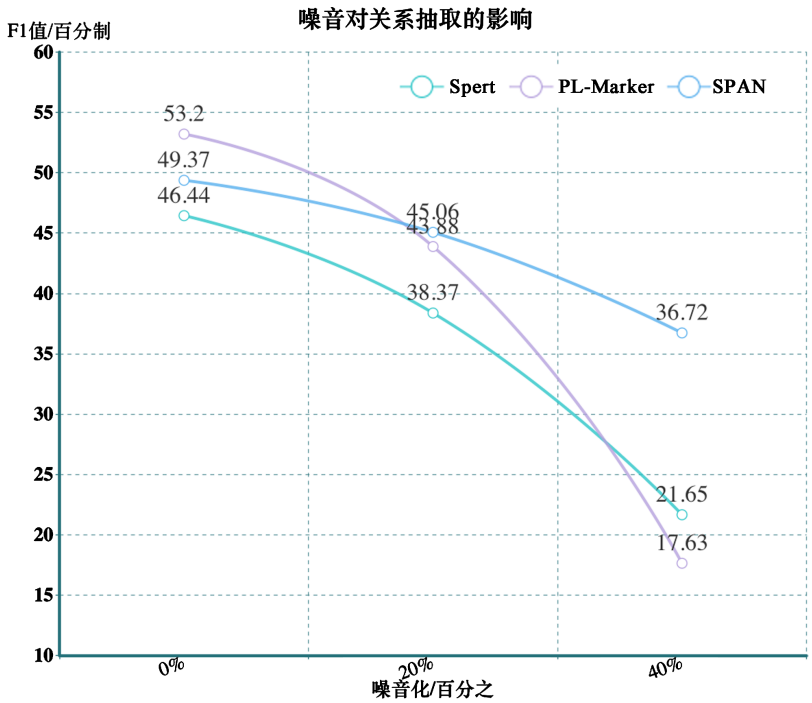

Figure 3. Influence of noise on model relation extraction
图 3. 噪音对模型关系抽取效果的影响

Table 3. The extraction effect is based on different pre-training models
表 3. 基于不同预训练模型对抽取效果

Dataset	Model	Entity F1	Relation F1
ADE	$SPAN_{SpanBert}$	90.20	79.53
	BERT	87.92	70.34
CoNLL04	$SPAN_{SpanBert}$	89.02	71.58
	BERT	86.56	68.13
SciERC	$SPAN_{SpanBert}$	68.74	49.37
	BERT	65.43	45.71

损失函数方面，当我们使用最经典的交叉熵损失来训练模型的时候，发现效果并没有太大区别，可以推出 poly loss 损失函数对本模型并没有带来更加出色的表现，但并不代表其可有可无，正如上方介绍那样，poly loss 本身设计是为了应对不同的任务和场景。在普通的联合抽取任务中，面对不同的数据集，会有不同的表现，正常情况下其等效于交叉熵损失，但当面对少样本数据集或其他预料较少的数据集时，调整加权的参数，模型就能够更好的应对类别不平衡的问题，可能会带来意想不到的效果。

5. 结论

本文提出了一种基于 Span-BERT 的实体关系联合抽取的模型。该模型基于 Span 的抽取方式来抽取实体并进行关系分类，借助强大的预训练语言模型 Span-BERT 和负采样策略取得了不错的效果。具体来说基于 Span 的抽取方式能够解决一些传统模型遇到实体关系重叠等问题，且不像基于表格抽取那么臃肿，是一个更轻量级的抽取模型。得益于基于 Span 进行预训练，结合温差 Soft-Max 和负采样策略，在面对以 Span 为基本单位的实体关系联合抽取的任务时，获得了更高的精度，在性能和速度有所提升的同时，模型获得了更好的鲁棒性，面对质量较差或其他领域的数据时，也能拥有不俗的效果。

参考文献

[1] Wei, Z.P., Su, J.L., Wang, Y., Tian, Y. and Chang, Y. (2020) A Novel Cascade Binary Tagging Framework for Relational Triple Extraction.

[2] Vinyals, O., Fortunato, M. and Jaitly, N. (2015) Pointer Networks. *Advances in Neural Information Processing Systems*, **28**, 1-9.

[3] Joshi, M., Chen, D.Q., Liu, Y.H., Weld, D.S., Zettlemoyer, L. and Levy, O. (2020) Spanbert: Improving Pre-Training by Representing and Predicting Spans. *Transactions of the Association for Computational Linguistics*, **8**, 64-77. https://doi.org/10.1162/tacl_a_00300

[4] Socher, R., Huval, B., Manning, C.D. and Ng, A.Y. (2012) Semantic Compositionality through Recursive Matrix-Vector Spaces. *Proceedings of the* 2012 *Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Jeju, 12-14 July 2012, 1201-1211.

[5] Zhang, D.X. and Wang, D. (2014) Relation Classification via Convolutional Deep Neural Network. *International Conference on Computational Linguistics*, Dublin, August 2014, 2335-2344.

[6] Xu, Y., Mou, L.L., Li, G., Chen, Y.C., Peng, H. and Jin, Z. (2015) Classifying Relations via Long Short Term Memory Networks along Shortest Dependency Paths. Association for Computational Linguistics, Lisbon. https://doi.org/10.18653/v1/D15-1206

[7] Zheng, S.C., Wang, F., Bao, H.Y., Hao, Y.X., Zhou, P. and Xu, B. (2017) Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme. *Proceedings of the* 55*th Annual Meeting of the Association for Computational Linguistics*, **1**, 1227-1236. https://doi.org/10.18653/v1/P17-1113

[8] Yuan, Y., Zhou, X.F., Pan, S.R., Zhu, Q.N., Song, Z.L. and Guo, L. (2021) A Relation-Specific Attention Network for

Joint Entity and Relation Extraction. In: Zhou, Z.H., Ed., *International Joint Conference on Artificial Intelligence*, Yokohama, 7-15 January 2021, 4054-4060. https://doi.org/10.24963/ijcai.2020/561

[9] Dai, D., Xiao, X.Y., Lyu, Y.J., Dou, S., She, Q.Q. and Wang, H.F. (2019) Joint Extraction of Entities and Overlapping Relations Using Position-Attentive Sequence Labeling. *The AAAI Conference on Artificial Intelligence*, Honolulu, 27 January-1 February 2019, 6300-6308. https://doi.org/10.1609/aaai.v33i01.33016300

[10] Zhang, H. (2023) SpanRE: Entities and Overlapping Relations Extraction Based on Spans and Entity Attention.

[11] Luan, Y., Wadden, D., He, L.H., Shah, A., Ostendorf, M. and Hajishirzi, H. (2019) A General Framework for Information Extraction Using Dynamic Span Graphs. https://doi.org/10.18653/v1/N19-1308

[12] Wang, Y.C., Yu, B.W., Zhang, Y.Y., Liu, T.W., Zhu, H.S. and Sun, L.M. (2020) TPLinker: Single-Stage Joint Extraction of Entities and Relations through Token Pair Linking. Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, December 2020, 1572-1582. https://doi.org/10.18653/v1/2020.coling-main.138

[13] Eberts, M. and Ulges, A. (2021) Span-Based Joint Entity and Relation Extraction with Transformer Pre-Training.

[14] Wang, Y., Liu, X.X., Hu, W.X. and Zhang, T. (2022) A Unified Positive-Unlabeled Learning Framework for Document-Level Relation Extraction with Different Levels of Labeling. *Proceedings of the* 2022 *Conference on Empirical Methods in Natural Language Processing*, Abu Dhabi, December 2022, 4123-4135. https://doi.org/10.18653/v1/2022.emnlp-main.276

[15] Ye, Z.-X. and Ling, Z.-H. (2019) Multi-Level Matching and Aggregation Network for Few-Shot Relation Classification. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, July 2019, 2872-2881.

[16] Dutordoir, V., Saul, A., Ghahramani, Z. and Simpson, F. (2022) Neural Diffusion Processes.

[17] Lee, K., Salant, S., Kwiatkowksi, T., Parikh, A., Das, D. and Berant, J. (2017) Learning Recurrent Span Representations for Extractive Question Answering.

[18] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2019) Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT* 2019, Minneapolis, 2-7 June 2019, 4171-4186.

[19] He, C.Y., Annavaram, M. and Avestimehr, S. (2020) Group Knowledge Transfer: Federated Learning of Large Cnns at the Edge. *Advances in Neural Information Processing Systems*, **33**, 14068-14080.

[20] Leng, Z.Q., Tan, M.X., Liu, C.X., Cubuk, E.D., Shi, X.J., Cheng, S.Y., Anguelov, D., *et al*. (2022) Polyloss: A Polynomial Expansion Perspective of Classification Loss Functions. 10*th International Conference on Learning Representations* (*ICLR* 2022), 25-29 April 2022, 1-16.

[21] Gurulingappa, H., Rajput, A.M., Roberts, A., Fluck, J., Hofmann-Apitius, M. and Toldo, L. (2012) Development of a Benchmark Corpus to Support the Automatic Extraction of Drug-Related Adverse Effects from Medical Case Reports. *Journal of Biomedical Informatics*, **45**, 885-892. https://doi.org/10.1016/j.jbi.2012.04.008

[22] Roth, D. and Yih, W.-T. (2004) A Linear Programming Formulation for Global Inference in Natural Language Tasks. In: *Proceedings of the* 8*th Conference on Computational Natural Language Learning* (*CoNLL*-2004) *at HLT-NAACL* 2004, Association for Computational Linguistics, Boston, 1-8.

[23] Luan, Y., He, L.H., Ostendorf, M. and Hajishirzi, H. (2018) Multi-Task Identification of Entities, Relations, and Coreference for Scientific Knowledge Graph Construction. *Proceedings of the* 2018 *Conference on Empirical Methods in Natural Language Processing*, Brussels, October-November 2018, 3219-3232. https://doi.org/10.18653/v1/D18-1360

[24] Theodoropoulos, C., Henderson, J., Coman, A.C. and Moens, M.-F. (2021) Imposing Relation Structure in Language-Model Embeddings Using Contrastive Learning. *Proceedings of the* 25*th Conference on Computational Natural Language Learning*, November 2021, 337-348. https://doi.org/10.18653/v1/2021.conll-1.27

[25] Ye, D.M., Lin, Y.K., Li, P. and Sun, M.S. (2021) Packed Levitated Marker for Entity and Relation Extraction.

[26] Lan, Y.S., He, G.L., Jiang, J.H., *et al*. (2023) Complex Knowledge Base Question Answering: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, **35**, 11196-11215. https://doi.org/10.1109/TKDE.2022.3223858

[27] Qiao, D., Dai, C.C., Ding, Y.Y., Li, J.T., Chen, Q. and Chen, W.L. (2022) SelfMix: Robust Learning against Textual Label Noise with Self-Mixup Training. *Proceedings of the* 29*th International Conference on Computational Linguistics*, Gyeongju, October 2022, 960-970.