

基于共享特征的联合实体关系抽取网络

郝增昊^{1,2,3}, 赵晶^{1,2,3}

(1. 齐鲁工业大学(山东省科学院) 山东省计算中心(国家超级计算济南中心) 算力互联网与信息安全教育部重点实验室; 2. 齐鲁工业大学(山东省科学院) 计算机科学与技术学部 大数据应用技术山东省工程研究中心; 3. 山东省工业网络和信息网络安全重点实验室 山东省基础科学研究中心(计算机科学), 山东 济南 250353)

摘要: 在信息抽取流程中, 联合实体关系抽取占据着关键地位, 该任务整合了命名实体识别与关系抽取两项核心子任务。目前的联合抽取方法仍存在特征误分类和子任务交互不足的问题。为此, 提出一种基于共享特征的联合实体关系抽取网络SJERE。该网络采用并行编码的方法来获取命名实体特征表示和关系特征表示, 有效避免了特征重叠和特征误分类问题。为实现两个子任务之间信息的双向交互, SJERE网络构造了共享特征模块。该模块允许模型通过实体信息辅助进行关系预测, 从而使实体识别和关系抽取两个子任务相互促进、相互增强。在NYT、WebNLG和SciERC 3个基准数据集上与现有联合模型进行比较, SJERE网络展现出一定的性能优势。

关键词: 信息抽取; 联合实体关系抽取; 并行编码; 共享特征; 双向交互

DOI: 10.11907/tjdc.241626

开放科学(资源服务)标识码(OSID):



中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(2025)009-0041-07

Shared Feature-Based Joint Entity and Relation Extraction Network

HAO Zenghao^{1,2,3}, ZHAO Jing^{1,2,3}

(1. Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Ji'nan), Qilu University of Technology (Shandong Academy of Sciences); 2. Shandong Engineering Research Center of Big Data Applied Technology, Faculty of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences); 3. Shandong Provincial Key Laboratory of Industrial Network and Information System Security, Shandong Fundamental Research Center for Computer Science, Ji'nan 250353, China)

Abstract: In the information extraction process, joint entity relationship extraction plays a key role, which integrates two core sub tasks: named entity recognition and relationship extraction. The current joint extraction methods still suffer from feature misclassification and insufficient interaction between subtasks. Therefore, a joint entity relationship extraction network SJERE based on shared features is proposed. This network adopts a parallel encoding method to obtain named entity feature representations and relationship feature representations, effectively avoiding feature overlap and feature misclassification problems. To achieve bidirectional information exchange between two subtasks, the SJERE network constructed a shared feature module. This module allows the model to assist in relationship prediction through entity information, thereby promoting and enhancing the two sub tasks of entity recognition and relationship extraction. Compared with existing joint models on three benchmark datasets, NYT, WebNLG, and SciERC, the SJERE network exhibits certain performance advantages.

Key Words: information extraction; joint entity and relation extraction; parallel encoding; shared feature; bidirectional interaction

收稿日期: 2024-07-22

扫描二维码阅读全文:



基金项目: 山东省自然科学基金—联合基金项目(ZR2022LZH008); 齐鲁工业大学(山东省科学院)研究生教改项目(YJG23YB002); 山东省本科教学改革研究项目重点项目(Z2023314); 济南市市校融合发展战略工程项目(JNSX2023045)

作者简介: 郝增昊(2000-), 男, 齐鲁工业大学(山东省科学院)计算机科学与技术学部硕士研究生, 研究方向为实体关系抽取; 赵晶(1979-), 女, 博士, 齐鲁工业大学(山东省科学院)计算机科学与技术学部副教授、硕士生导师, 研究方向为智能优化算法、机器学习、图像处理。本文通讯作者: 赵晶。

0 引言

实体关系抽取是自然语言处理领域中的关键任务,其目的是通过实体识别从非结构化或半结构化的文本中识别出头实体和尾实体,然后通过关系抽取对实体之间的关系进行预测,并将其表示为实体—关系三元组的形式,即(<实体,关系,实体>)。实体关系抽取可将自然语言文本转换为结构化数据,使计算机能够更好地理解文本中所描述的实体及其关系。该任务对于知识图谱构建、机器阅读、文本摘要和问答系统等下游自然语言处理任务具有重要的意义^[1-3]。

1 相关研究

目前,关于实体关系抽取技术的研究已取得一定进展,相关方法大致可分为以下几个类别。

1.1 基于标注的方法

该类方法通常采用二元标注序列来确定实体的起始位置或建立实体之间的关系。例如,文献[4]提出的BiRTE通过双向框架(s2o和o2s)提取实体对,利用共享编码器促进特征流动;同时采用二进制标记预测实体边界概率,并结合双线性模型分类实体关系。文献[5]提出的CasRel模型采用级联的二元标注框架,先提取头实体,再提取关系和尾实体,从而实现关系三元组的抽取。文献[6]提出的SGNet模型通过软剪枝构建全连接图,结合图卷积网络(Graph Convolutional Network, GCN)整合节点信息,利用GlobalPointer解码器将三元组转为五元组(主体头尾、关系、宾语头尾)进行联合评分与解码,从而进行实体关系抽取。

1.2 基于跨度的方法

该类方法通过识别不同的跨度来确定实体及其类型,然后对跨度对进行关系分类操作,以获得结构化的关系三元组。例如,文献[7]将命名实体识别视为一个跨度分类问题,通过构造动态跨度图生成跨度表示。文献[8]提出的SPE模型将跨度级信息集成到预训练模型中,用于实体关系提取。文献[9]提出的SpERT模型在BERT嵌入的基础上进行轻量级推理,同时结合实体识别、局部上下文表示和强负样本训练,实现了跨度级的联合实体及其关系抽取。

1.3 基于生成的方法

该类方法使用生成模型直接从文本中生成实体之间的关系标签,而不是简单地从预定义的关系中选择。例如,文献[10]提出基于Seq2Seq和拷贝机制的端到端模型,直接生成关系三元组。文献[11]提出基于Seq2Seq和强化学习的实体关系抽取模型,编码器将句子编码,解码器逐步生成关系或拷贝实体,直至生成停止标记。其通过

REINFORCE优化抽取顺序,奖励正确生成的三元组,从而更好地实现关系三元组的抽取操作。文献[12]将关系抽取任务视为序列到序列的问题,使用预训练模型BART(Bidirectional and Auto-Regressive Transformers)为原始句子生成关系三元组^[13]。文献[14]提出的SPN模型使用一种基于集合预测的方法来进行实体关系抽取,利用transformer的非自回归解码器输出三元组。文献[15]提出的EmRel是一种基于生成的三元组抽取模型,通过构建统一表示空间并融合实体、关系与上下文信息,利用Tucker分解生成候选三元组并判断其有效性,实现端到端抽取,从而实现关系三元组的抽取。

1.4 基于表填充的方法

该类方法通常为每个关系维护一个表,其中的每个条目表示实体对是否具有这种类型的关系。例如,文献[16]提出的TpLinker模型将联合抽取任务转化为表填充的单阶段模型,通过“握手标记方案”构建3个标记矩阵,标注实体边界及其关系,并展平成序列用于训练与解码。文献[17]提出的UNIRE模型通过统一表填充过程中实体关系抽取任务的标签空间实现关系三元组的抽取。文献[18]提出的PFN模型设计了分区过滤网络,通过对不同任务分区进行划分整合来进行实体关系提取。文献[19]提出的DiReclRel模型将三元组抽取转化为特定关系的表填充任务,通过连接矩阵在“头至尾”图中直接判断实体对关系,实现一步抽取,提升了鲁棒性与效率。

在早期阶段,实体关系抽取任务被分解为两个独立的任务^[20]。首先是实体识别任务,其从大规模非结构化文本中识别实体的边界与类型;其次是关系抽取任务,其将实体识别的结果作为输入,通过从输入的实体信息中提取上下文特征来预测实体之间是否存在某种关系。实体识别阶段的错误会传输到关系抽取任务中,导致误差传播和偏差暴露等现象,从而影响关系抽取性能。此外,由于这两个独立任务之间缺乏交互,其中的文本信息得不到充分利用。

为解决命名实体识别与关系抽取任务之间缺乏交互的问题,许多研究者对联合实体关系抽取展开深入研究^[5,9,17-18,21]。联合实体关系抽取模型旨在使用单一模型抽取实体及其关系,通过整合实体识别与关系抽取任务之间的潜在关联特征来促进两个子任务之间关键信息的交换,从而解决流水线方法中存在的问题。虽然目前的联合建模方法可以通过学习实体识别与关系抽取任务之间的依赖来捕获更复杂的实体及其关系信息,但是其仍存在一些局限性。例如,文献[9]和[17]提出的方法由于需要获得两个子任务之间的共享特征表示,会出现特征误分类的情况,即当获取实体特征表示时会掺杂一些关系特征表示,反之,在获取关系特征表示时,也会掺杂一些实体特征。此外,目前大多数联合模型只利用实体信息来辅助关系抽取任务,忽略了关系信息也可以辅助实体识别的事实

或只利用少量局部关系信息,这会导致两个子任务之间出现交互不足的现象^[5,9,17]。

为解决子任务交互不足和特征误分类的问题,本文提出基于共享特征的联合实体关系抽取网络(Shared Feature-Based Joint Entity and Relation Extraction Network, SJERE)。在实体关系提取过程中,首先,使用BERT获得句子的特征表示,使用两个独立的多层感知器从句子的特征表示中提取实体和关系的特征表示。由于这两个多层感知器是独立运行的,模型的特征误分类问题得到有效改善;其次,将位置矩阵嵌入到实体关系的特定表示中,以捕获空间语义信息,利用一维卷积获得实体关系特征表示之间的表面和显式语义关系;再次,使用二维卷积捕获实体关系特征表示之间的深层和潜在语义关系;最后,对一维卷积和二维卷积的结果进行加权与融合,得到共享特征矩阵,从而实现实体与关系的双向交互。在3个基准数据集上进行了大量实验,结果表明,与基线模型相比,所提模型展现出一定的性能优势,在SciERC数据集上的消融实验亦证实了模型各模块的有效性。

2 SJERE 模型构建

2.1 问题定义

SJERE模型在统一框架下将联合抽取任务划分为命名实体识别单元和关系抽取单元,然后通过表填充方法对实体和关系进行分类建模。首先,给定一个包含 K 个token的句子 $Z = \{z_1, z_2, z_3, \dots, z_k\}$;其次,创建一组预定义的实体类型 E 以及一组预定义的关系类型 R 。使用 $W = \{w_i\}_{i=1}^N$ 来表示句子 Z 中所有可能的实体,为句子 Z 中的每个实体 $w_i \in W$ 预测一个实体类型 $e \in E$ 。对于每一个预测的实体对 $w_i, w_j \in W$,还需要预测实体对之间的关系 $r \in R$ 。联合

实体识别和关系提取任务的输出为一个三元组集合 $X = \{((w_i, e_i), (w_j, e_j), r) | w_i, w_j \in W, e_i, e_j \in E, r \in R\}$ 。

2.2 模型结构

SJERE模型由编码器模块、共享特征模块和分类模块3个部分组成。结构如图1所示。

2.2.1 编码器模块

将句子 $Z = \{z_1, z_2, z_3, \dots, z_k\}$ 输入到编码器模块中。该模块使用BERT作为编码器对输入的句子 Z 进行编码,得到句子的特征表示^[22]。公式为:

$$C = \text{BERT}(z_1, z_2, z_3, \dots, z_k) \quad (1)$$

式中: $C = \{c_1, c_2, \dots, c_k | c_i \in \mathbb{R}^d\}$ 为通过BERT编码处理得到的特征表示; k 为句子 Z 编码后的特征向量个数; d 为特征向量对应的维度。

2.2.2 共享特征模块

通常情况下,命名实体识别和关系抽取两个任务需要不同的上下文表示,前者更强调词汇级特征,后者更重视实体之间的上下文表示。如果直接对两个子任务使用相同的特征表示,可能会导致其无法抓取各自的有效特征,导致特征误分类。SJERE模型的共享特征模块可有效避免该问题的发生。其通过并行编码的方式将特征输入到共享特征模块中,即使用两个不同的多层感知器分别获得实体识别和关系抽取任务的特定特征表示。表示为:

$$C_{NER} = \text{MLP}_{NER}(C) \quad (2)$$

$$C_{RE} = \text{MLP}_{RE}(C) \quad (3)$$

式中: $C_{NER} = \{c_1^e, c_2^e, \dots, c_k^e | c_i^e \in \mathbb{R}^d\}$ 为命名实体的特征表示,即命名实体初始特征; $C_{RE} = \{c_1^r, c_2^r, \dots, c_k^r | c_i^r \in \mathbb{R}^d\}$ 为关系的特征表示,即关系初始特征。

为捕获空间中词向量之间的语义关系,嵌入一个位置矩阵 D 来获得词向量之间的相对距离,将两个子任务的初始特征与相对位置矩阵 D 结合起来作为卷积层的输入。

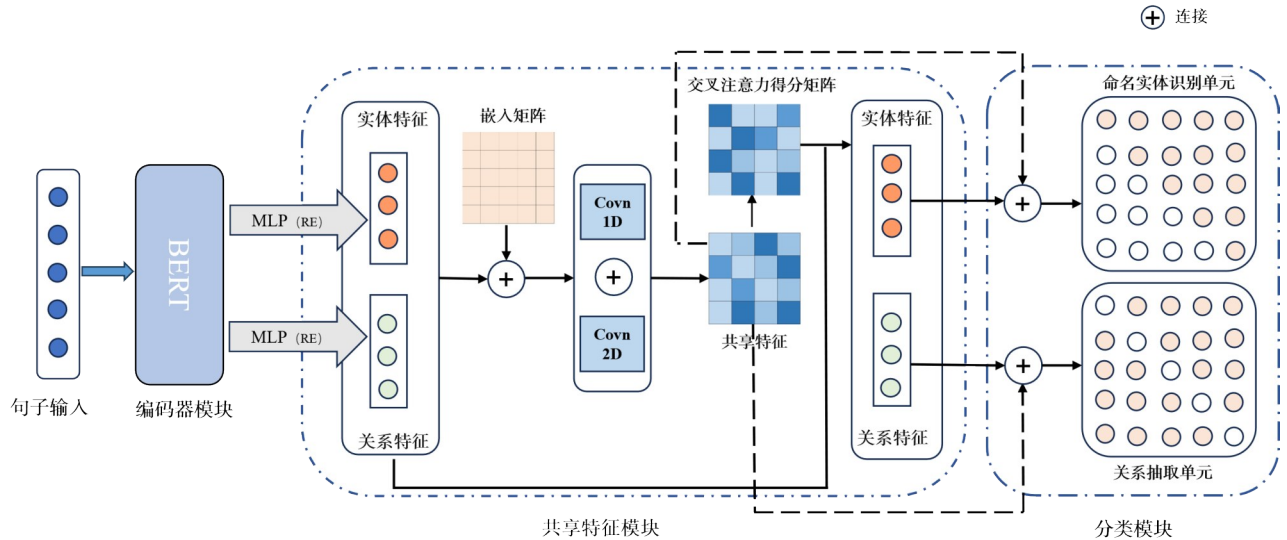


Fig. 1 SJERE model structure

图1 SJERE模型结构

受 Yan 等^[18]研究的启发,本文引入一种特征共享机制,通过注意力权重将有价值的信息传递给每个子任务。在共享特征模块中,利用一维卷积来捕获实体和关系的表面语义特征和显式特征,利用二维卷积来捕获实体和关系的深层上下文关系和潜在特征,对二者进行加权融合,获得两个子任务的共享特征矩阵。表示为:

$$[C_{NER}; C_{RE}; D] \quad (4)$$

$$C_{share} = \alpha \cdot \text{Conv1d}(G) + \beta \cdot \text{Conv1d}(G) \quad (5)$$

式中: G 表示命名实体特征、关系特征与位置矩阵相互拼接的特征矩阵; α 和 β 为可以通过训练调整的权重系数; C_{share} 表示卷积后得到的共享特征矩阵。使用前馈神经网络和 softmax 激活函数获得 C_{share} 的交叉注意力得分矩阵。表示为:

$$CA = FN(C_{share}) \quad (6)$$

$$\gamma = \text{softmax}(CA) \quad (7)$$

$$\delta = \text{softmax}(CA^T) \quad (8)$$

式中: γ 表示实体到关系的注意力权重,即关系特征对实体的重要性; δ 表示关系到实体的注意力权重,即实体特征对关系的重要性。

使用共享特性机制实现两个子任务之间的交互,并将交互结果整合到初始特征中。表示为:

$$a_i^e = c_i^e + \sum_{j=1}^k \gamma_{ij} c_j^r \quad (9)$$

$$a_i^r = c_i^r + \sum_{j=1}^k \delta_{ij} c_j^e \quad (10)$$

式中: c_i^e 为位置 i 处命名实体的特征表示; c_j^r 为位置 j 处的关系特征表示; γ_{ij} 为命名实体 i 处特征表示与关系 j 处特征表示的相关性; a_i^e 为交互后的命名实体特征,即 i 处的实体特征与各位置的关系特征进行加权融合得到更加丰富的特征表示,其他同理。 $c_i^e, c_j^e \in C_{NER}$, $c_i^r, c_j^r \in C_{RE}$ 。

2.2.3 分类模块

分类模块包括命名实体识别单元和关系抽取单元两个部分。该模块将这两个子任务转换为表填充问题来解决,采用二维表来表示实体标签和关系标签。在命名实体识别任务中,模型可以识别句子中的实体跨度。在关系抽取任务中,模型只需要标识实体的头词,而不需要对整个实体进行分类。将命名实体和关系的特征表示与共享特征相结合,作为命名实体识别和关系抽取任务的输入。表示为:

$$v_{ij}^e = [a_i^e; a_j^e; c_{ij}^{share}] \quad (11)$$

$$v_{ij}^r = [a_i^r; a_j^r; c_{ij}^{share}] \quad (12)$$

式中: v_{ij}^e 为位置 i 与位置 j 之间的命名实体特征表示; v_{ij}^r 为位置 i 与位置 j 之间的关系特征表示; c_{ij}^{share} 为位置 i 与位置 j 之间的共享特征表示, c_{ij}^{share} 属于 C_{share} 。

对于词对 (z_i, z_j) ,采用联合建模的方式同时预测该实体对所属的实体类别标签 e 以及其关系类型标签 r 的概率 $P(e|z_i, z_j)$ 和 $P(r|z_i, z_j)$ 。计算公式为:

$$P(e|z_i, z_j) = \text{sigmoid}(M_2 \sigma(M_1 v_{ij}^e + b_1) + b_2) \quad (13)$$

$$P(r|z_i, z_j) = \text{sigmoid}(M_4 \sigma(M_3 v_{ij}^r + b_3) + b_4) \quad (14)$$

式中: M 和 b 为可通过学习不断调整的参数; $\sigma(\cdot)$ 为激活函数。

将命名实体识别和关系抽取任务视为多标签分类任务,使用二元交叉熵作为该任务的损失函数。计算公式为:

$$\mathcal{L}_{NER} = -\sum_{i \leq j} \sum_{e \in E} \mathbb{I}(e = 1) \log P(e|z_i, z_j) + \mathbb{I}(e = 0) \log(1 - P(e|z_i, z_j)) \quad (15)$$

$$\mathcal{L}_{RE} = -\sum_{i \neq j} \sum_{r \in R} \mathbb{I}(r = 1) \log P(r|z_i, z_j) + \mathbb{I}(r = 0) \log(1 - P(r|z_i, z_j)) \quad (16)$$

式中: $\mathbb{I}(\cdot)$ 为指示函数。

本文模型的损失函数 $\mathcal{L}$ 由命名实体识别任务损失函数 $\mathcal{L}_{NER}$ 和关系抽取任务损失函数 $\mathcal{L}_{RE}$ 组成,可以通过调整损失函数对整个模型进行微调以提高性能。表示为:

$$L = \mathcal{L}_{NER} + \mathcal{L}_{RE} \quad (17)$$

3 实验方法与结果分析

3.1 实验环境与参数设置

实验在显卡为 NVIDIA GeForce RTX 3090 GPU 的服務器上运行,使用的开发语言为 Python,深度学习框架为 Pytorch。初始词向量维度设置为 786,样本序列最大长度设置为 128。SciERC 数据集和其他数据集的训练批次大小分别设置为 4 和 8,初始学习率分别设置为 $1e-4$ 和 $2e-4$ 。dropout 设置为 0.1,用于缓解过拟合,训练过程中采用梯度裁剪防止梯度爆炸。优化器选择 Adam,训练轮数设置为 100,并保留在验证集上 F1 值最高的模型参数作为最终结果。为保证实验结果的可靠性,在每个数据集上使用不同的随机种子进行 3 次独立实验,取 3 次实验结果的最优值作为性能评估数据。

3.2 数据集

在 NYT、WebNLG 和 SciERC 3 个实体关系提取数据集上评估模型性能。其中,NYT 数据集由《纽约时报》的新闻文章构建而成,包含大量新闻文本。每个文本均包括实体(如人员、组织、地点等)及其关系的注释。标注的关系可以是人与人之间的关系,如雇佣关系或家庭关系;也可以是人与组织、地点等之间的关系,如工作场所关系或隶属关系^[23]。WebNLG 数据集主要来自于 Wikipedia 的结构化数据,包括人物、地点和事件等实体以及国籍、出生日期或事件描述等相关属性。其由数千个三元组及相应的自然语言描述组成,涵盖了广泛的主题和属性^[24]。SciERC 数据集常用于科学事件提取任务,包含科学文献中的实体和事件信息,主要涉及实体之间的各种事件关系^[25]。

各数据集的统计信息见表 1。采用文献^[18]中的分割

方法对数据进行预处理。对于 NYT 和 WebNLG 数据集, 由于缺乏实体类型标注, 统一将实体类型设为“NONE”, 且其标注仅包含实体尾部信息。SciERC 数据集则提供完整实体边界(头部和尾部), 并采用 10 折交叉验证进行评估。3 个数据集均被划分为训练集、验证集和测试集。

Table 1 Statistical information of each dataset

表 1 各数据集统计信息

数据集	训练集	验证集	测试集	实体类型	关系类型
NYT	56 195	5 000	5 000	/	24
WebNLG	5 019	500	703	/	170
SciERC	1 861	275	551	6	7

3.3 基线方法

在 NYT 和 WebNLG 数据集上, 选择 BiRTE^[4]、Cas-Rel^[5]、SGNet^[6]、SPN^[14]、EmRel^[15]、TpLinker^[16]、PFN^[18]和 DiRecRel^[19]等基线模型与本文模型进行比较。在 SciERC 数据集上, 选择 SPE^[8]、SpERT^[9]、UNIRE^[17]和 PFN^[18]等基线模型与本文模型进行比较。

3.4 评价指标

采用精确率(Precision, P)、召回率(Recall, R)和 F1 分数评价各模型性能。具体来说, 对于命名实体识别任务, 只有一个实体的类型和边界均正确时才认为识别的实体是正确的。对于关系抽取任务, 只有当类型、边界和关系类型匹配时才认为预测的实体-关系三元组是正确的。使用部分匹配模型评估 NYT 和 WebNLG 数据集, 仅注释数据集中实体的尾词。对 SciERC 数据集则使用严格的匹配策略, 考虑了注释实体的完整部分。为确保与现有方法比较的公平性, NYT 和 WebNLG 数据集使用 BERT 对数据进行编码处理^[22]; SciERC 数据集则使用 SciBERT 对数据进行编码处理^[26]。

3.5 实验结果与分析

3.5.1 比较实验

所有基线模型通过运行基线代码或参考相应的论文来得到评价指标数据。各模型在 NYT 和 WebNLG 数据集上的实验结果见表 2。可以看出, SJERE 模型在 NYT 数据集上的召回率和 F1 值为最佳, 其中召回率比次优模型

BiRTE 提升 0.2%, F1 值比次优模型 DirecRel 模型提升 0.4%, 但精确率比最优模型 DirecRel 模型降低 0.2%。这是由于 NYT 数据集中含有较多重叠三元组, 而 SJERE 模型在处理重叠三元组方面相较于现有模型没有得到很大提升, 其并未专门设计应对重叠三元组的优化措施。SJERE 模型在 WebNLG 数据集上的精确率和召回率为最佳, 其中精确率比次优模型 EmRel 提升 0.5%, 召回率比次优模型提升 1.0%, 而 F1 值比最优模型 PFN 降低 0.2%, 这是由于该数据集的重叠三元组数量远小于 NYT 数据集。

Table 2 Performance comparison of various models on NYT and WebNLG datasets

表 2 各模型在 NYT 和 WebNLG 数据集上的性能比较 (%)

模型	NYT			WebNLG		
	P	R	F1	P	R	F1
CsaRel	89.8	88.2	89.0	88.3	84.6	86.4
TpLinker	91.4	92.6	92.0	88.9	84.5	86.7
PFN	—	—	92.4	—	—	93.6
EmRel	91.7	92.5	92.1	92.6	92.7	92.6
SGNet	91.2	91.4	91.3	91.8	91.9	91.9
BiRTE	91.9	93.7	92.8	89.0	89.5	89.3
SPN	93.3	91.7	92.5	92.5	92.2	92.3
DirecRel	93.6	92.2	92.9	91.0	89.0	90.0
SJERE	93.4	93.9	93.3	93.1	93.7	93.4

各模型在 SciERC 数据集上的实验结果见表 3。可以看出, SJERE 模型的精确率、召回率和 F1 值均为最佳, 证实了其处理复杂数据时的优秀性能。

在 SciERC 数据集上对本文模型与 PFN 模型进行可视化比较。结果如图 2 所示。可以看出, PFN 模型明显无法

Table 3 Performance comparison of various models on the SciERC dataset

表 3 各模型在 SciERC 数据集上的性能比较 (%)

模型	SciERC		
	P	R	F1
SpERT	40.5	36.8	38.6
SPE	—	—	34.6
PFN	40.6	36.5	38.4
UNIRE	37.3	36.6	36.9
SJERE	41.7	37.2	39.4

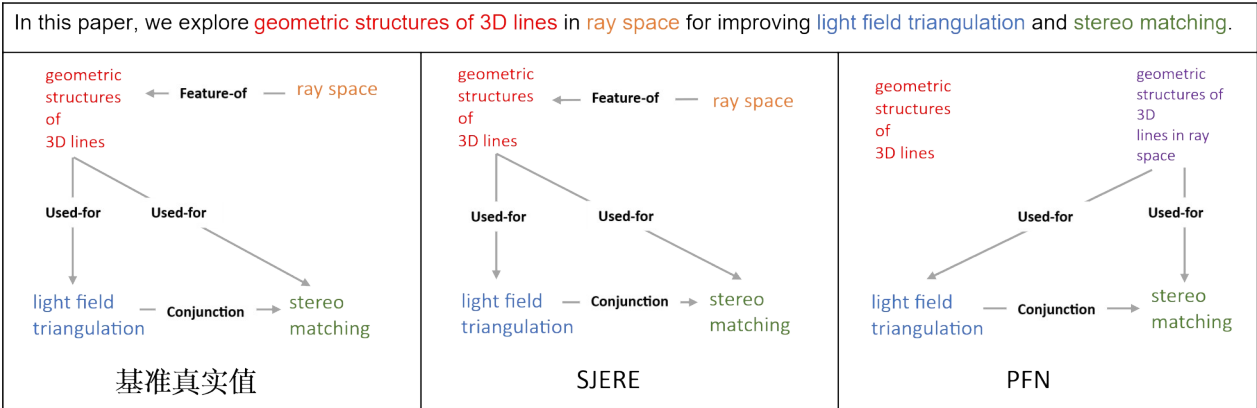


Fig. 2 Visualized comparison

图2 可视化比较

捕捉实体“ray space”和关系三元组(“ray space”、“Feature-of”、“geometric structure of 3D lines”),甚至产生实体“ray space”与“geometric structure of 3D lines”的关联错误。相比之下,本文模型成功识别了给定句子中的所有实体和关系。

3.5.2 消融实验

研究模型各个部分对性能的影响,包括位置矩阵嵌入、共享特征机制、共享表示和卷积层。采用命名实体识别和关系提取两个子任务在 SciERC 数据集上的 F1 分数对不同消融情况下的模型性能进行评估。结果见表 4。可以看出,在没有位置矩阵嵌入的情况下,命名实体识别和关系提取任务的 F1 值分别下降 0.3% 和 0.4%,表明位置嵌入在捕获词间空间关系方面是有效的。在没有纳入共享表示的情况下,两个子任务的 F1 值分别下降 0.2% 和 0.7%,表明两个子任务之间存在着依赖性和相关性。当只使用一维卷积时,两个子任务的 F1 值分别下降 0.6% 和 0.5%,说明深层语义信息和潜在特征对两个子任务的性能有一定影响。缺少共享特征模块会导致两个子任务的 F1 值分别下降 1.0% 和 1.3%,下降幅度最大,说明该模块有效促进了两个子任务之间的双向交互。

Table 4 Ablation experiment result on the SciERC dataset

表 4 在 SciERC 数据集上的消融实验结果 (%)

设置	命名实体识别	偏差	关系提取	偏差
完整状况	69.2	—	39.4	—
缺少嵌入矩阵	68.9	0.3	39.0	0.4
缺少共享表示	69.0	0.2	38.7	0.7
缺少二维卷积	68.6	0.6	38.9	0.5
缺少共享特征机制	68.2	1.0	38.1	1.3

共享特征模块的层数对模型性能亦有一定影响。同样在 SciERC 数据集上使用两个子任务的 F1 分数对 SJERE 模型性能进行评估,不同共享特征模块层数下两个子任务的 F1 分数见表 5。可以看出,当共享特征模块有 3 层时,模型性能达到最佳。当增加至 4 层时,模型性能反而开始下降。这是由于随着共享特征模块层数的增加,过拟合的概率也会增加,从而导致较低层次中的一些重要特征丢失,影响模型性能。

Table 5 The impact of shared feature module layers on model performance

表 5 共享特征模块层数对模型性能的影响 (%)

共享特征层数	命名实体识别	关系提取
N=1	68.9	38.6
N=2	69.0	38.9
N=3	69.2	39.4
N=4	68.3	38.8

4 结语

本文提出一种基于卷积的联合实体关系提取共享特征网络 SJERE。该网络采用并行编码的方法,使用两个独

立的多层感知器来获取命名实体识别和关系提取两个子任务的不同特征表示,有效避免了两个子任务之间的特征误分类。同时,在并行编码的基础上引入一种共享特征机制,通过整合从一维和二维卷积中获得的特征映射以形成共享特征,促进命名实体识别与关系提取任务之间的交互,增强了模型理解和表示文本,以及捕获词对级关系的能力。在 NYT、WebNLG、SciERC 3 个数据集上的实验结果表明, SJERE 模型的精确率、召回率和 F1 分数与基线模型相比展现出一定优势。本文模型虽然比现有联合模型有了一定改进,但在面临重叠三元组的情况时提升有限,后续工作将重点改进模型架构和训练策略,以更有效地处理重叠三元组,提高识别效果。

参考文献:

- [1] RIEDEL S, YAO L, MCCALLUM A, et al. Relation extraction with matrix factorization and universal schemas [C]//Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2013: 74–84.
- [2] GUPTA V, LEHAL G S. A survey of text summarization extractive techniques[J]. Journal of Emerging Technologies in Web Intelligence, 2010, 2 (3): 258–268.
- [3] DIEFENBACH D, LOPEZ V, SINGH K, et al. Core techniques of question answering systems over knowledge bases: a survey[J]. Knowledge and Information Systems, 2018, 55: 529–569.
- [4] REN F, ZHANG L, ZHAO X, et al. A simple but effective bidirectional framework for relational triple extraction [C]//Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, 2022: 824–832.
- [5] 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.
- [6] LIANG J, HE Q, ZHANG D, et al. Extraction of joint entity and relationships with soft pruning and GlobalPointer[J]. Applied Sciences, 2022, 12 (13): 6361.
- [7] LUAN Y, WADDEN D, HE L, et al. A general framework for information extraction using dynamic span graphs [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019: 3036–3046.
- [8] WANG Y, SUN C, WU Y, et al. Pre-training entity relation encoder with intra-span and inter-span information [C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 2020: 1692–1705.
- [9] EBERTS M, ULGES A. Span-based joint entity and relation extraction with transformer pre-training [DB/OL]. <https://arxiv.org/abs/1909.07755>.
- [10] 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 Linguistics, 2018: 506–514.
- [11] ZENG X, HE S, ZENG D, et al. Learning the extraction order of multiple relational facts in a sentence with reinforcement learning [C]//Proceed-

- ings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019: 367–377.
- [12] CABOT P L H, NAVIGLI R. REBEL: relation extraction by end-to-end language generation [C]//Findings of the Association for Computational Linguistics, 2021: 2370–2381.
- [13] LEWIS M, LIU Y, GOYAL N, et al. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020: 7871–7880.
- [14] SUI D, ZENG X, CHEN Y, et al. Joint entity and relation extraction with set prediction networks [J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(9): 12784–12795.
- [15] XU B, WANG Q, LYU Y, et al. EmRel: joint representation of entities and embedded relations for multi-triple extraction [C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2022: 659–665.
- [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] WANG Y, SUN C, WU Y, et al. UniRE: a unified label space for entity relation extraction [C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 2021: 220–231.
- [18] YAN Z, ZHANG C, FU J, et al. A partition filter network for joint entity and relation extraction [C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021: 185–197.
- [19] SHANG Y M, HUANG H, SUN X, et al. Relational triple extraction: one step is enough [C]//31st International Joint Conference on Artificial Intelligence, 2022: 4360–4366.
- [20] ZHONG Z, CHEN D. A frustratingly easy approach for entity and relation extraction [C]//2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021: 50–61.
- [21] WANG J, LU W. Two are better than one: joint entity and relation extraction with table-sequence encoders [C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 2020: 1706–1721.
- [22] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019: 4171–4186.
- [23] RIEDEL S, YAO L, MCCALLUM A. Modeling relations and their mentions without labeled text [C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2010: 148–163.
- [24] GARDENT C, SHIMORINA A, NARAYAN S, et al. Creating training corpora for nlg micro-planning [C]//55th Annual Meeting of the Association for Computational Linguistics, 2017: 179–188.
- [25] LUAN Y, HE L, OSTENDORF M, et al. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction [C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018: 3219–3232.
- [26] BELTAGY I, LO K, COHAN A. SciBERT: a pretrained language model for scientific text [C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019: 3615–3620.

(责任编辑:尹晨茹)