

基于对比学习与梯度惩罚的 实体关系联合抽取模型

张强, 曾俊玮, 陈锐

(东北石油大学 计算机与信息技术学院, 黑龙江 大庆 163318)

摘要: 针对使用全局指针网络进行实体关系抽取时特征信息不明显的实体关系类型数据稀疏问题, 以及数据中存在的类别不平衡和错误标注问题, 提出一种基于对比学习和梯度惩罚方法并使用改进的 RoBERTa 预训练模型的实体关系联合抽取模型, 在阿里天池中文医疗信息处理评测基准数据集 CBLUE2.0 上进行实验的结果表明, 该模型相比全局指针网络效果更优, 能更有效完成复杂数据的实体关系抽取。

关键词: 实体关系抽取; 对比学习; 梯度惩罚; RoBERTa 预训练模型; 全局指针网络

中图分类号: TP18 **文献标志码:** A **文章编号:** 1671-5489(2024)05-1155-08

Entity-Relation Joint Extraction Model Based on Contrastive Learning and Gradient Penalty

ZHANG Qiang, ZENG Junwei, CHEN Rui

(School of Computer and Information Technology, Northeast Petroleum University,
Daqing 163318, Heilongjiang Province, China)

Abstract: Aiming at the problem of sparse entity relationship type data with unclear feature information when using global pointer networks for entity relationship extraction, as well as the problem of class imbalance and incorrect labeling in the data, we proposed a entity-relation joint extraction model based on contrastive learning and gradient penalty methods while utilizing an enhanced RoBERTa pre-trained model. Experimental results on the Alibaba Tianchi Chinese medical information processing benchmark CBLUE2.0 dataset show that this model outperforms the global pointer network, and can more effectively extract entity relationship from complex data.

Keywords: entity relation extraction; contrastive learning; gradient penalty; RoBERTa pre-trained model; global pointer network

0 引言

实体关系抽取是信息抽取^[1]的一个子任务, 广泛应用于自然语言处理领域, 其目的是从非结构化文本中抽取“主体-关系-客体”形式的三元组, 是构建知识图谱、语义搜索引擎和推荐系统的关键任

收稿日期: 2023-08-20.

第一作者简介: 张强(1982—), 男, 汉族, 博士, 教授, 从事神经网络及智能进化算法的研究, E-mail: dqpi_zq@163.com. **通信作者简介:** 曾俊玮(1998—), 男, 汉族, 硕士研究生, 从事自然语言处理的研究, E-mail: 1275915941@qq.com.

基金项目: 国家自然科学基金(批准号: 42002138)、黑龙江省研究生课程思政建设项目(批准号: YJSKCSZ_202303)和黑龙江省优秀青年教师基础研究支持计划项目(批准号: YQJH2023073).

务. 如图 1 所示, 三元组中的实体关系类别主要分为三类: 1) Normal, 即三元组之间没有重叠现象; 2) EPO(entity pair overlap), 即三元组之间共享一对实体对; 3) SEO(single entity overlap), 即三元组之间仅共享一个实体. 在极端情况下 EPO 和 SEO 可能同时出现. 实体关系抽取任务中的重难点任务是如何处理重叠三元组的问题, 即如何处理 EPO 和 SEO 数据.

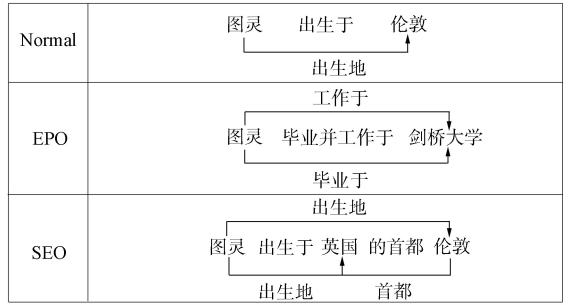


图 1 三元组重叠关系类型示例

Fig. 1 Example of triple overlap relationship types

本文模型基于改进的 RoBERTa 预训练模型和全局指针网络模型, 以获得所有可能关系的主客体起始位置和结束位置, 预测主客体对的关系, 并组成三元组, 可在一定程度上解决 EPO 和 SEO 问题. 但受制于训练集数据, 一旦复杂数据出现特征信息不明显且稀疏的问题, 就很可能导致无法抽取复杂的重叠三元组, 因此本文采用对比学习方法提升模型的泛化能力, 以更好地理解 and 处理语义关系. 又考虑到数据的错误标注以及实体关系类型数据不均衡可能导致的模型性能下降, 引入了梯度惩罚项, 以提升模型的对抗鲁棒性. 在数据集 CBLUE2.0 上的实验结果表明, 本文模型能有效提升复杂数据环境下的三元组抽取性能.

1 实体关系抽取模型相关理论

1.1 实体关系抽取

实体与关系的联合抽取问题传统方法分为两个独立的子任务进行处理^[2], 即命名实体识别(named entity recognition, NER)和关系抽取(relation extraction, RE)^[3], 这类模型称为流水线模型. 流水线模型存在误差传递^[4]、忽略两个子任务之间的联系和信息冗余等问题.

针对上述问题, 为提升实体关系抽取的效率, 增强模型应对重叠三元组的能力, 同时抽取实体与关系的模型研究备受关注. Zheng 等^[5]提出了一种新的标注模式, 基于这种模式将联合抽取转换为标注问题. Zeng 等^[6]提出了一种端到端神经模型和复制机制组成的方法, 能直接从原始文本中抽取关系事实, 提升模型在不确定环境下的鲁棒性. Fu 等^[7]用将文本建模为关系图的方式联合处理实体和关系的抽取任务. Wei 等^[8]将关系视为头实体映射到尾实体的函数, 提出了一种端到端的级联二元标记框架 CasRel. Su 等^[9]基于全局指针的思想, 通过点积型注意力与多标签交叉熵^[10]实现了用统一的方式处理嵌套和非嵌套命名实体识别.

1.2 对比学习

实体关系抽取既要考虑实体对之间的语义相似度, 又要面对样本不均衡的问题. 这通常是由于负样本远多于正样本导致的, 而对比学习方法旨在通过数据扩增等方法构成正样本对与负样本对并进行相似度对比, 以强化语义表示学习的过程, 从而达到缓解样本不均衡问题, 优化样本数据的特征表示, 提升模型泛化能力的目的.

对比学习起源于同类数据对比的思想^[11], 以及自监督学习中的实例判别任务^[12], 其核心思想是通过将正样本(相似样本)与负样本(不相似样本)进行对比学习样本的特征表示. Mikolov 等^[13]提出了 Word2Vec 模型, 探索词语和短语的分布式表示及其组合性, 将对比学习方法引入自然语言处理领域. Yang 等^[14]通过对比训练样本与衍生样本之间的差异, 使模型更关注长距离依赖, 解决了神经机器翻译中的漏译错误问题. 李晓等^[15]将对比学习机制引入论元语义的表示学习过程, 利用了高困惑度样本的信息, 通过对比学习提高隐式关系的表示能力, 提升论元语义编码器的鲁棒性, 对抗训练梯度惩罚. 高怡等^[16]考虑到小样本环境, 引入对比学习成功提升了模型在小样本环境下的效果.

1.3 梯度惩罚

实体关系抽取任务中由于模型通常较复杂, 因此容易在训练集上过拟合, 导致最终抽取时效果欠佳. 同时考虑到训练数据中可能存在的微小扰动, 可使用梯度惩罚减少模型过拟合的风险, 提高对抗

动的鲁棒性. 梯度惩罚^[17]是一种在生成对抗网络(GAN)中用于改善训练稳定性和生成样本质量的技术, 它旨在解决原始 GAN 训练中的模式崩溃和梯度消失问题, GAN 模型在各领域中都得到了广泛应用. 冯永等^[18]提出了 Wasserstein 图像循环生成对抗网络模型, 并添加了梯度惩罚以解决模型优化困难的问题. 王先武等^[19]针对传统数字岩心重构技术存在的问题, 提出了一种基于带梯度惩罚深度卷积生成对抗网络. Fedus 等^[20]通过在生成对抗网络中使用梯度惩罚增加生成样本的真实性和多样性, 从而改进文本生成任务中的缺失词填充. Ross 等^[21]对输入梯度进行正则化, 使深度神经网络模型对输入数据的微小扰动具有强鲁棒性.

2 模型设计

2.1 模型结构

本文使用经过改进的 RoBERTa 预训练模型作为编码器, 以提升模型语义表示学习与句子级别的理解能力, 适应实体关系联合抽取任务的需求, 捕捉更多词语与句子之间的语义信息和关联性. 并采用融合对比学习的全局指针网络, 在利用 Dropout 模块实现数据扩增、增加样本多样性的基础上, 抽取头尾实体并对预测其可能的关系类型. 同时在损失函数中增加梯度惩罚项, 提升对抗鲁棒性. 模型整体框架如图 2 所示.

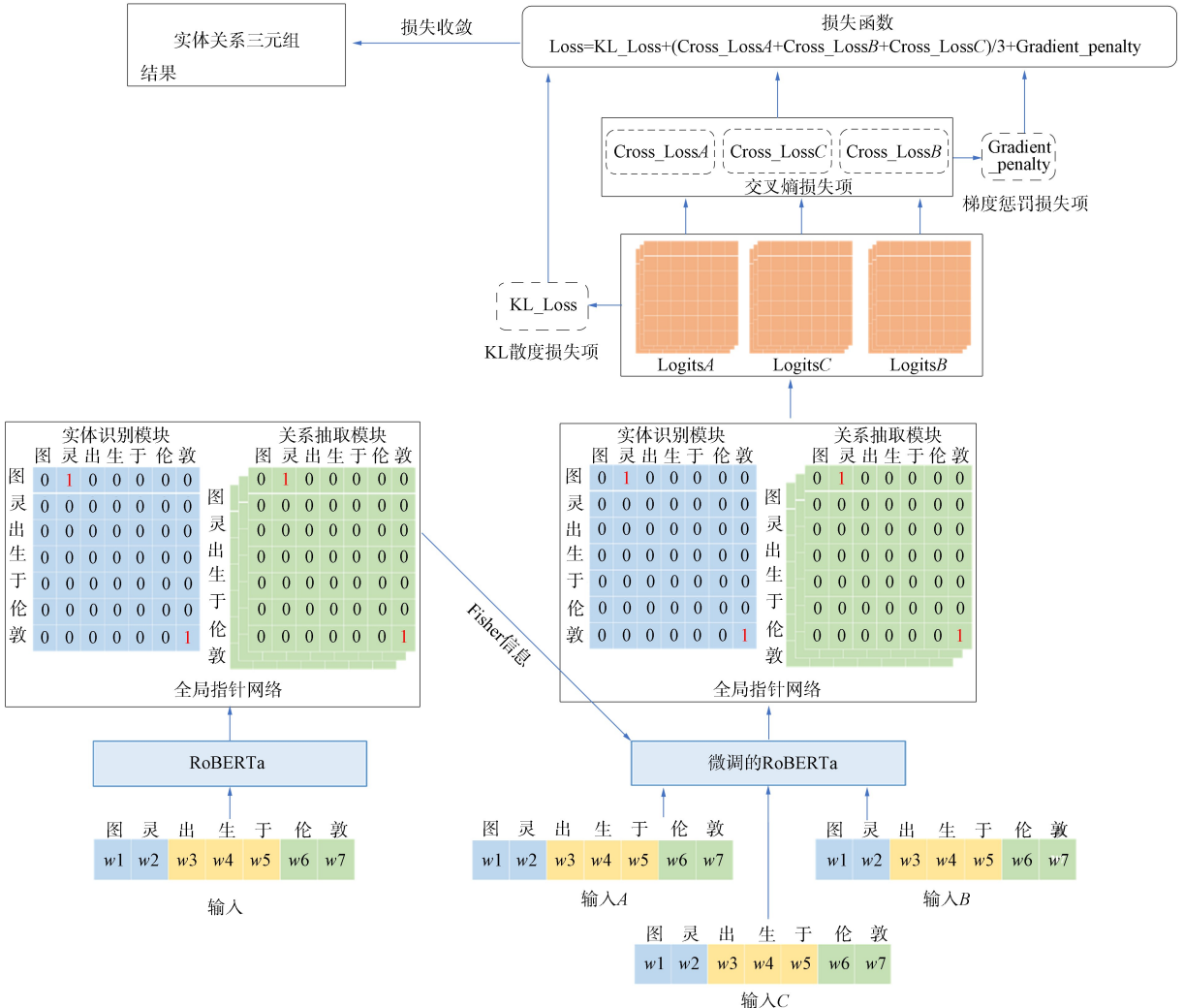


图 2 本文模型架构

Fig. 2 Architecture of proposed model

2.2 基于 Child-Tuning 优化的 RoBERTa 预训练模型

RoBERTa 是基于 BERT 改进的预训练模型, 其通过大规模数据的训练学习到通用的语言表示,

可为实体关系抽取提供更好的语义表示学习能力,减少数据需求.而微调可进一步调整预训练模型的参数以适应具体的下游任务,提升模型性能和效率.但预训练模型的规模与参数量相比于下游任务的训练数据差距过大,因此直接微调易导致过拟合现象,从而影响模型的最终性能.

针对上述问题, Xu 等^[22]提出了一种在仅更新预训练模型中部分网络参数的方法——Child-Tuning,即在预训练模型中选择一个子网络进行优化,从而在一定程度上避免过拟合风险.本文使用 Adam 优化器对 RoBERTa 进行微调,在获取下游任务网络的所有参数基础上,计算其 Fisher 信息:

$$F^{(i)}(w) = \frac{1}{|D|} \sum_{j=1}^D \left(\frac{\partial \log p(y_j | x_j; w)}{\partial w^{(i)}} \right)^2, \tag{1}$$

其中 D 为所有样本的数量, x, y 为输入和输出.并按计算结果对参数的重要性进行排序,获得其中有权重衰减参数的前 90% 作为子网络用于 Adam 优化器.

2.3 基于对比学习的改进全局指针网络

全局指针网络是指针网络的一种优化设计,旨在解决指针网络实体起止位置分布识别导致的训练预测不一致性问题,其能实现多头识别嵌套主客体的头实体对与尾实体对,全局指针网络识别实体如图 3 所示.在此基础上,通过头实体对与尾实体对预测结果的交集判断其可能存在的关系.图 4 为全局指针网络判断实体对关系的示意图.



图 3 全局指针网络实体识别示意图

Fig. 3 Schematic diagram of global pointer network entity recognition

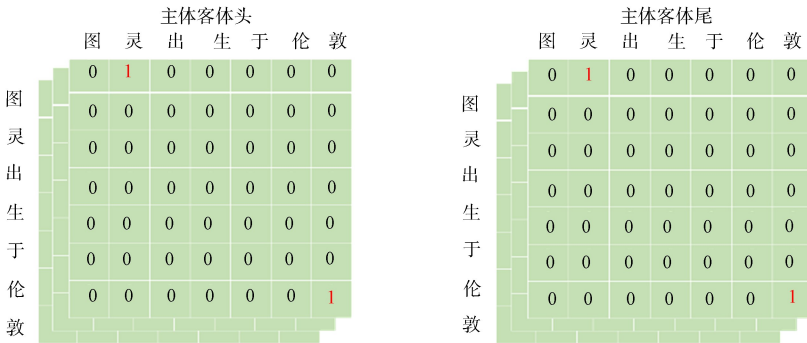


图 4 全局指针网络关系抽取示意图

Fig. 4 Schematic diagram of global pointer network relation extraction

全局指针网络在非嵌套环境下效果较好,但在嵌套(即重叠三元组)情形下,数据量和实体特征的独特性对实体关系抽取效果的影响被扩大^[23],当数据集中特征不明显的实体关系类型出现数据稀疏的情况时,可能会存在一批完全无法抽取的实体关系类型.为解决样本特征信息不明显、样本数据稀疏、泛化性差的问题,本文引入对比学习强化语义表示学习的过程,使语义相似的样本相距更近,使语义不相似的样本相距更远,以提升语义表示层面的效果,降低模型对数据量的依赖,并提升在重叠三元组环境下的泛化性.

对比学习的标准流程是将样本 A 通过数据扩增产生的结果均视为正样本,而将同一个批次内除 A 外的样本均视为负样本.而在自然语言处理领域中,由于自然语言丰富的语义结构和复杂的上下文依

赖关系, 数据扩增很可能会引入语义失真或矛盾.

本文将同一样本重复 3 份: 样本 c 通过预训练模型且不进行 Dropout 操作, 样本 a 和 b 通过预训练模型且进行 Dropout 操作. 因为 Dropout 随机增加乘性噪声的特点, 因此即使是同一样本, a 和 b 两次输出的特征向量也不完全相同. 且 3 个输出在语义表示上有相似性, 可被视为一组正样本, 再加上同一批次中剩余的样本, 即获得了对比学习所需的正负样本. 由于模型使用了 RoBERTa 作为预训练模型, 而 RoBERTa 中自带有 Dropout 操作, 因此正样本的实现过程即为将同一样本输入 RoBERTa 中 3 次.

获得正样本组后, 通过对称 KL(Kullback-Leibler)散度分别对比 ab, ac 两组样本对预测结果的概率分布并加权求和, 将其作为损失函数的一部分, 其中对称 KL 散度公式为

$$D(p, q) = \sum_{i=1}^n (p_i - q_i)(s_i - t_i). \tag{2}$$

在未经激活函数处理模型原始输出 Logits 的层面上, 对称 KL 散度公式为

$$D(p, q) = \sum_{i=1}^n (f(s_i) - f(t_i))(s_i - t_i) \leq f(s) - f(t), \quad s - t > 0, \tag{3}$$

其中 f 为 Softmax 变换.

考虑到全局指针网络的预测结果并非一个概率分布, 再结合所使用的多标签交叉熵损失函数, 使用 Sigmoid 作为激活函数, 则对称 KL 散度公式可表示为

$$D(p, q) = \sum_{i=1}^n (\sigma(s_i) - \sigma(t_i))(s_i - t_i) \leq \sigma(s) - \sigma(t), \quad s - t > 0. \tag{4}$$

因此, 全局指针网络上加上对比学习的损失函数为

$$\text{Loss} = \text{loss}_{\text{cross}} + \alpha \text{loss}_{\text{KL}}, \tag{5}$$

其中 α 为 KL 散度的超参数.

2.4 基于梯度惩罚项的改进损失函数

计算模型所用训练集的先验分布概率, 发现训练集数据存在一定程度的类别分布不均衡问题, 并考虑到训练集中可能存在错误标注引发的噪声影响, 因此在损失函数中添加梯度惩罚项, 以提升模型的对抗鲁棒性.

本文模型的输入梯度采用预训练模型最后一个隐藏层的输出计算, 并考虑到对比学习中的正样本对影响, 分别计算正样本所对应的梯度惩罚项并加权求和, 计算公式为

$$\text{GP} = \frac{1}{2} \sum_i (\partial(\text{loss}_{\text{cross}}) / \partial(\text{hidden}_i))^2. \tag{6}$$

之后在基于对比学习的全局指针网络损失函数上, 增加输入梯度惩罚项, 三部分共同学习, 可表示为

$$\text{Loss} = \text{loss}_{\text{cross}} + \alpha \text{loss}_{\text{KL}} + \beta \text{loss}_{\text{GP}}, \tag{7}$$

其中 β 为梯度惩罚项的超参数.

3 实验对比分析

3.1 数据处理及评价指标

本文选择在数据集 CBLUE^[24]基础上扩充生成的数据集 CBLUE2.0 进行实验. 数据集的关系集合包含 53 种关系, 其中 10 组同义词子关系, 43 组其他子关系. 数据集开放训练集 14 339 条数据, 验证集 3 585 条数据, 由于未公开测试集数据, 因此本文选用原来的验证集数据作为测试集.

根据训练集样本数据中句子的重叠模式, 将句子划分为 Normal, EPO 和 SEO 三类, 其中 Normal 有 5 811 条数据, SEO 有 8 339 条数据, EPO 有 189 条数据. 数据集中有 59.5% 的重叠三元组.

本文模型性能评价指标采用召回率 R 、精确率 P 和 F_1 值, 其公式分别为

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{8}$$

$$P = \frac{TP}{TP + FP}, \quad (9)$$

$$F_1 = \frac{2RP}{R + P}, \quad (10)$$

其中 TP 表示三元组中被正确抽取出的三元组数量, FN 表示未被抽取出的三元组数量, FP 表示错误抽取出的三元组数量, R 表示正确抽取的三元组在实际三元组中的比例, P 表示正确抽取的三元组在所有抽取三元组中的比例, F_1 值为一种综合衡量召回率和精确率的指标.

3.2 参数设置

本文使用 Pytorch 实现模型, 在数据集 CBLUE2.0 上的批处理大小为 8, 选用 RoBERTa_{l12_zh} 作为基本的预训练模型, 句子输入的最大长度为 256, 采用经过优化的 BertAdam 优化器, 初始学习率为 0.000 02, Warmup 设为 0.1, 并设置了仅适用于下游任务相关的部分参数的权重衰减为 0.01. 在 Google Colab p100 GPU 上进行实验, 训练 100 代, 最终选择 F_1 值最优的模型. 实验的其他参数设置如下: $\alpha=1$, $\beta=0.1$, Child-Tuning-rate 为 0.9.

3.3 结果对比分析

本文模型与数据集提供的任务基线模型对比结果列于表 1.

表 1 本文模型与基线模型对比结果

Table 1 Comparison results between proposed model and baseline models

| 模型 | F_1 | 模型 | F_1 |
|-----------------------|-------|----------------|-------|
| BERT-base | 54.0 | ALBERT-xxlarge | 47.6 |
| BERT-wwm-ext-base | 54.0 | ZEN | 50.1 |
| RoBERTa-large | 54.4 | MacBERT-base | 53.2 |
| RoBERTa-wwm-ext-base | 53.7 | MacBERT-large | 51.6 |
| RoBERTa-wwm-ext-large | 55.9 | PCL-MedBERT | 49.1 |
| ALBERT-tiny | 35.9 | 本文 | 62.8 |

为评估本文模型每部分的贡献进行消融实验, 对比结果列于表 2. 由表 2 可见, 分别去掉 Child-Tuning D、对比学习、梯度惩罚项时 F_1 值都会下降, 说明每一部分的加入都对模型性能提升有贡献, 从而验证了本文模型的有效性.

表 2 消融实验结果

Table 2 Results of ablation experiment

| 模型 | F_1 | P | R | 模型 | F_1 | P | R |
|-----------------|-------|------|------|-------|-------|------|------|
| 本文模型 | 62.8 | 65.0 | 60.7 | -梯度惩罚 | 62.4 | 64.9 | 60.0 |
| -Child-Tuning D | 62.5 | 64.7 | 60.4 | -对比学习 | 60.8 | 60.5 | 61.1 |

下面将本文的对比学习方法与目前流行的 R-drop^[25] 方法进行对比, 对比实验分别在全局指针网络、全局指针网络+梯度惩罚、全局指针网络+梯度惩罚+Child-Tuning D 微调 3 个模型上进行, 结果列于表 3. 由表 3 可见, 在数据集 CBLUE 上, 对以全局指针网络为基础的模型, 本文方法总体性能提升优于 R-drop 方法, 证明了本文对比学习方法的有效性.

表 3 本文对比学习方法与 R-drop 方法实验结果对比

Table 3 Comparison of experimental results between proposed contrastive learning method and R-drop method

| 模型 | 本文对比学习方法 | | | R-drop 方法 | | |
|----------------------------|----------|------|------|-----------|------|------|
| | F_1 | P | R | F_1 | P | R |
| 全局指针网络 | 62.4 | 64.9 | 60.0 | 61.6 | 62.2 | 61.0 |
| 全局指针网络+梯度惩罚 | 62.5 | 64.7 | 60.4 | 61.8 | 62.9 | 60.7 |
| 全局指针网络+梯度惩罚+Child-Tuning D | 62.8 | 65.0 | 60.7 | 61.9 | 63.0 | 60.8 |

为进一步验证本文模型处理重叠三元组的能力, 在消融实验的基础上, 将测试集中的样本数据按重叠模式划分为 Normal, EPO 和 SEO 三类, 并分别评价 3 种重叠类型数据下的模型性能. 其中测试集数据的分类结果列于表 4, 在不同重叠模式下的模型性能结果列于表 5. 由表 4 和表 5 可见, 在 SEO

和 Normal 的情形下, 改进模型的效果均优于基本全局指针网络, 且在 SEO 重叠情形下的准确率明显高于基础模型, 而 EPO 类型的效果有所下降的部分原因是由于其训练集和测试集数据远少于 SEO 和 Normal 类, 导致训练效果不佳且预测结果参考性下降。

表 4 测试集数据三元组重叠类型统计

Table 4 Statistics of triple overlap types in test set data

| 三元组重叠类型 | 数量 | 三元组重叠类型 | 数量 |
|---------|-------|---------|----|
| Normal | 1 521 | EPO | 40 |
| SEO | 2 024 | | |

表 5 不同重叠类型下的模型性能对比

Table 5 Model performance comparison under different overlap types

| 模型 | SEO | | | Normal | | | EPO | | |
|---------------|-------|------|------|--------|------|------|-------|------|------|
| | F_1 | P | R | F_1 | P | R | F_1 | P | R |
| 本文模型 | 66.1 | 70.6 | 62.1 | 48.6 | 43.8 | 54.7 | 61.2 | 75.3 | 51.5 |
| -Child-Tuning | 65.3 | 70.0 | 61.2 | 50.5 | 45.0 | 57.6 | 61.4 | 73.1 | 53.0 |
| -梯度惩罚项 | 65.3 | 70.3 | 61.0 | 49.4 | 44.5 | 55.5 | 63.9 | 77.8 | 54.1 |
| -对比学习 | 64.0 | 65.9 | 62.2 | 46.8 | 40.2 | 56.0 | 64.5 | 74.1 | 57.1 |

综上所述, 本文基于对比学习与梯度惩罚改进全局指针网络, 在不影响空间复杂度、时间复杂度略上升的情形下, 使其能在数据集类别不均衡、存在噪声干扰的环境下, 获得足够的泛化性和对抗鲁棒性. 使用基于 Child-Tuning 的策略微调预训练模型, 缓解了过拟合风险并使预训练模型与下游任务模型更匹配. 虽然本文对基础模型的改进有一定的性能提升, 使模型在数据集 CBLUE 上的预测结果得到了提高. 但依然存在对数据样本过于稀疏的复杂实体重叠语句(例如 EPO 重叠关系的样本)无法取得理想的效果问题。

参 考 文 献

[1] 衡红军, 苗菁. 语义与句法信息加强的二元标记实体关系联合抽取 [J]. 计算机工程, 2023, 49(4): 77-84. (HENG H J, MIAO J. Joint Extraction of Binary Tagging Entity Relation for Enhanced Semantic and Syntactic Information [J]. Computer Engineering, 2023, 49(4): 77-84.)

[2] 贾宝林, 尹世群, 王宁朝. 基于门控多层感知机的端到端实体关系联合抽取 [J]. 中文信息学报, 2023, 37(3): 143-151. (JIA B L, YIN S Q, WANG N C. An End-to-End Joint Extraction of Entity and Relation Based on MLPs with Gating [J]. Journal of Chinese Information Processing, 2023, 37(3): 143-151.)

[3] 张仰森, 刘帅康, 刘洋, 等. 基于深度学习的实体关系联合抽取研究综述 [J]. 电子学报, 2023, 51(4): 1093-1116. (ZHANG Y S, LIU S K, LIU Y, et al. Joint Extraction of Entities and Relations Based on Deep Learning: A Survey [J]. Acta Electronica Sinica, 2023, 51(4): 1093-1116.)

[4] 廖涛, 孙皓洁, 张顺香. 基于跨度和特征融合的实体关系联合抽取模型 [J]. 计算机工程, 2023, 49(6): 107-114. (LIAO T, SUN H J, ZHANG S X. Entity-Relation Joint Extraction Model Based on Span and Feature Fusion [J]. Computer Engineering, 2023, 49(6): 107-114.)

[5] ZHENG S C, WANG F, BAO H Y, et al. Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme [EB/OL]. (2017-06-07)[2023-01-05]. <https://arxiv.org/abs/1706.05075>.

[6] ZENG X R, ZENG D J, HE S Z, 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. [S.l.]: ACL, 2018: 506-514.

[7] FU T J, LI P H, MA W Y. 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. [S.l.]: ACL, 2019: 1409-1418.

[8] WEI Z P, SU J L, WANG Y, et al. A Novel Cascade Binary Tagging Framework for Relational Triple Extraction [EB/OL]. (2019-09-07)[2023-02-01]. <https://arxiv.org/abs/1909.03227>.

[9] SU J L, MURTADHA A, PAN S F, et al. Global Pointer: Novel Efficient Span-Based Approach for Named

- Entity Recognition [EB/OL]. (2022-08-05)[2023-01-15]. <https://arxiv.org/abs/2208.03054>.
- [10] SU J L, ZHU M R, MURTADHA A, et al. Zlpr: A Novel Loss for Multi-label Classification [EB/OL]. (2022-08-05)[2023-01-20]. <https://arxiv.org/abs/2208.02955>.
- [11] HADSELL R, CHOPRA S, LECUN Y. Dimensionality Reduction by Learning an Invariant Mapping [C]//2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). Piscataway, NJ: IEEE, 2006: 1735-1742.
- [12] WU Z R, XIONG Y J, YU S, et al. Unsupervised Feature Learning via Non-parametric Instance Discrimination [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 3733-3742.
- [13] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed Representations of Words and Phrases and Their Compositionality [C]//Proceedings of the 26th International Conference on Neural Information Processing Systems. New York: ACM, 2013: 3111-3119.
- [14] YANG Z H, CHENG Y, LIU Y, et al. Reducing Word Omission Errors in Neural Machine Translation: A Contrastive Learning Approach [C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. [S.l.]: ACL, 2019: 6191-6196.
- [15] 李晓, 洪宇, 窦祖俊, 等. 基于高困惑样本对比学习的隐式篇章关系识别 [J]. 中文信息学报, 2022, 36(11): 38-49. (LI X, HONG Y, DOU Z J, et al. Contrastive Learning with Confused Samples for Implicit Discourse Relation Recognition [J]. Journal of Chinese Information Processing, 2022, 36(11): 38-49.)
- [16] 高怡, 纪焘, 吴苑斌, 等. 基于标签增强和对比学习的鲁棒小样本事件检测 [J]. 中文信息学报, 2023, 37(4): 98-108. (GAO Y, JI T, WU Y B, et al. Robust Few Shot Event Detection Based on Label Augmentation and Contrastive Learning [J]. Journal of Chinese Information Processing, 2023, 37(4): 98-108.)
- [17] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved Training of Wasserstein Gans [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 5769-5779.
- [18] 冯永, 张春平, 强保华, 等. GP-WIRGAN: 梯度惩罚优化的 Wasserstein 图像循环生成对抗网络模型 [J]. 计算机学报, 2020, 43(2): 190-205. (FENG Y, ZHANG C P, QIANG B H, et al. GP-WIRGAN: A Novel Image Recurrent Generative Adversarial Network Model Based on Wasserstein and Gradient Penalty [J]. Chinese Journal of Computers, 2020, 43(2): 190-205.)
- [19] 王先武, 张挺, 吉欣, 等. 基于带梯度惩罚深度卷积生成对抗网络的页岩三维数字岩心重构方法 [J]. 计算机应用, 2021, 41(6): 1805-1811. (WANG X W, ZHANG T, JI X, et al. 3D Shale Digital Core Reconstruction Method Based on Deep Convolutional Generative Adversarial Network with Gradient Penalty [J]. Journal of Computer Applications, 2021, 41(6): 1805-1811.)
- [20] FEDUS W, GOODFELLOW I, DAI A M. Maskgan: Better Text Generation via Filling in the _ [EB/OL]. (2018-01-23)[2023-02-11]. <https://arxiv.org/abs/1801.07736>.
- [21] ROSS A S, DOSHI-VELEZ F. Improving the Adversarial Robustness and Interpretability of Deep Neural Networks by Regularizing Their Input Gradients [C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI, 2018: 1660-1669.
- [22] XU R X, LUO F L, ZHANG Z Y, et al. Raise a Child in Large Language Model: Towards Effective and Generalizable Fine-Tuning [EB/OL]. (2021-09-13)[2023-03-01]. <https://arxiv.org/abs/2109.05687>.
- [23] 甘子发, 咎红英, 关同峰, 等. CHIP 2020 评测任务 2 概述: 中文医学文本实体关系抽取 [J]. 中文信息学报, 2022, 36(6): 101-108. (GAN Z F, ZAN H Y, GUAN T F, et al. Overview of CHIP 2020 Shared Task 2: Entity and Relation Extraction in Chinese Medical Text [J]. Journal of Chinese Information Processing, 2022, 36(6): 101-108.)
- [24] ZHANG N Y, CHEN M S, BI Z, et al. Cblue: A Chinese Biomedical Language Understanding Evaluation Benchmark [EB/OL]. (2021-07-15)[2023-01-10]. <https://arxiv.org/abs/2106.08087>.
- [25] WU L, LI J, WANG Y, et al. R-Drop: Regularized Dropout for Neural Networks [J]. Advances in Neural Information Processing Systems, 2021, 34: 10890-10905.