

BGPNRE:一种基于BERT的全局指针网络实体关系联合抽取方法

邓亮, 齐攀虎, 刘振龙, 李敬鑫, 唐积强

引用本文

邓亮, 齐攀虎, 刘振龙, 李敬鑫, 唐积强. BGPNRE:一种基于BERT的全局指针网络实体关系联合抽取方法[J]. 计算机科学, 2023, 50(3): 42-48.

DENG Liang, QI Panhu, LIU Zhenlong, LI Jingxin, TANG Jiqiang. BGPNRE:A BERT-based Global Pointer Network for Named Entity-Relation Joint Extraction Method [J]. Computer Science, 2023, 50(3): 42-48.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于BERT和多特征融合嵌入的中文拼写检查](#)

Chinese Spelling Check Based on BERT and Multi-feature Fusion Embedding
计算机科学, 2023, 50(3): 282-290. <https://doi.org/10.11896/jsjcx.220100104>

[基于联合模型的端到端事件可信度识别](#)

End-to-End Event Factuality Identification with Joint Model
计算机科学, 2023, 50(2): 292-299. <https://doi.org/10.11896/jsjcx.211200108>

[EGOS-DST:对话现象感知和模式引导的一步对话状态追踪算法](#)

EGOS-DST:Efficient Schema-guided Approach to One-step Dialogue State Tracking for Diverse Expressions
计算机科学, 2022, 49(11A): 210900246-7. <https://doi.org/10.11896/jsjcx.210900246>

[基于论据边界识别的立场分类研究](#)

Stance Detection Based on Argument Boundary Recognition
计算机科学, 2022, 49(11A): 210800180-5. <https://doi.org/10.11896/jsjcx.210800180>

[基于深度学习与文本计量的技术趋势分析](#)

Analysis of Technology Trends Based on Deep Learning and Text Measurement
计算机科学, 2022, 49(11A): 211100119-6. <https://doi.org/10.11896/jsjcx.211100119>

BGNPRE:一种基于 BERT 的全局指针网络实体关系联合抽取方法

邓亮^{1,2,3} 齐攀虎⁴ 刘振龙⁴ 李敬鑫⁴ 唐积强⁵

1 中国科学院大学 北京 100049
2 中国科学院沈阳计算技术研究所 沈阳 110168
3 国家知识产权局专利局 北京 100083
4 中国科学院计算技术研究所 北京 100190
5 国家计算机网络应急技术处理协调中心 北京 100029
(deng_sipo@msn.com)

摘要 实体-关系联合抽取指从非结构化文本中联合抽取实体-关系三元组,是信息抽取和知识图谱构建的一项关键任务。文中提出了一种新的基于全局指针网络实体关系联合抽取方法 BGNPRE(BERT-based Global Pointer Network for Named Entity-Relation Joint Extraction),首先通过潜在关系预测模块预测文本中蕴含的关系,过滤掉不可能存在的关系,将实体抽取限制在预测的关系子集中;其次通过使用基于关系的全局指针网络,获取所有主客体实体的位置;最后通过全局指针网络通信模块,将主客体位置高效率地解码对齐成一个实体关系三元组。该方法避免了传统管道式方法存在的错误传播问题,同时也解决了关系冗余、实体重叠、Span 提取泛化不足等问题。实验结果表明,所提方法在多关系和重叠实体抽取上表现卓越,并且在 NYT 和 WebNLG 公共数据集上达到了最先进的水平。

关键词: 实体-关系联合抽取;BGNPRE;全局指针网络;BERT

中图法分类号 TP391

BGNPRE: A BERT-based Global Pointer Network for Named Entity-Relation Joint Extraction Method

DENG Liang^{1,2,3}, QI Panhu⁴, LIU Zhenlong⁴, LI Jingxin⁴ and TANG Jiqiang⁵
1 University of Chinese Academy of Sciences, Beijing 100049, China
2 Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China
3 China National Intellectual Property Administration, Beijing 100083, China
4 Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
5 National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing 100029, China

Abstract Named entity-relation joint extraction refers to extracting entity-relation triples from unstructured text. It's an important task for information extraction and knowledge graph construction. This paper proposes a new method--BERT-based global pointer network for named entity-relation joint extraction(BGNPRE). Firstly, the potential relation prediction module is used to predict the relations contained in the text, filters out the impossible relations, and limits the predicted relation subset for entity recognition. Then a relation-specific global pointer-net is used to obtain the location of all subject and object entities. Finally, a global pointer network correspondence component is designed to align the subject and object position into named entity-relation triples. This method avoids error propagation from pipeline model, and also solves the redundancy of relation prediction, entity overlapping, and poor generalization of span-based extraction. Extensive experiments show that our model achieves state-of-the-art performance on NYT and WebNLG public benchmarks with higher performance gain on multi relations and entities overlapping.

Keywords Named entity-relation joint extraction, BGNPRE, Global pointer network, BERT

1 引言

作为自然语言处理领域的一项基础核心技术,实体-关系抽取被广泛应用到知识图谱构建、关系抽取等任务中。实体-关系抽取指从非结构化文本中抽取〈主体(Subject),关系(Relation),客体(Object)〉形式的实体-关系三元组。目前,实体-关系抽取主要采用联合抽取方法,其可以通过共享实体和关系抽取参数来提高模型性能,并且避免了错误传播问题。

然而,当前方法仍面临诸多挑战。首先,对于多任务联合抽取学习方法,如 PRGC,关系三元组抽取分 3 个阶段进行,一方面将全局通信模块应用到特定关系的主客体实体抽取之后,会造成主客体对齐中的偏差暴露,另一方面基于序列标注的提取方案只关注非重叠实体的局部位置特征,导致泛化效果较差。同时,由于其主客体对齐机制不足,解码效率低,因此仅限于一次处理一个主体。其次,对于单任务学习的端到端联合抽取方法,如 OneRel,为了避免主客体对齐中的偏差,该方法利用了一个单阶段指针网络和一个相当复杂的单步解码器,但这导致了稀疏标签和低收敛率的问题,而且 Span 的提

取泛化能力仍然不足。

为了解决上述问题,本文提出了一种新的基于 BERT 的全局指针网络实体关系联合抽取方法 BGPNRE。该模型采用 BERT 作为编码器,使用潜在关系预测、特定关系全局指针网络识别和全局指针网络通信 3 个组件进行实体-关系三元组联合抽取。

实验结果表明,本文方法在两个英文公共数据集上以更高的效率优于目前最先进的方法。特别地,在单实体重叠(SEO)、实体对重叠(EPO)的复杂场景上(见表 1)的实验结果表明 BGPNRE 具有一定的优势。

表 1 Normal EntityPairOverlap(EPO)及 SingleEntityOverlap(SEO)重叠模式的样例

Table 1 Examples of Normal, EntityPairOverlap(EPO) and SingleEntityOverlap(SEO) Modes

	Texts	Triplets
Normal	The [United States] President [Joe Biden] will visit [Paris],[France].	(United States,president,Joe Biden) (France,contains,Paris)
SEO	The ruins are in the upper of fringes of the [Tigris] and Euphrates Valleys,near the [Iraq] border and with-in sight of Taurus Mountains of southern [Turkey].	(Turkey,contains,Tigris) (Iraq,contains,Tigris)
EPO	[Madrid] is a capital city of [Spain].	(Spain,contains,Madrid) (Spain,capital city,Madrid)

本文的主要贡献如下:

(1)提出了一个新的实体关系抽取的子任务分解方法,即关系判断、主客体实体抽取和主客体实体对齐。

(2)提出了一个新的端到端框架,并针对子任务设计了 3 个组件,大大缓解了冗余关系判断、实体重叠、Span 提取泛化能力不足等问题。

(3)在两个英文公共数据集上进行了广泛的实验,实验结果表明本文方法达到了最先进的性能,特别是在重叠三元组的复杂场景中。我们也进行了消融实验的研究和分析,证实了模型中每个组件的有效性。

本文第 2 节描述了相关工作;第 3 节介绍了本文方法;第 4 节详述了实验结果并对其进行讨论;最后总结全文。

2 相关工作

实体-关系抽取的主要方法分为流水线方法和联合抽取方法两类。流水线方法将任务分成命名实体识别^[1-2]和关系抽取^[3-6]两个独立的子任务,虽然灵活但忽略了两者之间的相关性,且易出现错误传播的情况。为了建立这两个任务之间的相关性,联合抽取模型引起了广泛的关注。先前基于特征的联合抽取模型^[7-10]需要一个复杂的特征工程,并依赖于各种 NLP 工具和繁琐的手工操作。

近年来,为了减少人工的参与,神经网络模型成为了研究的热点。Zheng 等提出了一种新的标注方案^[11],该方案统一了标注实体的角色和实体之间的关系,从而将联合提取任务转换为序列标注任务,但它未能解决重叠的问题。Bekoulis 等提出首先提取所有候选实体^[12],然后将每个实体对之间的关系预测为一个多头选择问题,它共享参数,但不联合解码。Nayak 等采用编码器-解码器架构和基于指针网络的解码方法^[13],在每个时间步长中都产生了一个完整的三元组。

为了解决实体重叠的问题,Wei 等提出了一个级联框架 CasRel^[14],首先在一个句子中识别所有可能的主体实体,

然后对于每个主体实体,应用基于 Span 的标记,基于每个关系来识别相应的客体。该方法导致了关系判断的冗余、基于局部 Span 的实体提取方案不具有鲁棒性。为了表示三元组的关系,Yuan 等提出了一种基于特殊关系的注意力网络^[15],为每个关系下的单词分配不同的权重,但该网络应用了一个启发式最近邻原理来组合实体对,这意味着最近的主体实体和客体实体将被组合成一个三元组,这显然不符合事实。同时,它也造成了关系冗余判断的问题。Wang 等提出了一种单阶段联合抽取框架 TPLinker^[16],该方法采用 token 对预测方案,执行两个 $O(n^2)$ 矩阵操作来提取实体,并将主体与每个句子的每个关系下的客体对齐,分别导致了关系冗余判断问题和主客体对齐的复杂性。为了解决关系冗余的问题,Sui 等提出了一种集合生成网络 SPN^[17],该方法融合句子编码表示和潜在关系表示生成三元组,并通过二位图执行 $O(n^3)$ 矩阵操作匹配出最优三元组,解决了冗余关系判断问题,但导致了主客体对齐的复杂性。Zheng 等提出了一种基于潜在关系和全局通信的三元组联合抽取框架 PRGC^[18],该方法设计了潜在关系预测、特定于关系的序列标注以及全局通信 3 个组件,解决了关系冗余问题,但是重叠实体的泛化能力不足的问题,仍然没有得到有效的解决。Shang 等提出了一种基于单步解码的单模块联合抽取框架 OneRel^[19],该方法通过一个基于评分的分类器和一个基于特定关系的角标记全局指针网络,实现了关系三元组的抽取,解决了关系冗余和实体重叠问题,但是这种全局指针网络对实体长度和跨度的敏感度低,导致了 Span 实体泛化能力不足的问题。

3 模型方法

针对实体-关系三元组抽取任务存在的关系冗余、实体重叠、Span 泛化能力不足的问题,本文利用一个多标签分类器和全局指针网络分别来标注关系、主客体实体在文本中的起始位置对和结束位置对、主体实体在文本中的起始结束位置

对以及客体实体在文本中的起始结束位置对,同时利用多个全局指针网络同时抽取实体对应的关系或者类型。全局指针网络是指针网络^[20]的一个变种,将起始位置作为一个整体进行识别,可以无差别地识别重叠实体和长实体,具有全局性

特征,效率高。受此特点启发,我们采用其解决实体重叠和 Span 泛化能力不足的问题。

模型的主要架构如图 2 所示,本文称该模型为基于 BERT 的全局指针网络(BGPNRE)。

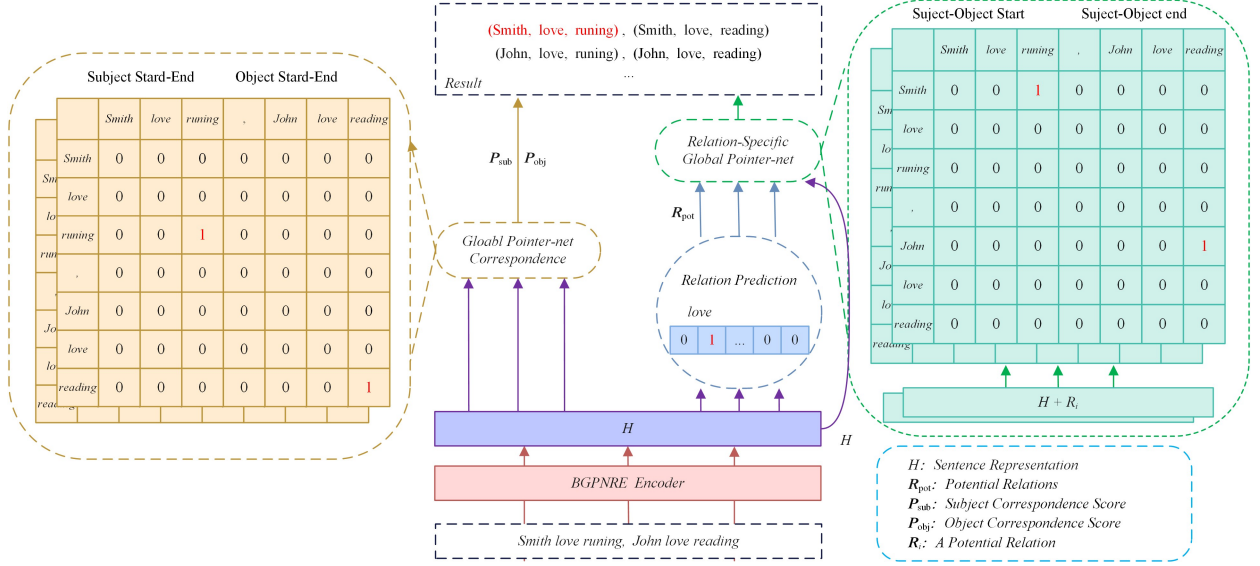


图 2 BGPNRE 的整体架构(电子版为彩图)

Fig. 2 Overall Architecture of BGPNRE

3.1 问题定义

输入是一个带有 n 个 token 的句子 $S = \{x_1, x_2, \dots, x_n\}$ 。期望输出的是关系三元组 $T(S) = \{(s, r, o) | s, o \in E, r \in R\}$, E 和 R 分别是实体和关系集合。本文将问题定义分割成 3 个子任务。

3.1.1 关系判断

对于给定的句子 S , 目标是识别句子中所包含的潜在关系。此任务的输出是 $Y_r(S) = \{r_1, r_2, \dots, r_m | r_i \in R\}$, m 是潜在关系子集的大小。

3.1.2 主客体实体抽取

对于给定的句子 S 和一个预测的潜在关系 r_i , 目标是使用全局指针网络预测主客体起始位置对 (s_h, o_h) 和主客体结束位置对 (s_t, o_t) 的对应得分, 这意味着只有正确的主客体起始位置对和主客体结束位置对才能得到很高的分数, 其他的则会得到很低的分数。 P_h 表示主客体起始位置全局指针网络的结果, P_t 表示主客体结束位置全局指针网络的结果, P_e 表示最终枚举匹配的输出。此任务的输出是 $Y_h(S, r_i | r_i \in R) = P_h$, $Y_t(S, r_i | r_i \in R) = P_t$, 最终通过枚举匹配得到的输出为 $Y_e(P_h, P_t | r_i \in R) = P_e$ 。

3.1.3 主客体对齐

对于给定的句子 S , 目标是预测出所有的主体起始结束位置对 (s_h, s_t) 和客体起始结束位置对 (o_h, o_t) , 将主客体对齐解码成三元组。这意味着只有正确的主体起始结束位置对和客体起始结束位置对才行走得得到很高的分数, 其他的则会得到很低的分数。 E_s 表示主体全局指针网络的矩阵, E_o 表示客体全局指针网络的矩阵。此任务的输出是 $Y_s(S) = E_s \in \mathbf{R}^{n \times n}$, $Y_o(S) = E_o \in \mathbf{R}^{n \times n}$ 。

3.2 模型编码端

本文采用 BERT 预训练语言模型^[18]作为编码器, 如图 2

所示, 输出为 $Y_{enc}(S) = \{h_1, h_2, \dots, h_n | h_i \in \mathbf{R}^{d \times 1}\}$, 其中 d 是嵌入的维度, n 是 token 的数量。BERT 模型采用多层双向的 Transformer 网络结构^[21]来做特征抽取的基本编码单元, 同时引入自注意力机制模块, 减小了对外部知识的依赖, 更擅长捕捉词语或者特征的内部相关性。

3.3 模型解码端

本文提出了一个全新的全局指针网络, 采用 3 个任务的组件作为模型的解码器, 从而联合学习解码出文本中的实体-关系三元组。

3.3.1 全局指针网络

常规的指针网络在进行实体识别时, 一般是用两个模块分别识别实体的起始位置和结束位置, 这会带来训练和预测的不一致。而全局指针网络就是针对这个不一致而设计的, 它将实体起始结束位置和实体类别视为一个整体去进行识别, 因此它更有“全局观”。

全局指针网络是将实体识别分为抽取和分类两个步骤, 抽取指抽取实体的片段, 分类则是确定每个实体的类型。基于此, 抽取相当于只有一种实体类型的实体识别步骤, 我们采用点积型注意力机制^[22]的方式和相对位置 RoPE^[23]来设计打分矩阵, 引入 RoPE 相对位置是为了增加对实体的长度和跨度敏感度, 以更好地分辨出真正的实体。而对于分类, 我们采用特征拼接和全连接层来完成。具体的步骤如下:

$$q_i = w_q \cdot h_i + b_q \quad (1)$$

$$q_j = w_q \cdot h_j + b_q \quad (2)$$

$$k_i = w_k \cdot h_i + b_k \quad (3)$$

$$k_j = w_k \cdot h_j + b_k \quad (4)$$

$$P_{i,j}^a = q_i^T u_i^T u_j k_j \quad (5)$$

$$P_{i,j}^m = w_e^T [q_i; k_i; q_j; k_j] + b_e^T \quad (6)$$

$$\mathbf{P}_{i,j}^{\alpha} = \mathbf{P}_{i,j}^{\alpha_{\text{er}}} + \mathbf{P}_{i,j}^{\alpha_{\text{ss}}} \quad (7)$$

输入是长度为 n 的句子经过编码后得到的矩阵 $[\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n]$, $\mathbf{h}_i \in \mathbf{R}^{d \times 1}$ 是输入句子的第 i 个 token 的编码表示, $\mathbf{h}_j \in \mathbf{R}^{d \times 1}$ 是输入句子的第 j 个 token 的编码表示, $\mathbf{q}_i, \mathbf{q}_j \in \mathbf{R}^{d \times 1}$ 是输入句子第 i 个或者第 j 个 token 的 query 向量, $\mathbf{k}_i, \mathbf{k}_j \in \mathbf{R}^{d \times 1}$ 是输入句子第 i 个或者第 j 个 token 的 key 向量, $\mathbf{u}_i, \mathbf{u}_j \in \mathbf{R}^{d \times 1}$ 是输入句子的第 i 个或者第 j 个相对位置向量, $\mathbf{P}_{i,j}^{\alpha_{\text{er}}}$ 和 $\mathbf{P}_{i,j}^{\alpha_{\text{ss}}}$ 分别是 α 类型的实体抽取打分矩阵和 α 类型的实体分类打分矩阵, $\mathbf{w}_q, \mathbf{w}_k \in \mathbf{R}^{d \times 1}$, $\mathbf{w}_u \in \mathbf{R}^{d \times 1}$ 是可训练的权重。

3.3.2 潜在关系预测

该组件如图 2 中的蓝色方框所示, R_{pot} 是潜在的关系。不同于之前的工作^[14-16]会对每个关系执行实体抽取,我们根据给定的一句话,首先预测出句子中可能存在的潜在关系,然后实体抽取只需要应用于潜在的关系。给定有 n 个 token 的句子的嵌入 $\mathbf{h} \in \mathbf{R}^{n \times d}$, 该组件中的每个元素均为:

$$\mathbf{h}_{\text{avg}} = \text{Avgpool}(\mathbf{h}) \in \mathbf{R}^{d \times 1} \quad (8)$$

$$\mathbf{P}_r = \sigma(\mathbf{W}_r \mathbf{h}_{\text{avg}} + \mathbf{b}_r) \quad (9)$$

Avgpool 是平均池化的操作^[24], $\mathbf{W}_r \in \mathbf{R}^{d \times 1}$ 是可训练的权重, σ 表示 sigmoid 函数。如图 2 所示,我们将其建模为一个多标签的二分类任务,如果概率超过某一阈值 θ_1 (默认设置为 0.5),则使用标记 1 分配相应的关系,否则使用标记 0。因此接下来只需要将基于特定关系的全局指针网络应用到预测的关系中,而不是所有的关系。

3.3.3 基于特定关系的全局指针网络

如图 2 中的绿色方框所示,我们得到了第 3.3.1 节中描述的几种基于潜在关系的句子表示。然后执行两个全局指针网络分别去预测主客体起始位置和主客体结束位置。我们分别抽取主客体起始位置对和主客体结束位置对的原因是为了解决实体重叠问题。此组件对每个 token 的操作如下:

$$\mathbf{P}_{i,j,k}^{\text{start}} = \text{GP}((\mathbf{h}_i + \mathbf{u}_k), (\mathbf{h}_j + \mathbf{u}_k)) \quad (10)$$

$$\mathbf{P}_{i,j,k}^{\text{end}} = \text{GP}((\mathbf{h}_i + \mathbf{u}_k), (\mathbf{h}_j + \mathbf{u}_k)) \quad (11)$$

其中, $\mathbf{u}_k \in \mathbf{R}^{d \times 1}$ 是可训练的嵌入矩阵 $\mathbf{U} \in \mathbf{R}^{d \times n_r^{\text{pot}}}$ 的第 k 个关系表示, n_r^{pot} 是句子中潜在关系子集的大小, $\mathbf{h}_i \in \mathbf{R}^{d \times 1}$ 是输入句子的第 i 个 token 的编码表示, $\mathbf{h}_j \in \mathbf{R}^{d \times 1}$ 是输入句子的第 j 个 token 的编码表示。

3.3.4 基于全局指针网络的通信

在基于特定关系的全局指针网络之后,我们获得了关于句子关系的所有可能的主客体起始位置和主客体结束位置,然后使用两个基于全局指针网络的通信矩阵确定正确的主客体对。需要注意的是,由于这两个基于全局指针网络的通信矩阵独立于关系,因此它可以与潜在关系预测同时学习。详细的过程如下:首先列举所有可能的主体起始结束位置对和客体起始位置对;然后分别检查每个主体起始结束位置对和客体起始结束位置对在全局矩阵中相应的分数,如果该值超过阈值 θ_2 ,则保留,否则将其过滤掉。

如图 2 中的橙色方框中的 Subject Start-End 类型的矩阵和 Object Start-End 类型的矩阵,给出一个有 n 个 token 的句子, Subject Start-End 类型和 Object Start-End 类型的全局

通信矩阵形状都是 $\mathbf{R}^{n \times n}$ 。Subject Start-End 类型的矩阵中的每个元素是主体起始结束位置对, Object Start-End 类型的矩阵中的每个元素是客体起始结束位置对,这两个通信矩阵的值越高,属于主体和客体的置信度就越高。举个主体的例子:“Smith”在第一行第一列,如果它是主体实体,它的值则会越高,客体与主体一致。Subject Start-End 类型和 Object Start-End 类型的全局通信矩阵中各元素的值如下:

$$\mathbf{P}_{i,j}^{\text{sub}} = \text{GP}(\mathbf{h}_i, \mathbf{h}_j) \quad (12)$$

$$\mathbf{P}_{i,j}^{\text{obj}} = \text{GP}(\mathbf{h}_i, \mathbf{h}_j) \quad (13)$$

其中, $\mathbf{h}_i \in \mathbf{R}^{d \times 1}$ 是输入句子的第 i 个 token 的编码表示, $\mathbf{h}_j \in \mathbf{R}^{d \times 1}$ 是输入句子的第 j 个 token 的编码表示。

3.4 训练策略

一般来说,我们把多标签分类任务当作多个二分类来处理,即使用 Sigmoid 函数得到每个类别标签的预测概率,然后通过阈值来选取模型对每个标签的预测结果,但是这样的做法常常面临数据不均衡的问题,而且最后的结果还会受阈值选取的影响。因此,我们在训练联合抽取模型时,受 Sun 等提出的 CircleLoss^[25]的启发,将“Softmax+交叉熵”推广到多标签分类任务。与多标签分类变成多个二分类法不同的是,本文方法将目标类别得分与非目标类别得分进行比较,且借助于 logsumexp 的良好性质,自动确定阈值,从而解决了数据不均衡的问题。损失函数可以分为 3 个部分,具体如下:

$$\ell_{\text{rel}} = \frac{-1}{n_r} \sum_{i=1}^{n_r} (y_i \log P_{\text{rel}} + (1 - y_i) \log(1 - P_{\text{rel}})) \quad (14)$$

$$\ell_{\text{relGlobal}} = \frac{-1}{2 \times n \times n \times n_r^{\text{pot}}} \sum_{t \in \{\text{start}, \text{end}\}} \sum_{k=1}^{n_r^{\text{pot}}} \sum_{j=1}^n \sum_{i=1}^n (\log(1 + \sum_{(i,j) \in P_k^t} e^{-P_{i,j,t}^k}) + \log(1 + \sum_{(i,j) \in N_k^t} e^{P_{i,j,t}^k})) \quad (15)$$

$$\ell_{\text{corGlobal}} = \frac{-1}{2 \times n \times n} \sum_{t \in \{\text{sub}, \text{obj}\}} \sum_{j=1}^n \sum_{i=1}^n ((\log(1 + \sum_{(i,j) \in P_t} e^{-P_{i,j,t}}) + \log(1 + \sum_{(i,j) \in N_t} e^{P_{i,j,t}})) \quad (16)$$

其中, n_r 是完整的关系集合的大小, n_r^{pot} 是句子中潜在关系子集的大小, P_k^t 是该样本的关系为 k 的主客体的 t 位置对集合, N_k^t 是该样本的关系为 k 的所有非实体和关系不为 k 的主客体 t 位置对集合, P_t 是该样本的类型是 t 实体的起始结束位置对, N_t 是该样本的所有非实体或类型非 t 的起始结束位置对,损失函数就是这 3 个部分的总和,具体如下:

$$\ell_{\text{total}} = \ell_{\text{rel}} + \ell_{\text{relGlobal}} + \ell_{\text{corGlobal}} \quad (17)$$

4 实验与分析

4.1 数据集

为了进行公平、全面的比较,我们根据 HBT^[26]和 TPlinker^[16]采用的两个公开数据集 NYT^[27]和 WebNLG^[28],来评估本文模型。这两个数据集有两个版本:1)标注实体的最后一个词;2)标注整个实体的跨度。NYT* 和 WebNLG* 标注实体的最后一个词, NYT 和 WebNLG 标注整个实体的跨度。为了进一步验证本文方法在重叠关系和多重关系中的抽取能力,我们根据 OneRel^[19],通过重叠的模式和三元组数量来分割统计测试集,具体数据集的统计结果如表 1 所列。

表 1 实验数据集的统计

Table 1 Statistics of experimental datasets

Dataset	Train	Valid	Test	Overlapping Pattern			Number of triplets		Relations
				Normal	SEO	EPO	N= 1	N> 1	
NYT*	56 195	4 999	5 000	3 266	1 297	978	3 244	1 756	24
WebNLG*	5 019	500	703	245	457	26	266	437	171
NYT	56 196	5 000	5 000	3 071	1 273	1 168	3 089	1 911	24
WebNLG	5 019	500	703	239	448	6	256	447	216

根据上文的描述,只有抽取的关系三元组与真实标注的关系三元组精确匹配时,才认为其是正确的,这就意味着实体的最后一个词或者整个主客体实体的跨度、关系必须是正确的。与此同时,我们展示了所有基线模型的精确率、召回率和 F1-score。

我们将本文方法与 11 个基线模型进行了比较,其中 PRGC^[18] 和 OneRel^[19] 是最先进的模型。除了具体说明外,所有基线模型的实验结果都直接取自 PRGC^[18] 和 OneRel^[19]。

4.2 实验细节

我们用 Keras 实现了模型,模型使用了 Adam^[29] 优化器,并且在 NYT 和 WebNLG 数据集上的批处理大小分别为 64 和 8。编码端和解码端的学习率分别为 0.00005 和 0.002,以便快速收敛。我们也设置了权重衰减^[30],其值为 0.005。

为了进行公平的比较,我们使用了英文版最基本的

BERT 预训练语言模型作为编码端,并将输入句子的最大长度设置为 128,这与之前的工作相同。我们的实验使用 1 块 NVIDIA Tesla P40 GPU,训练了 100 个 epoch,最后选择了 F1 指标最高的模型。

4.3 实验结果

4.3.1 总体结果

表 2 列出了本文模型与其他基线模型在所有数据集上的结果。我们可以观察到,本文模型在几乎所有的评估指标方面都优于对比模型,即使与最近的 OneRel^[19] 基线模型相比,本文方法也更具优势。同时为了验证我们的解码端的有效性,将编码端 BERT 的所有参数随机初始化。BGNRE_{random} 的性能表明,我们的解码端(比 PRGC_{random} 提升了 0.2%,比 CasRel_{random} 提升了 7%)即使没有利用预训练好的 BERT 语言模型,也仍然比其他方法更加具有竞争力。

表 2 BGNRE 与基线模型对比结果

Table 2 Comparison results of BGNRE and Baseline model

Model	NYT*			WebNLG*			NYT			WebNLG		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
NovelTagging ^[11]	—	—	—	—	—	—	32.8	30.6	31.7	52.5	19.3	28.3
CopyRE ^[31]	61.0	56.6	58.7	37.7	36.4	37.1	—	—	—	—	—	—
MultiHead ^[12]	—	—	—	—	—	—	60.7	58.6	59.6	57.5	54.1	55.7
GraphRel ^[32]	63.9	60.0	61.9	44.7	41.1	42.9	—	—	—	—	—	—
OrderCopyRE ^[33]	77.9	67.2	72.1	63.3	59.9	61.6	—	—	—	—	—	—
ETL-span ^[26]	84.9	72.3	78.1	84.0	91.5	87.6	85.5	71.7	78.0	84.3	82.0	83.1
WDec ^[13]	94.5	76.2	84.4	—	—	—	—	—	—	—	—	—
RSAN ^[15]	—	—	—	—	—	—	85.7	83.6	84.6	80.5	83.8	82.1
CasRel _{Random} ^[14]	81.5	75.7	78.5	84.7	79.5	82.0	—	—	—	—	—	—
CasRelBERT ^[14]	89.7	89.5	89.6	93.4	90.1	91.8	—	—	—	—	—	—
TPLinkerBERT ^[16]	91.3	92.5	91.9	91.8	92.0	91.9	91.4	92.6	92.0	88.9	84.5	86.7
SPNBERT ^[17]	93.3	91.7	92.5	93.1	93.6	93.4	92.5	92.2	92.3	—	—	—
PRGC _{Random} ^[18]	89.6	82.3	85.8	90.6	88.5	89.5	87.8	83.8	85.8	82.5	79.2	80.8
PRGCBERT ^[18]	93.3	91.9	92.6	94.0	92.1	93.0	93.5	91.9	92.7	89.9	87.2	88.5
OneRelBERT ^[19]	92.8	92.9	92.8	94.1	94.4	94.3	93.2	92.6	92.9	91.8	90.3	91.0
BGNRE _{Random}	89.7	82.6	86.0	90.6	88.9	89.7	87.6	84.2	85.9	83.1	78.8	80.9
BGNREBERT	93.1	92.9	92.9	94.8	93.5	94.1	93.3	92.6	92.9	91.6	90.7	91.1

(单位: %)

值得注意的是,本文模型在 NYT* 数据集上相比 OneRel_{BERT} 取得了 0.1% 的提升。其原因是,本文模型提出了新的全局指针网络,大大增加了重叠实体的识别,特别是优于 OneRel 中采用的全局指针网络方法。换句话说,与在每个关系下执行实体抽取的模型相比,使用本文的全局指针网络可以带来额外的提升。

4.3.2 复杂场景中的详细结果

根据之前的工作^[14-16],为了验证本文模型在面对不同重叠方式和拥有不同数量三元组的句子时的有效性,我们在 NYT* 和 WebNLG* 数据集上做了进一步的实验。

如表 3 所列,本文模型在同一个数据集的所有重叠方式上,超越了所有的基线模型。如表 4 所列,在句子中的三元组的数量不同的情况下,本文模型的结果在每个子集几乎上都

优于其他模型。总之,这两个进一步的实验充分证明了本文模型在复杂场景中的优势。

表 3 具有不同重叠模式的句子的 F1 指标对比

Table 3 Comparison of F1 indicators of sentences with different overlapping patterns

(单位: %)

Model	NYT*			WebNLG*		
	Normal	SEO	EPO	Normal	SEO	EPO
OrderCopyRE ^[33]	71.2	69.4	72.8	65.4	60.1	67.4
ETL-Span ^[26]	88.5	87.6	60.3	87.3	91.5	80.5
CasRel ^[14]	87.3	91.4	92.0	89.4	92.2	94.7
TPLinker ^[16]	90.1	93.4	94.0	87.9	92.5	95.3
SPN ^[17]	90.8	94.1	94.0	—	—	—
PRGC ^[18]	91.0	94.0	94.5	90.4	93.6	95.9
OneRel ^[19]	90.6	95.1	94.8	91.9	95.4	94.7
BGNRE	91.2	95.1	94.9	91.6	95.5	96.1

表 4 句子中不同三元组数量的 F1 指标对比
Table 4 Comparison of F1 indicators of different triples in sentences

(单位: %)

Model	NYT *					WebNLG *				
	N=1	N=2	N=3	N=4	N≥5	N=1	N=2	N=3	N=4	N≥5
OrderCopyRE ^[33]	71.7	72.6	72.5	77.9	45.9	63.4	62.2	64.4	57.2	55.7
ETL-Span ^[26]	88.5	82.1	74.7	75.6	76.9	82.1	86.5	91.4	89.5	91.1
CasRel ^[14]	88.2	90.3	91.9	94.2	83.7	89.3	90.8	94.2	92.4	90.9
TPLinker ^[16]	90.0	92.8	93.1	96.1	90.0	88.0	90.1	94.6	93.3	91.6
SPN ^[17]	90.9	93.4	94.2	95.5	90.6	89.5	91.3	96.4	94.7	93.8
PRGC ^[18]	91.1	93.0	93.5	95.5	93.0	89.9	91.6	95.0	94.8	92.2
OneRel ^[19]	90.5	93.4	93.9	96.5	94.2	91.4	93.0	95.9	95.7	94.5
BGPNRE	91.3	93.4	93.8	96.6	94.3	91.2	93.1	95.2	95.8	94.6

4.4 消融实验

为了证明本文方法各组件的有效性,我们进行了消融实验,具体结果如表 5 所列。

表 5 BGPNRE 的消融实验指标对比
Table 5 Comparison of ablation experimental indexes of BGPNRE
(单位: %)

Model		P	R	F1
NYT *	BGPNRE	93.1	92.9	92.9
	-Potential Relation Prediction	90.6	91.3	90.9
	-Rel-Spec Global Pointer	68.2	90.8	77.9
	-Global Pointer Correspondence	70.3	90.6	79.2
WebNLG *	BGPNRE	94.8	93.5	94.1
	-Potential Relation Prediction	87.3	88.6	87.9
	-Rel-Spec Global Pointer	40.3	90.5	55.8
	-Global Pointer Correspondence	65.3	91.2	76.1

4.4.1 潜在关系预测的影响

在去除潜在关系预测组件时,使用关系集中的每个关系来执行全局指针网络,以避免暴露偏差。如表 5 所列,在没有该组件的情况下,精度显著降低,因为句子中没有出现的关系,导致预测三元组的数量增加,特别是在 WebNLG 数据集中,关系数据集较大,会带来巨大的冗余关系。同时,随着句子中关系数量的增加,训练和推理时间增加了 3~4 倍。实验结果证明了潜在关系预测组件的有效性,这不仅有利于提高模型的准确性,而且有利于提高效率。

4.4.2 基于特定关系的全局指针网络的影响

在去除基于特定关系的全局指针网络组件时,根据 OneRel,采用基于传统的全局指针网络的方案,执行一个多标签分类来识别主客体实体的起始结束位置,如表 5 所列,基于传统的全局指针网络的方案的结果显著下降。基于传统的全局指针网络的方案会出现明显的越界问题,虽然识别了正确的关系,但忽略了它们对主客体实体起始结束位置信息的敏感程度。这是因为模型倾向于学习一个实体的位置,而不是理解底层的起始结束位置语义。然而,我们所使用的全局指针网络方案在这两种情况下都表现良好,证明了我们的全局指针网络方案更具有优势。

4.4.3 基于全局指针网络的通信的影响

为了进行比较,我们使用 Zheng 等^[11]和 Yuan 等^[15]提出的启发式最近邻原理来组合主客体对方法进行验证。如表 5 所列,在没有全局对应的情况下,精度也显著下降,这是因为当模型失去了该组件施加的约束时,预测的数量会随着许多不匹配的对而增加。本实验证明了基于全局指针网络的通信

在主客体对齐任务中的有效性,并且优于启发式最近邻原理。

结束语 本文提出了一种基于 BERT 的全局指针网络实体关系联合抽取方法,它能够发现共享多个实体的重叠关系,同时不受实体嵌套的影响,而且大大减轻了冗余关系的判断,解决了 Span 提取泛化能力不足等问题。实验结果表明,该模型在两个公共数据集上达到了先进的水平,并成功地处理了许多复杂的场景。虽然该模型取得了先进的效果,但是在解决暴露偏差问题方面仍然略显不足,后续仍然会探索新的方法来解决该问题。

参 考 文 献

[1] SANG E,MEULDER F. Language-independent named entity recognition[C]//Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003. 2003;142-147.

[2] RATINOV L,ROTH D. Design challenges and misconceptions in named entity recognition[C]//Proceedings of the Thirteenth Conference on Computational Natural Language Learning. 2009;147-155.

[3] ZELNKO D,AONE C,RICHARDELLA A. Kernel methods for relation extraction[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing-Volume 10. 2002;77-78.

[4] RAZVAN C,RAYMOND J. A shortest path dependency kernel for relation extraction[C]//Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing. 2005;724-731.

[5] PAWAR S,GIRISH K,BHATTACHARYYA P. Relation extraction: A survey[J/OL]. <https://doi.org/10.48550/arXiv.1712.05191>.

[6] WANG Z,WEN R,CHEN X,et al. Finding in influential instances for distantly supervised relation extraction[J/OL]. <https://doi.org/10.48550/arXiv.1712.05191>.

[7] YU X,LAM W. Jointly identifying entities and extracting relations in encyclopedia text via agraphical model approach[C]//The 28th International Conference on Computational Linguistics. 2010;1399-1407.

[8] LI Q,JI H. Incremental joint extraction of entity mentions and relations[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics(Volume 1:Long Papers). 2014;402-412.

[9] MIWA M,SASAKI Y. Modeling joint entity and relation extrac-

tion with tablerepresentation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014;1858-1869.

[10] REN X, WU Z, HE W, et al. Cotype: Joint extraction of typed entities and relations with knowledge bases[C]//Proceedings of the 26th International Conference on World Wide Web. 2017: 1015-1024.

[11] ZHENG S, WANG F, BAO H, et al. Joint extraction of entities and relations based on a noveltatagging scheme[C] // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics(Volume 1: Long Papers). 2017;1227-1236.

[12] BEKOULIS G, DEIEU J, DEMEESTER T, et al. Joint entity recognition and relation extraction as a multi-head selection problem[J]. Expert System with Applications, 2018, 114: 34-45.

[13] NAYAK T, TOU H. Effective modeling of encoder-decoder architecture for joint entity and relation extraction[C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2020: 8528-8535.

[14] 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. 2020;1476-1488.

[15] YUAN Y, ZHOU X, PAN S, et al. A relation-specific attention network for joint entity and relation extraction [C] // International Joint Conference on Artificial Intelligence. 2020;4054-4060.

[16] WANG Y, YU B, ZHANG Y, et al. TPLinker: Single-stage joint extraction of entities and relations through token pair linking [C]//Proceedings of the 28th International Conference on Computational Linguistics. 2020;1572-1582.

[17] SUI D, CHEN Y, LIU K, et al. Joint Entity and Relation Extraction with Set Prediction Networks [J/OL]. <https://doi.org/10.48550/arXiv.2011.01675>.

[18] ZENG X, HE S, ZENG D, et al. PRGC: Potential Relation and Global Correspondence Based Joint Relational Triple Extraction [C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics. 2021;6225-6235.

[19] SHANG Y, HUANG H, MAO X. OneRel: Joint Entity and Relation Extraction with One Module in One Step[C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2022: 11285-11293.

[20] VINYALS O, FORTUNATO M, JAITLEY N. Pointer networks [C] // Advances in Neural Information Processing Systems. 2015;2674-2682.

[21] DEVLIN J, CHANG M, LEE K. Bert: Pre-training of deep bidirectional transformers for language understanding[C] // Annual Conference of the North American Chapter of the Association for Computational Linguistics. 2019;4171-4186.

[22] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C] // Advances in Neural Information Processing Systems. 2017;6000-6010.

[23] SU J, LU Y, PAN S, et al. RoFormer: Enhanced Transformer with Rotary Position Embedding[J/OL]. <https://doi.org/10.48550/arXiv.2104.09864>.

[24] LIN M, CHEN Q, YAN S. Network in network[C]//2nd International Conference on Learning Representations. 2014;14-16.

[25] SUN Y, CHENG C, ZHANG Y, et al. Circle Loss: A Unified Perspective of Pair Similarity Optimization[J/OL]. <https://doi.org/10.48550/arXiv.2002.10857>.

[26] YU B, ZHANG Z, SU J, et al. Joint extraction of entities and relations based on a novel decomposition strategy[C]//24th European Conference on Artificial Intelligence—ECAI 2020. 2019.

[27] RIEDEL S, YAO L, MCCALLUM A. Modeling relations and their mentions without labeled text [C]// Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases. 2010.

[28] GARDENT C, SHIMORINA A, NARAYAN S, et al. Creating training corpora for NLG microplanners[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics(Volume 1: Long Papers). 2017;179-188.

[29] DIEDERIK P, BA J. Adam: A method for stochastic optimization[C]//3rd International Conference on Learning Representations. 2015.

[30] LOSHCHILOV I, HUTTER F. Fixing weight decay regularization in Adam [J/OL]. <https://doi.org/10.48550/arXiv.1711.05101>.

[31] ZENG X, ZENG D, HE S, et al. Extracting relational facts by an end-to-end neural model with Copy Mechanism[C] // Proceedings of the 56th Annual Meeting of the Association for Computational(Volume1: Long Papers). 2018;506-514.

[32] FU T, LI P, MA W. GraphRel: Modeling text as Relational graphs for joint entity and relation extraction[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019;1409-1418.

[33] ZENG X, HE S, ZENG D, et al. Learning the extraction order of multiple relational facts in a sentence with reinforcement learning[C]//Proceedings of the 56th Annual Meeting of the Association for Computational(Volume1: Long Papers). 2018;506-514.



DENG Liang, born in 1980, postgraduate. His main research interests include deep learning and knowledge graph.



QI Panhu, born in 1990, postgraduate. His main research interests include deep learning and natural language Processing.