基于 Span 方法和多叉解码树的实体关系抽取

张 鑫,冼广铭*,梅灏洋,周岑钰,刘赢方
(华南师范大学 软件学院,广东 佛山 528225)

摘 要:实体关系抽取作为自然语言处理领域的一项关键技术,在构建知识图谱、信息检索等领域有着极为重要的意义。然实体关系抽取模型普遍存在词与词之间依赖性运用不足、实体识别效果低下以及单解码带来的三元组强行执行某种不必要顺序的问题。为了解决这三个方面的问题,提升模型的性能,提出了一种新的实体关系抽取模型。该模型首先运用提取特征能力更强的 BERT 预训练模型获取句子表征,然后采用图卷积神经网络来增强实体与关系之间的依赖关系,再使用对实体提取能力更强的 Span 方法(识别实体的神经网络方法)进行实体抽取,最后采用深度多叉解码树实施并行解码得到相应的关系三元组。在 CoNLL04、ADE 数据集上的实验结果表明,与其他的关系抽取基线模型相比,该模型的 F1 值具有较好的提升,同时也验证了该文模型的有效性与泛化能力。

关键词:实体识别;关系抽取;深度学习;预训练模型;多叉解码树;图神经网络
中图分类号:TP391 **文献标识码**:A **文章编号**:1673-629X(2023)05-0152-07
doi:10.3969/j.issn.1673-629X.2023.05.023

Entity Relation Extraction Based on Span Method and Multi-fork Decoding Tree

ZHANG Xin,XIAN Guang-ming*,MEI Hao-yang,ZHOU Cen-yu,LIU Ying-fang
(School of Software,South China Normal University,Foshan 528225,China)

Abstract:As a key technology in the field of natural language processing,entity relation extraction is of great significance in the construction of knowledge graphs,information retrieval and other fields. However,the entity relation extraction model generally has the problems of insufficient application of dependencies between words,low entity recognition effect,and the forced execution of an unnecessary order of triples brought by single decoding. In order to solve three problems and improve the performance of the model,a new entity relation model is proposed. The model first uses the BERT pre-training model with stronger feature extraction ability to obtain sentence representation,and then uses graph convolutional neural network to enhance the dependency between entities and relationships. The Span method (Neural Network Methods for Recognizing Entities),which has stronger entity extraction ability,is used for entity extraction. Finally,a deep multi-fork decoding tree is used to implement parallel decoding to obtain the corresponding relationship triples. The experiments on the CoNLL04 and ADE datasets show that compared with other relation extraction baseline models,the F1 value of the proposed model has a better improvement. And it also verifies the effectiveness and generalization ability of the proposed model.

Key words:entity recognition;relationship extraction;deep learning;pre-trained model;multi-fork decoding tree;graph neural network

0 引 言

作为自然语言处理领域的一个重要任务,实体关系抽取受到了广泛的关注和研究。早期,基于规则[1]或本体[2]的实体关系抽取方法过度依赖行业专家进行的大规模的模式匹配规则,在跨领域中可移植性差,极大耗费了人力物力。后来,随着传统机器学习的发展,以统计学为基础的机器学习方法显著提高了实体关系抽取的召回率和跨领域能力。

近些年,随着深度学习的发展,实体关系抽取的性能得到了极大的提升[3-5]。自从 Hinton G 等[6]提出深度学习方法以来,研究人员将深度学习应用到实体关系抽取任务中,取得了相当优异的效果。2012 年,Socher R 等[7]运用循环神经网络[8](Recurrent Neural Networks,RNN)进行实体关系抽取,并在该任务中融入了句子特征,然其忽略了位置信息的重要性;Zhang 等[9]使用循环神经网络时虽然引入了位置信息,但循

收稿日期:2022-07-24 **修回日期**:2022-11-24
基金项目:国家自然科学基金(61070015)
作者简介:张 鑫(1996-),男,CCF 会员(F7689G),硕士研究生,研究方向为自然语言处理;通讯作者:冼广铭(1975-),男,博士,副教授,CCF 会员(12559M),研究方向为人工智能。

环神经网络存在远距离依赖问题;Zeng 等[10]首次使用卷积神经网络(Convolutional Neural Networks, CNN)进行关系抽取,提取出更加丰富的特征,但是由于受到卷积核大小的影响,无法很好地提取出语义特征;李青青等[11]设计一种 Attention 机制的多任务模型,通过共享信息编码提升实体关系抽取的性能;李卫疆等[12]使用 Bi-RNN 解决词与词之间的依赖关系,并融入位置、语法、句法和语义信息进行实体关系抽取;Zheng 等[13]将 BiRNN 和 CNN 模型融合成为联合抽取模型,通过共享 BiLSTM 的编码层,运用 LSTM 与 CNN 进行解码,解决了信息冗余问题,然而都无法很好地解决复杂实体中的实体重叠问题。并且以上所有方法由于使用单解码模型,都受到单解码强行执行某种顺序而带来的局限性。

随着预训练模型的兴起,以往的循环神经网络和卷积神经网络等逐渐淡去,新的神经网络 Transformer[14]逐渐受到广大研究者的青睐。对于 Transformer 等预训练模型,由 Tang 等[15]实验结果表明,相较循环、卷积等模型,Transformer 在综合特征提取能力和语义表征能力上有较高提升。而近期,基于 Transformer 的模型 BERT(Bidirectional Encoder Representation from Transformers)[16]凭借其合理的设计,在自然语言处理领域取得了重大突破,同时也备受实体关系抽取研究人员的青睐。其中 Wei 等[17]和 Fan 等[18]利用 BERT 预训练模型在关系抽取领域取得不错的效果。

针对实体关系抽取中的依赖信息不足、重叠实体获取效果低下、解码顺序问题进行研究,主要贡献为:一是使用 GCN 融入语句的句法特征,增强相关实体与关系之间的依赖性;二是使用基于 span 的方法,筛选出对应的实体信息;三是利用深度多叉解码树进行解码操作,最终得到相应的关系三元组。

1　模型框架

该文提出的基于 Span[19]和 DMFDT 的实体关系抽取模型主要由四个部分组成,即使用预训练模型 BERT 编码层、基于 GCN[20]的句子依赖性增强、基于 Span 的实体获取、深度多叉解码树抽取关系三元组。模型的实现流程如图 1 所示。

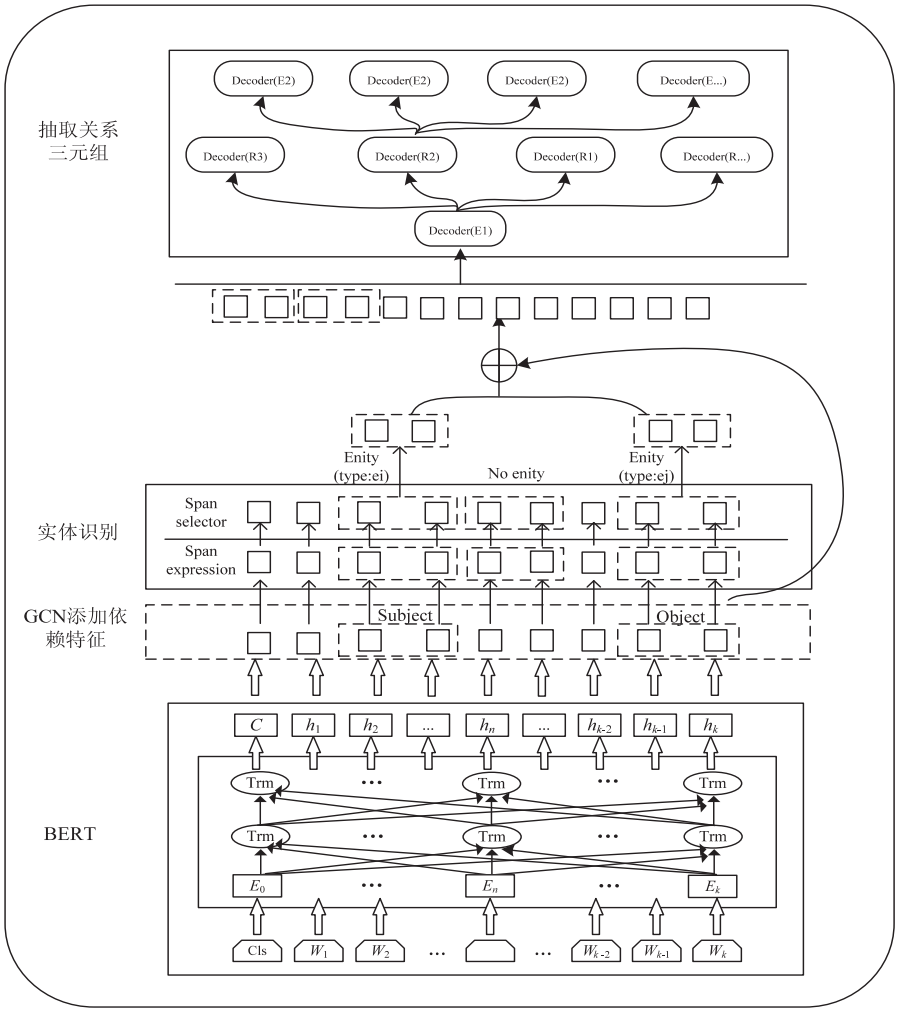

图 1　基于 Span 和深度多叉解码树的实体关系抽取模型框架

1.1 BERT 编码层

BERT 预训练模型主要包含输入层和多层编码层,其中输入层由位置编码(PosEmbedding)和词编码(Token Embedding)等组成,并为其添加[CLS]和[SEP]标志位,也正因如此使得 BERT 具有较强的综合特征提取能力和语义表征能力。该文使用 BERT 编码器提取句子特征,为下游任务提供后续的 H(隐藏层向量)。公式如下:

$$H = \mathrm{BERT}(W) \tag{1}$$

其中,$W = \{w_1, w_2, \cdots, w_n\}$ 为输入的句子;$H = [h_1, h_2, \cdots, h_n]$ 为相应位置词对应输出的隐藏层向量。

1.2 基于 GCN 的句子依赖性增强

对于 BERT 模型输出的隐藏层向量 H,通过融合句法特征可以很好地利用词与词之间的依赖信息,然而在过往的实体关系抽取模型中往往忽略实体与关系本身在句法上的联系。该文综合前人的研究成果,将可以有效利用句法依赖信息的图神经网络(GCN)融入其中,以提高模型的抽取性能。对于给定的句子,使用 Stanford CoreNLP 工具生成相应的句法依存树[21],然后使用 GCN 运行依赖关系图,将相应的依赖信息融合到编码中。

1.2.1 使用 GCN 标记图

对于依赖关系图,$G = (N, \mathrm{Ed})$,其中 N 和 Ed 表示节点(Node)和它们之间边(Edges)的集合。将任意的隐藏向量 h_i $(1 < i < n)$ 作为一个节点,其中边代表由第 u 个节点 $h_u \in N$ 到第 v 个节点 $h_v \in N$ 的依赖关系,使用 $L_{uv} \in \mathrm{Ed}$ 表示,从而节点 u 到节点 v 表示为 (h_u, h_v, L_{uv})。在使用 GCN 时,只考虑相邻节点的情况下,得到一个新的第 u 到 v 节点融合依赖边信息的隐藏 v 节点 h_{v_side}:

$$h_{v\_\mathrm{side}} = F\left(\sum_{u \in N(v)} (w_{Luv} h_u + b_{Luv})\right) \tag{2}$$

其中,w_{Luv} 为权重,b_{Luv} 为偏置,F 为非线性激活函数。

1.2.2 边处理

在生成的图中,可能会存在错误的边需要丢弃。因此,需要对生成的每一条边进行打分。通过打分,高分的边得到保留,低分的边被丢弃。用于计算最终边 (h_u, h_v, L_{uv}) 取舍的公式为:

$$T_{uv} = \mathrm{Score}(h_u w_{Luv} + b_{Luv}) \tag{3}$$

其中,w_{Luw} 为权重,$b_{Luw} = \sqrt{b^2 - 4ac}$ 为偏置,$\mathrm{Score}(\cdot)$ 为 Sigmoid 函数。

通过保留有效边,得到含有句法信息的隐藏 v 节点表示 h_{v_side}:

$$h_{v\_\mathrm{side}} = F\left(\sum_{u \in N(v)} (T_{uv} * (W_{Luc} h_u + b_{Luv}))\right) \tag{4}$$

由此,对于每一个词编码 h_i,通过 GCN 的编码为 h_i^{Gcn},最终得到增强依赖的隐藏表示为:

$$h_i^{Gcn} = F\left(\sum_{u \in N(v)} (T_{uv} * (W_{Luc} h_u^{Gcn} + b_{Luv}))\right) \tag{5}$$

最后,将通过 GCN 处理后得到的依赖信息编码添加到 BERT 的输出层中,得到最终的 token 表达 $h_i^{Gcn\_S}$。

$$h_i^{Gcn\_S} = [h_i^{Gcn}; h_i] EC^s \tag{6}$$

将所有的边信息和 BERT 输出连接生成新的序列表达 h^{Gcn_s}。

1.3 基于 Span 的实体获取

随着基于 Span 的实体关系抽取方法被提出,极大提升了过往基于 BIO/BILOU 标签方法[22]对重叠实体提取的效果。如在“food poisoning”中识别出“food”。基于 Span 的方法:任意标记的子序列(Span)都作为潜在的实体,例如:“Jones was diagnosed with food poisoning”中{“Jones”,“Joneswas”,…,“food”,“food poisoning”}等均可以作为潜在实体。通过 Span selector 得出相应的实体。基于 Span 的实体获取流程如图 2 所示。

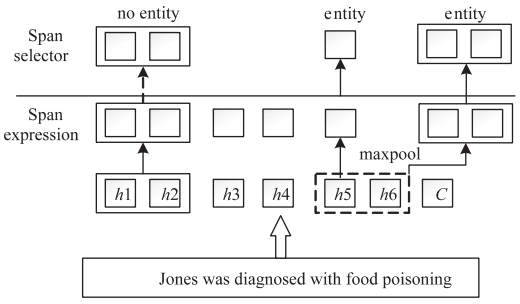

图 2 基于 Span 实体获取流程

1.3.1 Span 表示

对于任意输入的标记子序列(Span)作为 Span 选择器的输入。令标记序列 $S = (h_i^{Gcn\_s}, h_{i+1}^{Gcn\_s}, \cdots, h_{i+k}^{Gcn\_s})$(其中 $0 < i + k < n$ 且 $k < 4$,设定实体最多不会超过 4 个词)作为一个候选 Span。对于存在的预定义实体类集合,假设为 E,例如:“人”“组织”“地名”等实体类型。对于候选标记序列 S,Span 分类器将 S 映射到集合 EU{none}中的一类。none 表示不构成实体跨度。若 S 属于 E,则 S 为实体,否则为 none,不为实体。

使用融合函数 g 生成候选 Span 表达式。且在已有研究上发现最大池化的效果最好。

$$E(s) = g(h_i^{Gcn\_s}, h_{i+1}^{Gcn\_s}, \cdots, h_{i+k}^{Gcn\_s}) \tag{7}$$

根据前人研究,连接标记 C(cls,表达上下文语义信息)对实体类型表达具有强有力的作用。因此,融入标记 C。表达式如下:

$$X^s = E(s) \circ C \tag{8}$$

其中,X^s 作为 Span 的最终表达,$\circ$ 表示连接符号。

1.3.2 Span 分类器

将得到的 Span 表达输入到 softmax 分类器中,得到相应的实体类型。

$$EC^s = \mathrm{softmax}(X^s) \tag{9}$$

其中, EC^s 表示实体类型得分。

通过查看得分最高的类,得到 Span 分类器输出的每一个 Span 属于那一个类型的实体或者非实体类。通过这种方法,过滤掉 none 类型的 Span,留下一组 Spans,令它们构成实体 $X_i^E \in E$。从而避免了之前序列标注过程中重叠实体无法识别的现象。

将构成实体的 Spans 连接,表达相应的实体集合:

$$X = \sum_{1}^{n} [X_1^E; X_2^E; X_3^E; \cdots] \tag{10}$$

然后将增强依赖性的上下文语句信息与实体信息进行连接,作为解码部分的输入。

$$h^{\text{concat}} = [X; h^{Gcn\_s}] \tag{11}$$

1.4　深度多叉解码树抽取关系三元组

深度多叉解码树一改以往单解码的方式,运用多解码的方式更好地解决了单解码带来的模型容易记忆和过度拟合训练集中频繁出现的三元组顺序的问题。通过并行解码三元组(Triad),很好地解决了三元组执行顺序问题。例如:"Qian Xuesen was born in Shang hai, and graduated from Massachusetts nstitute of Technology andCalifornia Institute of Technology."。如果解码时三元组的执行顺序为{Triad1, Triad2, Triad3}(单解码执行顺序如图 3 所示)。但是{Triad2, Triad1, Triad3}{Triad2, Triad3, Triad1}也是正确的。然而对于单解码情况下只能执行其中一种顺序。并且对于单解码情况,若出现如{Triad2, Triad3}这种顺序的数据,由于训练出来的模型将会高度拟合{Triad1, Triad2, Triad3}这种顺序,从而直接由 Triad3 结束,进而忽略了 Triad1,在后期的应用中出现无法拟合三元组 1 的现象。因而,放弃以往的单解码的方式,使用多叉解码树的方式,解码关系三元组顺序如图 4 所示。

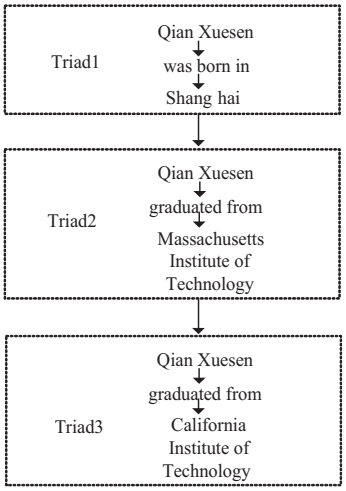

图 3　单解码关系强制顺序执行

在解码期间,对于不同的解码层使用相同的输入编码,不同的输出层。使用在 Span 方法处理中得到的序列 h_v^{concat} 作为解码层的输入序列。

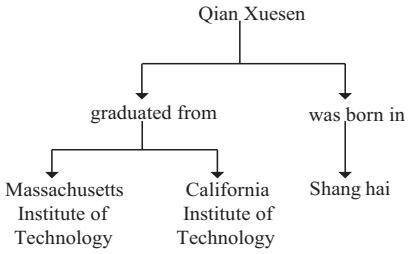

图 4　多叉解码树关系执行顺序

对于实体 1 解码层,从含有实体表达 h_v^{concat} 中直接解码选择出实体 1。先在序列上进行线性变换,后再做最大池化 Max(·) 操作。

$$e_1 = \text{sig}(\text{Max}(h^{\text{concat}} w_{e_1} + b_{e_1})) \tag{12}$$

其中, w_{e_1} 为权重, b_{e_1} 为偏置, sig 为 sigmoid 函数。

对于关系解码层,将中间表达向量 h_v^{concat} 输入到关系预测层,通过在整个序列上进行两层二分类操作,来预测关系的开始位置($\text{Possible}_{\text{rs}}$)和结束位置($\text{Possible}_{\text{re}}$),得到相应的关系:

$$\text{Possible}_{\text{re}} = \text{Sig}(h^{\text{concat}} w_r + b_r) \tag{13}$$

$$\text{Possible}_{\text{rs}} = \text{Sig}(h^{\text{concat}} w_r + b_r) \tag{14}$$

其中, w_r 为权重, b_r 为偏置。

对于实体 2 解码层与实体 1 类似,在整个序列上预测实体 1 和关系 R 对应的实体 2,最终得到构成的三元组。

$$e_2 = \text{Sig}(\text{Max}(h^{\text{concat}} w_{e_2} + b_{e_2})) \tag{15}$$

其中, w_{e_2} 为权重, b_{e_2} 为偏置。

1.5　损失函数

损失函数作为评价模型输出值(预测值)与真实目标的相似程度。因此,选择合适的损失函数对模型的性能来说至关重要,不同的模型损失函数一般也不一样。通过对比多个损失函数的,该模型最终使用 Hinge 损失函数,它的标准形式如下所示:

$$L(y, f(x)) = \max(0, 1 - yf(x)) \tag{16}$$

其中, $f(x)$ 是预测值,Hinge 损失函数的特点:使用 Hinge 损失函数如果被分类正确,损失为 0,否则损失为 $1 - yf(x)$; $f(x)$ 在-1 到 1 之间,使得分类器并不过度打分,让某个正确分类的样本距离分割线超过 1 并不会有任何奖励,从而使得分类器更加专注于整体的误差, y 是目标值(-1 或 1);具有较高的健壮性。

该模型使用的 Hinge 损失函数为 MultiMarginLoss,公式如下:

$$\text{Loss}(x, y) = \frac{\sum_i \max(0, w[y] * (\text{margin} - x[y]))^p}{\text{X.size}(0)} \tag{17}$$

其中, x 为神经网络的输出, y 是真实的类别标签, w 为每一类可传入相应的权值,margin 默认为 1。

2 实 验

2.1 数据集

为了测试该模型的性能,使用实体关系抽取领域公开数据集 CoNLL04、ADE 进行实验。CoNLL04 数据集包含了从新闻中提取出来的带有注释的命名实体和关系的句子。其包含 LOC、ORG、PER、OTHERS 四种实体类型和 Located in、Work for、Organization based in、Live in、Kill 五种关系。ADE 数据集包含了从医学报告中提取出的 4 272 个句子和 6 281 个关系。它包含了一种单一的关系 Adverse Effect 和两种实体类型 Adverse Effect、Drug。

2.2 实验环境与参数设置

实验采用 Pytorch 框架,使用谷歌云盘和谷歌 Colab 作为实验环境,使用 Python 编程语言。

在模型训练过程中,将 batch_size 设置为 32;根据数据集合理设置语句的长度(max_length);根据日常训练过程中损失函数的收敛情况,设置失活率(drop out)和学习率(learning rate)分别为 0.1、1e-4;设置 hidden_size 为 768。使用 Multi-MarginLoss 损失函数和 Adam(Adaptive Moment Estimation)算法优化模型参数。

2.3 基线模型与评价指标

将所提模型与目前该领域主流基线模型进行比较。其中用于比较的模型如下:

(1)Global Optimization[23]:将双向 LSTM 和全局优化结合在一起,命名实体识别和关系抽取同时进行。

(2)Multi-turn QA[24]:将实体关系抽取任务转换为多问答任务,即将实体和关系的提取转换为从上下文识别答案跨度的任务。

(3)Multi-head+AT[25]:将对抗训练应用到联合实体关系抽取模型当中。

(4)Relation-Metric[26]:结合 CNN(卷积神经网络)和 metric learning(度量学习)的思想应用到端到端的关系抽取任务中。

(5)Biaffine Attention[27]:提出了一种端到端神经网络的抽取模型。其采用 BiLSTM-CRF 体系结构进行实体识别,使用双注意力机制的关系分类。

(6)Replicating Multihead with AT[28]:使用 CRF 将实体识别和关系抽取任务建模为多头选择任务。

(7)SpERT[29]:一种基于预训练模型 BERT 和 Spaner 的实体关系联合抽取模型,是一种和文中同样使用 BERT 和 Span 方法的关系抽取模型,但相对于文中模型未使用依赖增强和多叉解码方法。

其中 Multi-head+AT、Biaffine Attention、RMWA 模型使用不同的 RNN-CRF 方法进行关系抽取;SpERT 使用基于 BERT 和 Span 的方法进行关系抽取;模型 Global Optimization、Multi-turn QA、RelationMetric 使用关系抽取的其他方法进行关系抽取。

为了评估文中模型的优劣性,实验结果主要采用准确率 P(precision)、召回率 R(recall)以及 F1 进行评估。其中计算准确率、召回率、F1 值之前,首先要得到 TP(预测为真且实际正确的样本数)、FN(预测为假且实际正确的样本数)、FP(预测为真且实际为假的样本数)、TN(预测为假且实际为假的样本数)。然后进行如下计算:

$$P = \frac{TP}{TP + FP} \tag{18}$$

$$R = \frac{TP}{TP + FN} \tag{19}$$

$$F1 = \frac{2 * P * R}{P + R} \tag{20}$$

2.4 实验结果与分析

为了更好地展现文中模型的效果,在数据集 CoNLL04、ADE 上进行了对比实验和消融实验。

2.4.1 模型对比实验分析

文中模型以 BERT 模型作为词编码层,基于 Span 的方式获取实体信息,并通过 GCN 增强词与词之间的依赖性,最后运用深度为多叉解码树对实体关系三元组进行解码,得到实体与对应的关系。损失函数值与训练次数如图 5 所示,可以直观地看到,随着训练次数的增加,模型的 loss 逐渐减少,当 loss 值分别为 0.3、0.28 左右时,模型在 CoNLL04、ADE 数据集上趋于收敛。

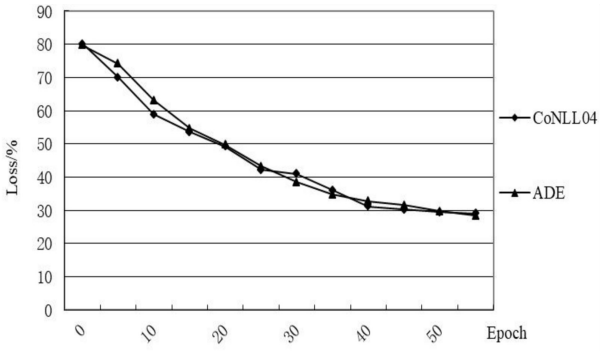

图 5 损失函数值与训练次数变化

图 6 展示的为文中模型、Global Optimization 模型、SpERT 模型在数据集 CoNLL04 与 ADE 上 F1 值随着训练次数增加的变化曲线。可以看到,在 CoNLL04 数据集上 Global Optimization 模型的最优 F1 值明显小于文中模型和 SpERT,由此说明了基于预训练模型和 Span 方法的优越性。在数据集 CoNLL04、ADE 上,对于 SpERT 模型,虽然也取得了不错的效果,但相比文中模型 F1 值分别低了 3.13 百分点、1.54 百分点,由此可见文中模型使用依赖增强和多叉解码树的优

越性。

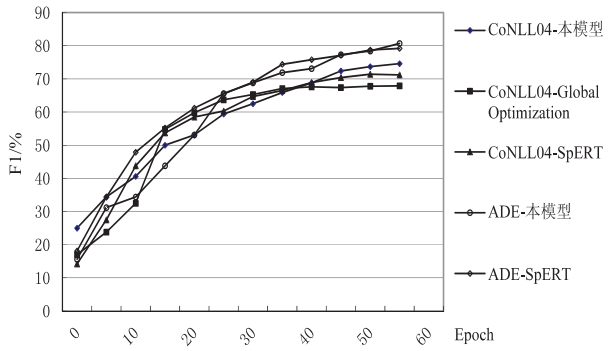

图 6　文中模型、Global Optimization、SpERT 模型的 F1 与迭代次数变化

由表 1 中的实验结果对比可知,在 Co-NLL04 数据集上,文中模型相比 Multi-head+AT、Biaffine Attention、RMWA 等 RNN-CRF 类型网络模型的 F1 值分别高出 12.57 百分点、10.2 百分点、12.72 百分点,说明使用预训练模型和 Span 方法要明显优于使用 RNN-CRF 方法的模型。通过文中模型与 Global Optimization、Multi-turn QA、RelationMetric 等其他关系抽取方法对比结果可知,文中模型在 CoN-LL04 数据集上效果要比以上三种模型分别高出 6.7 百分点、5.7 百分点、12.56 百分点,在数据集 ADE 上文中模型也要比 Multi-turn QA、Rel-ationMetric 模型分别高出 5.18 百分点和 3.14 百分点,由此也说明了即使在一些优化方法的作用下,文中模型依然具有相当大的优势。根据文中模型与 SpERT 模型的实验结果对比可以看出,文中模型使用 GCN 增强依赖性下的多叉解码树方法的提取效果在 CoNLL04、ADE 数据集上要比 SpERT 模型高出 3.14 百分点、1.46 百分点(GCN、多叉解码树各自优势见消融实验)。由此可见,文中模型具有极好的优越性。

表 1　实体关系抽取对比实验结果

Dataset	Model	Precision/%	Recall/%	F1/%
CoNLL04	Global Optimization	-	-	67.90
	Multi-turn QA	69.20	68.20	68.90
	Multi-head+AT	-	-	62.03
	Relation-Metric	63.75	60.43	62.04
	Biaffine Attention	-	-	64.40
	RMWA	65.81	57.59	61.88
	SpERT	73.04	70.00	71.47
	文中模型	70.28	79.50	74.60
ADE	Multi-head+AT	-	-	75.52
	Relation-Metric	63.75	60.43	77.29
	SpERT	78.09	80.43	79.24
	文中模型	81.20	80.20	80.70

文中模型性能较优主要因为使用实体提取能力更强的 Span 方法和更具优越性的多叉解码方法,以及使用了综合特征能力提取更强的预训练模型和 GCN 进行词与词之间依赖增强。

2.4.2　消融实验与分析

为了验证提出的多叉解码树和使用 GCN 融入依赖特征的效果,做了以下消融实验:

(1) Baseline (SpERT):不使用多叉解码树和 GCN,使用单解码的方式也即模型 SpERT 的效果。

(2) Self-GCN:文中模型只进行依赖增强但仍使用单解码的方式。

(3) Self-DMFDT:不融入依赖增强信息,将单解码换为多叉解码树进行解码。

(4) Self-GCN-DMFDT(文中模型):将 2 中的单解码的方式换为多叉解码树的方式。

从表 2 可见,在 CoNLL04、ADE 数据集上,Self-GCN 相比于 Baseline (SpERT)模型分别提升 1.06 百分点和 0.09 百分点,由此证明了通过 GCN 进行词与词之间的依赖增强,模型的性能可以得到一定的提升。而 Self-DMFDT 相比于 Baseline(SpERT)模型的实验结果证明,通过使用多叉解码树进行并行解码在两种数据集上可以使模型性能提升 1.88 百分点、0.9 百分点,由此也证明了多叉解码树的优越性。对于融合了两种方法的模型也即文中模型相比 Baseline (SpERT) 在两种数据集上达到了 3.13% 和 1.54%。由此证明使用依赖增强和多叉解码方法对模型的效果有一个不错的提升。

表 2　消融实验结果

Dataset	Model	Precision/%	Recall/%	F1/%
CoNLL04	Baseline(SpERT)	73.04	70.00	71.47
	Self-GCN	74.50	70.66	72.53
	Self-DMFDT	73.65	73.05	73.35
	Self-GCN-DMFDT (文中模型)	70.28	79.50	74.60
ADE	Baseline(SpERT)	78.09	80.43	79.24
	Self-GCN	80.10	78.57	79.33
	Self-DMFDT	80.06	80.22	80.14
	Self-GCN-DMFDT (文中模型)	81.20	80.20	80.70

3　结束语

该文提出一种基于 Span 方法和深度多叉解码树的实体关系抽取模型。该模型通过利用句法依赖信息提升下游实体识别和关系抽取的效果。下游任务中基于 Span 方法更好地获取重叠实体的同时,使用多叉解

码树成功地避免了单解码的执行顺序问题。实验结果表明,该方法在 CoNLL04、ADE 数据集上明显比先前方法优越。

参考文献:

[1] POPOVSKI G,KOCHEV S,KOROUSIC-SELJAK B,et al. FoodIE:a rule-based named-entity recognition method for food information extraction[C]//Proceedings of the 8th international conference on pattern recognition applications and methods. Prague:ICPRAM,2019:915-922.

[2] LIMA R,ESPINASSE B,FREITAS F. Ontoilper:an ontology-and inductive logic programming-based system to extract entities and relations from text[J]. Knowledge and Information Systems,2018,56(1):223-255.

[3] QIAO B,ZOU Z,HUANG Y,et al. A joint model for entity and relation extraction based on BERT[J]. Neural Computing and Applications,2022,34(5):3471-3481.

[4] SHANG Y M,HUANG H,MAO X. OneRel:joint entity and relation extraction withone module in one step[C]//Proceedings of the AAAI conference on artificial intelligence. Vancouver:AAAI,2022:11285-11293.

[5] 何阳宇,易晓宇,唐 亮,等. 基于 BLSTM-ATT 的老挝语军事领域实体关系抽取[J]. 计算机技术与发展,2021,31(5):31-37.

[6] HINTON G E,SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science,2006,313(5786):504-507.

[7] SOCHER R,HUVAL B,MANNING C D,et al. Semantic compositionality through recursive matrix-vector spaces[C]//Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning. Jeju Island:EMNLP,2012:1201-1211.

[8] 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.

[9] ZHANG Dongxu,WANG Dong,LIU Rong. Relation classification via recurrent neural network[R]. Beijing:Tsinghua University,2015.

[10] ZENG D,LIU K,LAI S,et al. Relation classification via convolutional deep neural network[C]//Proceedings of COLING 2014,the 25th international conference on computational linguistics:technical papers. Dublin:COLING,2014:2335-2344.

[11] 李青青,杨志豪,罗 凌,等. 基于多任务学习的生物医学实体关系抽取[J]. 中文信息学报,2019,33(8):84-92.

[12] 李卫疆,李 涛,漆 芳. 基于多特征自注意力 BLSTM 的中文实体关系抽取[J]. 中文信息学报,2019,33(10):47-56.

[13] ZHENG S,HAO Y,LU D,et al. Joint entity and relation extraction based on a hybrid neural network[J]. Neurocomputing,2017,257:59-66.

[14] VASWANI A,SHAZEER N,PARMAR N,et al. Attention is all you need[C]//Proceedings of the 31st international conference on neural information processing systems. California:NIPS,2017:6000-6010.

[15] TANG G,MÜLLER M,GONZALES A R,et al. Why self-attention? a targeted evaluation of neural machine translation architectures[C]//Proceedings of the 2018 conference on empirical methods in natural language processing. Brussels:Association for Computational Linguistics,2018:4263-4272.

[16] KENTON J D M W C,TOUTANOVA L K. BERT:pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of NAACL-HLT. Minneapolis:NAACLHLT,2019:4171-4186.

[17] WEI Z,SU J,WANG Y,et al. A novel cascade binary tagging framework for relational triple extraction[C]//Proceedings of the 58th annual meeting of the association for computational linguistics. Washington:Association for Computational Linguistics,2020:1476-1488.

[18] FAN J,LIU X,DONG S,et al. Enriching pre-trained language model with dependency syntactic information for chemical-protein interaction extraction[C]//China conference on information retrieval. [s. l.]:Springer,2020:58-69.

[19] ZUO M,ZHANG Y. A span-based joint model for extracting entities and relations of bacteria biotopes[J]. Bioinformatics,2022,38(1):220-227.

[20] VASHISHTH S,JOSHI R,PRAYAGA S S,et al. RESIDE:improving distantly-supervised neural relation extraction using side information[C]//Proceedings of the 2018 conference on empirical methods in natural language processing. Brussels:EMNLP,2018:1257-1266.

[21] 温 政,段利国,李爱萍. 基于最短依存路径与神经网络的关系抽取[J]. 计算机工程与设计,2019,40(9):2672-2676.

[22] 郑余祥,左祥麟,左万利,等. 基于时间关系的 Bi-LSTM+GCN 因果关系抽取[J]. 吉林大学学报:理学版,2021,59(3):643-648.

[23] ZHANG M,ZHANG Y,FU G. End-to-end neural relation extraction with global optimization[C]//Proceedings of the 2017 conference on empirical methods in natural language processing. Copenhagen:EMNLP,2017:1730-1740.

[24] LI X,YIN F,SUN Z,et al. Entity-relation extraction as multi-turn question answering[C]//Proceedings of the 57th annual meeting of the association for computational linguistics. Florence:Association for Computational Linguistics,2019:1340-1350.

[25] BEKOULIS I,DELEU J,DEMEESTER T,et al. Adversarial training for multi-context joint entity and relation extraction[C]//EMNLP2018,the conference on empirical methods in

(下转第 166 页)