

基于预训练大语言模型 BERT 的关系抽取方法*

唐梓烨¹⁾

¹⁾ (东南大学 人工智能学院, 南京 中国 211189)

摘要 关系抽取是从非结构化文本中提取实体之间关系的核心任务, 传统的流水线方法将命名实体识别与关系抽取分开处理, 而近年的端到端方法则通过联合建模实现更好的性能。本研究提出了一种基于预训练大语言模型 BERT 的关系抽取方法, 针对中文医疗领域的关系分类任务进行了实验。我们采用了 Chinese-RoBERTa-wwm-ext 作为预训练模型, 并设计了多种文本编码方式, 如实体标记法和问答式编码等, 通过微调模型提升了关系抽取的准确性。实验结果表明, 基于 Chinese-RoBERTa-wwm-ext 的模型在关系分类任务中表现优异。与其他预训练模型 (BioBERT、PCL-MedBERT) 相比, 我们的方法在 Micro-F1 值上分别提高了 49.4% 和 1.5%, 验证了在本任务中中文领域特定优化和实体标记方式的重要性。此外, 我们还通过类别增强策略有效缓解了难类别问题, 进一步提高了模型在难以区分的类别上的表现。该研究为中文医疗文本中的关系抽取提供了方法, 具有较高的实际应用价值。我们的代码开源在: <https://github.com/Warma10032/Relation-Extraction>

关键词 关系抽取; BERT; 预训练语言模型; 中文医疗文本; 关系分类

中图法分类号 TP183

Relationship Extraction Method Based on Pre-trained BERT

Tang Ziye¹⁾

¹⁾(Department of Artificial Intelligence, Southeast University, Nanjing China 211189)

Abstract

Relation extraction is a core task of extracting relationships between entities from unstructured text. Traditional pipeline methods treat named entity recognition and relation extraction as separate processes, while recent end-to-end approaches achieve better performance through joint modeling. This study proposes a relation extraction method based on the pre-trained large language model BERT, and conducts experiments on the relation classification task in the Chinese medical domain. We used Chinese-RoBERTa-wwm-ext as the pre-trained model and designed several text encoding methods, such as entity tagging and question-answer encoding. By fine-tuning the model, we improved the accuracy of relation extraction. Experimental results show that the model based on Chinese-RoBERTa-wwm-ext performs excellently in relation classification tasks. Compared to other pre-trained models (BioBERT, PCL-MedBERT), our method improves the Micro-F1 score by 49.4% and 1.5%, respectively, demonstrating the importance of Chinese domain-specific optimization and entity tagging in this task. Furthermore, we effectively mitigated the "hard class" problem through a class enhancement strategy, further improving the model's performance on difficult-to-distinguish categories. This research provides new methods and ideas for relation extraction in Chinese medical texts, with significant practical application value. You can find our code in <https://github.com/Warma10032/Relation-Extraction>

Key words

Relation Extraction; BERT; Pre-trained Language Models; Chinese Medical Texts; Relation Classification

*本文为东南大学2024秋季学期自然语言处理课程大作业报告

¹⁾唐梓烨, 男, 58122310, 213221785@seu.edu.cn

1 引言

从非结构化文本中提取实体及其关系是信息提取中的一个基本问题。这个问题可以分解为两个子任务：命名实体识别(Sang 等人, 2003[1]; Ratinov 等人, 2009[2]) 和关系提取(Zelenko 等人, 2002[3]; Bunescu 等人, 2005[4])。

早期的工作采用了流水线方法, 训练一个模型来提取实体 (Florian 等人, 2004[5]), 另一个模型来分类它们之间的关系 (Zhou 等人, 2005[6]; Chan 等人, 2011[7])。近来流行的端到端评估方法, 由系统共同模拟这两项任务 (Li 等人, 2014[8]; Miwa 等人 [9], 2016; Zhang 等人, 2017[10]; Wadden 等人, 2019[11]; Wang 等人, 2020[12])。长期以来, 人们一直认为联合模型可以更好地捕获实体和关系之间的交互, 并有助于缓解错误传播问题。

最近, 预训练大语言模型的出现, 一些研究者提出了一种更加简单的方法。该方法学习建立在深度预训练语言模型上的两个编码器 (Devlin 等人, 2019[13]; Beltagy 等人, 2019[14]; Lan 等人, 2020[15])。这两个编码器分别是用于命名实体识别的实体模型和关系抽取的关系模型, 它们是独立训练的, 关系模型仅依赖实体模型来提供输入特征。实体模型建立在单词级别的表示之上, 而关系模型建立在给定特定实体对的上下文表示之上。这是一种高效的流水线方法, 但它在两个任务之间又引入了一些联系, 使用实体模型的输出帮助关系模型进行关系抽取, 是一种联合的思想。

2 相关工作

传统上, 提取文本中实体之间的关系被研究为两个独立的任务: 命名实体识别和关系提取。基于流水线的方法进行关系抽取的主要流程可以描述为: 针对已经标注好目标实体对的句子进行关系抽取, 最后把存在实体关系的三元组作为预测结果输出。一些基于流水线方法的关系抽取模型被陆续提出, 其中, 采用基于 RNN, CNN 及其改进模型的网络结构, 因其高精度获得了学术界的大量关注。

2.1 基于流水线的方法

(1) 基于 RNN 模型的实体关系抽取方法

基于 RNN 模型进行关系抽取的方法由 Socher

等人[16]于 2012 年首次提出, 该方法为分析树中的每个节点分配一个向量和一个矩阵, 其中, 向量用于捕获组成部分的固有语义, 而矩阵则用于刻画节点如何影响其相邻单词或短语的含义。此矩阵向量 RNN 模型能够学习命题逻辑和自然语言中的操作符语义, 有效解决了传统单词向量空间模型无法捕捉长短语的构成语义的问题, 从而在更深层次上推动了自然语言理解的发展。

Hashimoto 等人[17]在 2013 年提出了基于句法树的递归神经网络模型, 与 Socher 等人[16]提出的模型不同的是, Hashimoto 的模型未采用计算复杂的词依存矩阵, 而是通过引入词性 (POS) 标签、短语类别和句法头等附加特征增强语义表示。此外, 该模型还引入了平均参数机制, 为目标任务中关键短语分配更高的权重。实验证明, 增加特征和引入平均参数均显著提升了模型的性能。

进一步的研究中, Cai 等人[18]于 2016 年提出了一种基于最短依赖路径 (SDP) 的深度学习关系抽取模型, 即双向递归卷积神经网络 (BRCNN) [18]。该模型将卷积神经网络 (CNN) 与基于长短期记忆网络 (LSTM) 的双向递归神经网络结合, 以充分挖掘 SDP 中的依存关系信息。其中, 双向 LSTM 用于对最短依存路径的全局模式进行编码, 而卷积层则捕捉依存关系链接中相邻词语的局部特征。这种融合增强了模型对实体对之间关系方向的分类能力, 进一步推动了关系抽取领域的发展。

(2) 基于 CNN 模型的实体关系抽取方法

Zeng 等人[19]在 2014 年首次提出了使用 CNN 进行关系抽取, 并提出了基于卷积深度神经网络 (CDNN) [19]的方法。该模型通过提取词汇层次和句子层次的特征, 将所有单词标记作为输入, 省去了复杂的预处理步骤, 从而有效规避了传统预处理系统可能带来的特征误差传播问题, 显著提升了模型性能的鲁棒性。

在此基础上, Xu 等人[20]于 2015 年进一步提出了一种基于依存分析树的卷积神经网络关系抽取模型[20]。该模型将输入文本映射至依存分析树, 同时设计了一种创新的负采样策略: 首先, 利用依存路径学习关系的方向性; 其次, 使用负采样方法学习主体和客体的位置分布。具体而言, 模型以从客体到主体的最短依存路径作为负样本输入, 借此缓解实体对距离较远时依存分析树可能引入的无关信息对模型性能的影响。这一策略显著提升了关

系抽取任务的准确性。

进一步的研究中, Wang 等人[21]于 2016 年提出了一种新颖的基于注意力机制的 CNN 架构, 旨在捕获与指定实体相关的多层次注意力信息, 并通过池化操作聚合与特定关系相关的注意力特征。该方法使模型能够在句子异构的情况下自动识别与目标关系类别相关的句子片段。模型引入的基于注意力的池化方法能够提取出对关系分类最为显著的 N-gram 短语。此外, 混合注意力机制在聚合过程中能高效地抽取与关系分类相关的三元组字段。实验结果表明, 这种方法在提升模型性能的同时, 显著增强了对细粒度语义线索的捕获能力。

2.2 基于联合学习的方法

在过去的几年里, 由于基于流水线方法存在错误累积传播、忽视子任务间关系依赖、产生冗余实体等问题, 联合模型逐渐开始受到重视。我们将现有的联合模型分为两类: 结构化预测和多任务学习。

(1) 结构化预测

结构化预测方法将这两项任务放入一个统一的框架中, 尽管它可以通过多种方式进行表述。Li[8]等人在 2014 年提出基于动作的系统, 用于联合抽取实体和关系, 该系统在识别新实体的同时能够建立与之前实体的链接; Zhang 等人[10]在 2017 年采用了 Miwa 等人[22]在 2014 年提出的表格填充方法, 将两个任务整合到统一框架中, 通过特定的方式同时处理两个任务, 以实现联合抽取; Katiyar 等人[23]在 2017 年运用基于序列标记的方法进行联合抽取; Sun 等人[24]在 2019 年提出基于图的方法来联合预测实体和关系类型, 通过构建图结构更好地捕捉实体和关系之间的复杂依赖关系; Li 等人[25]在 2019 年将关系抽取任务转换为多轮问答问题, 以一种新的视角尝试联合抽取。这些结构化预测方法都需要在推理时解决全局优化问题, 并借助束搜索或强化学习进行联合解码, 以提高模型对实体和关系联合抽取的准确性。

(2) 多任务学习

多任务学习本质上为实体和关系构建了两个单独的模型, 并通过参数共享一起优化它们。Miwa 等人[9]在 2016 年提出为实体识别和关系抽取分别构建模型, 其中实体预测使用序列标记模型, 关系抽取使用基于树的 LSTM 模型, 并且两个模型共享一个 LSTM 层来获取上下文词表示, 通过这种参数共享方式优化两个模型的性能。Bekoulis 等人

[26]在 2018 年提出的方法类似, 也是构建两个单独模型, 但将关系分类建模为多标签头选择问题, 同样通过共享参数来提升整体性能。这些多任务学习方法在解码时仍保留一定的管道特性, 即先提取实体, 再将关系模型应用于已提取的实体上。

3 任务分析

本次大作业我选择的任务是中文医疗关系分类, 旨在从中文医疗文本中识别并分类不同类型的医疗关系, 这些关系通常涉及药物、症状、疾病、治疗方法等实体之间的联系。例如, 该任务可能需要识别并分类“药物-治疗疾病”、“症状-诊断疾病”或“疾病-引发并发症”等关系类型, 是个多分类问题。通过训练模型识别并分类这些关系, 可以帮助构建医疗知识图谱, 支持智能诊断、医学研究和健康管理等应用, 有助于提升医疗数据的利用效率和准确性。数据例子如下所示:

表 1 数据示例

句子 (sentence)	头实体 (h)	尾实体 (t)	关系 (r)
(二) 婴儿上消化道出血常见原因: 吞入母血、反流性食管炎、应激性溃疡、胃炎、出血性疾病以及 Mallory-Weiss 综合征。			
上消化 道出血	应激性 溃疡	病因	

具体来说, 每条数据包括: 句子 id、句子、头实体、尾实体、关系。

关系类别有 10 类, 分别为: 临床表现、药物治疗、同义词、病因、并发症、病理分型、实验室检查、辅助治疗、相关(导致)、影像学检查。

从数据集可以看出, 我们本次任务关注于关系抽取部分, 数据中为我们提供了句子准确的头实体和尾实体, 我们需要做的是判断这个实体对在句子中的关系。

因此我决定采用基于流水线的方法, 主要参考 SOTA 模型[27]的关系抽取模块, 将预训练大语言模型经过微调训练后用于我们的关系分类任务中。并在文本嵌入方面尝试不同的嵌入方法, 分析不同文本嵌入方法的效果。

4 方法介绍

4.1 预训练大语言模型BERT

BERT[28]是一种基于 Transformer 架构的预训练语言表示模型,旨在通过深度双向学习来理解语言。BERT 的目的是解决传统模型中词汇表示的局限性,即它们通常是基于独立语境的,无法有效捕捉词汇在不同句子中的多样含义。BERT 通过预训练任务 Masked Language Model (MLM) 和 Next Sentence Prediction (NSP) 来训练模型,使其能够学习到词汇在不同上下文中的表征。

BERT 的训练分为两个阶段:预训练 (pre-training) 和微调 (fine-tuning)。在预训练阶段,BERT 使用大量的无标注数据,在 MLM 和 NSP 两种预训练任务上进行训练。MLM 任务通过随机掩盖输入句子中的一些单词,并让模型预测这些掩盖单词的原始内容,而 NSP 任务则是预测两个句子是否是连续的。使用这两种方法可以捕捉文本中词语和句子级别的上下文表示。微调阶段则是在预训练模型的基础上,使用特定任务的标注数据对模型进行进一步训练,以使模型更好地适应特定的任务。

BERT 的主要创新点在于其双向训练策略,这使得模型能够同时考虑上下文中的左侧和右侧信息,从而更好地理解词汇的上下文含义。此外,BERT 还采用了 Transformer 架构,这使得模型能够有效地处理长距离依赖问题,并在大规模数据集上进行训练。

4.2 BERT的多种下游任务

BERT 的下游任务可以分为四大类:文本分类、命名实体识别、问答系统和文本生成。BERT 通过微调预训练模型来适应特定的下游任务。

(1) 文本分类

在文本分类任务中,BERT 的[CLS]标记的输出被用作分类的依据,因为[CLS]标记包含了整个句子的上下文信息。为了适应文本分类任务,BERT 的输入数据需要进行特定的预处理,例如将文本分割成 tokens,然后添加特殊标记如[CLS]和[SEP],并进行微调以适应分类任务。

(2) 命名实体识别

命名实体识别 (NER) 是识别文本中特定类型实体(如人名、地名、组织名等)的任务。BERT 在 NER 任务中的应用涉及到将文本输入模型中,按顺序输出预测的各种实体类别的标签。在处理

NER 任务时,BERT 的每个 token 的输出都被用来预测该 token 是否属于某个特定的实体类别,以及属于哪个类别。

(3) 问答系统

BERT 在问答系统中的应用主要体现在其能够捕捉到问题与答案之间的上下文关系。通过微调,BERT 可以学习到如何根据问题去预测答案的起始和结束位置,从而实现对问答对的处理。

(4) 文本生成

BERT 也可以用于生成文本摘要,即从一个长文本中提取出最重要的信息。这通常涉及到使用 BERT 的解码器部分或者基于 BERT 的序列到序列模型,通过训练模型学习如何从输入文本生成摘要。

我们本次任务是关系分类,因此我们可以利用 BERT 在文本分类中的优秀能力。当然关系分类与文本分类还存在一定的区别与联系,关系分类关注的是识别文本中实体之间的关系,而文本分类则是对整个文本的情感或主题进行分类。在关系分类中,BERT 的输出不仅包括[CLS]标记的输出,还包括实体对应的标记的输出,因此我们可以设计一些实体的标记如[E1], [E1]供 BERT 进行学习,BERT 对于这些标记的输出被用于关系分类。

4.3 不同领域的BERT

BERT 的不同优化版本针对特定领域的数据进行了预训练,以提高模型在这些领域的性能。

BioBERT[29]是在生物医学文献和临床笔记上预训练的英文 BERT 模型,改进点在于它能够更好地捕捉生物医学领域的专业术语和上下文。BioBERT 在医学领域的命名实体识别和关系抽取任务中表现出色。BioBERT 的开发对生物医学文本处理领域具有重要意义,它不仅保留了 BioBERT 在处理生物医学文本方面的强大能力,还大大减小了模型的规模,使得它可以在资源受限的环境中使用。

Chinese-RoBERTa-wwm-ext[30]是由哈工大讯飞联合实验室发布的针对中文的 BERT 预训练模型,其改进点在于采用了全词掩码技术,相较于字级掩码,能更有效地学习词级语义,提升中文自然语言处理的效果。这种模型特别适合中文文本的处理,因为它能够更好地理解和处理中文的语言特点,该模型在多项基准测试上获得了进一步性能提升。

PCL-MedBERT 是由鹏城实验室智能医学研究

小组提出的一种医学预训练语言模型，它在医学问题匹配和命名实体识别方面具有优异的性能。该模型在 6.5 亿字符的中文临床自然语言文本语料上基于 BERT 模型进行预训练，主要贡献在于提出了一种新的预训练算法，并验证了其在生物医学领域内的有效性，同时展示了该算法的通用性，为将来将此类方法扩展到其他领域提供了可能。PCL-MedBERT 通过领域自适应预训练，持续提高在生物医学领域的任务表现，并且在多个医疗任务上超越了 Google MedBERT[31]。

4.4 我们的方法

我们主要使用 Chinese-RoBERTa-wwm-ext[30] 作为预训练大语言模型，因为它是基于中文语料库训练的语言模型，特别适合处理中文自然语言任务；而且该模型采用全词掩码技术，对完整词语进行掩码训练，提升了对词级别信息的理解能力，尤其是在领域特定的命名实体识别和关系抽取任务中表现更佳。同时我们尝试了 5 中不同的文本编码模式，包括尝试改变文本与实体的输入顺序，尝试添加自定义实体标签[E1], [E1]等，分析在不同编码模式的微调下，模型的能力变化。

表 2 不同文本编码方式实现

文本编码方式	具体实现
--------	------

5 结果分析

5.1 不同预训练BERT的结果

表 3 不同预训练 BERT 的分数

预训练 BERT	Macro-precision	Macro-recall	Macro-F1	Micro-F1
BioBERT(entity_marked1)	46.2%	40.3%	35.4%	40.3%
Chinese-RoBERTa-wwm-ext(entity_marked2)	90.5%	89.7%	89.8%	89.7%
PCL-MedBERT(QA)	89.3%	88.2%	88.3%	88.2%

可以看出 BioBERT 在所有指标上表现显著低于其他两种模型。虽然 BioBERT 是专门针对生物医学领域预训练的模型，但由于其主要针对英文语料预训练，而本任务处理的是中文医疗文本，因此在语料适配性上存在明显不足。Chinese-RoBERTa-wwm-ext 取得了全指标最高分，同时 entity_marked2 编码模式进一步优化了实体标记方式，通过更贴合中文语境的标注形式提升了模型对关系分类任务的适配性。PCL-MedBERT 是针对中文医疗文本专门预训练的模型，在医疗问题匹配和命名实体识别方面表现出色，

Basic1	[CLS] {h} [SEP] {sentence} [SEP] {t}
Basic2	[CLS] {h} [SEP] {t} [SEP] {sentence}
QA	[CLS] {h} 和 {t} 之间的关系是什么？ [SEP] {sentence}
Entity_ marked1	[CLS]{sentence-part1}[E1]{h}/[E1]{sentence-part2} [E2]{t}/[E2]{sentence-part3}[SEP]
Entity_ marked2	[CLS]{sentence-part1}[实体 1]{h}/[实体 1]{sentence-part2}[实体 2]{t}/[实体 2]{sentence-part3}[SEP]

在实验中，我们关注到模型在区分并发症、相关（导致）这两类时的效果明显低于其他类别。因此我们进行了在这两个分类效果较差的类别上的增强训练，包括设计了一个含权重 loss 和数据增强方法，最终我们可以得到一个在不同类别上性能更加平衡的模型。

含权重的交叉熵损失：

$$\mathcal{L} = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C \omega_c \cdot y_c^{(n)} \cdot \log(p_c^{(n)})$$

其中：N 是一个 batch 中训练样本的总数。 $p_c^{(n)}$ 是第 n 个样本对类别 c 的预测概率。 $y_c^{(n)}$ 是第 n 个样本的真实标签。 ω_c 是类别 c 的权重。

其对专业术语的理解和关系抽取的能力较为全面。QA 编码模式通过显式地构建“问题-答案”形式，有助于模型捕捉实体间的关系，对于没有专门在文本分类任务上进行优化的模型，这种编码方式一般效果更好。

5.2 不同文本编码方式的结果

表 4 Chinese-RoBERTa-wwm-ext 不同文本编码方式的分数

编码方式	Macro-precision	Macro-recall	Macro-F1	Micro-F1
Basic1	88.2%	86.6%	86.8%	86.6%
Basic2	89.3%	88.2%	88.4%	88.2%
QA	90.2%	89.2%	89.2%	89.2%
Entity_marked1	89.8%	88.9%	89.0%	88.9%
Entity_marked2	90.5%	89.7%	89.8%	89.7%

可以看出，不同的编码方式对模型性能有一定影响。实体标记方式的优势在于明确标注了实体范围，减少了模型在实体定位上的不确定性。同时 Entity_marked2 进一步利用中文特有的语言结构增强了实体间关系的捕捉能力。QA 编码方式通过明确的任务提示，将关系分类转化为问答形式，显著提升了模型对任务的理解能力。Basic2 的表现优于 Basic1，表明实体顺序的调整对模型捕捉关系信息有一定影响。将尾实体提前可能更有利于模型理解实体间关系。

5.3 类别增强策略对结果的影响

表 5 Chinese-RoBERTa-wwm-ext 采用加权 Loss 对分数的提高

	Macro-precision	Macro-recall	Macro-F1	Micro-F1	类间 F1 极差
普通 Loss (Baseline)	89.2%	88.1%	88.6%	87.8%	25.1%
加权 Loss	89.1% (-0.1%)	88.5% (+0.4%)	88.8% (+0.2%)	88.4% (+0.6%)	17.9% (-7.2%)

可以看出加权 Loss 对类别权重的调整可能更关注少数类别，导致模型在部分类别上的预测准确性略有下降。模型更关注难以分类的少数类别，因此召回率略有提升。Macro-F1 提升幅度较小，表明加权 Loss 在全局平衡性上有一定提升，但影响有限。Micro-F1 提升较显著，说明模型在整体分类任务上更准确。极差从 25.1% 降至 17.9%，表明加权 Loss 有效平衡了不同类别的分类性能。

加权 Loss 增加了弱类别的权重，使模型在少数类别上的表现显著改善，同时对大多数类别的性能影响较小。极差的显著下降是分类任务均衡性提升的重要体现，尤其在类别性能不均衡的任务中具有实用价值。

表 6 Chinese-RoBERTa-wwm-ext 采用样本增强对分数的提高

	Macro-precision	Macro-recall	Macro-F1	Micro-F1	类间 F1 极差
--	-----------------	--------------	----------	----------	----------

加权 Loss (Baseline)	89.1%	88.5%	88.8%	88.4%	17.9%
加权 Loss+弱样	90.5%	89.7%	89.8%	89.7%	17.5%
本数据增强	(+1.4%)	(+1.2%)	(+1.0%)	(+1.3%)	(-0.4%)

可以看出加权 Loss 在前一阶段已经展示出良好的效果，尤其是在平衡不同类别的性能方面。但加权 loss+弱样本数据增强仍能在性能上得到一定的提升。弱样本数据增强策略显著提升了模型在 Precision、Recall、F1 等全局性能指标上的表现。

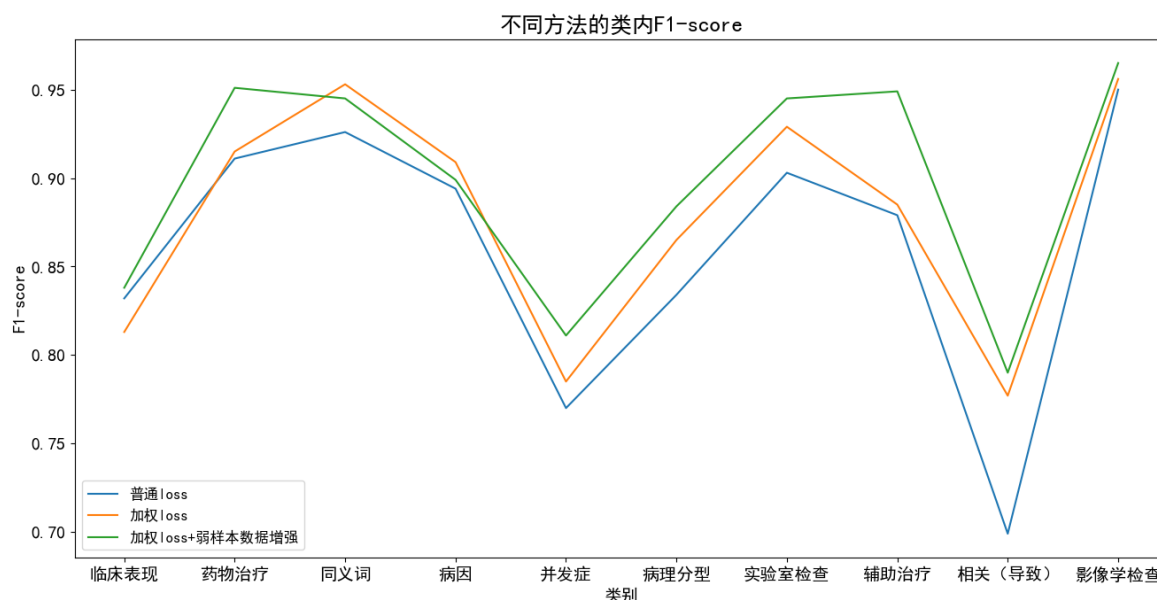


图1 采用不同方法的关系模型分类的内类 F1-score

6 总结

本文提出了一种基于预训练大语言模型 BERT 的关系抽取方法，专注于中文医疗领域的关系分类任务。通过采用 Chinese-RoBERTa-wwm-ext 作为预训练模型，并结合多种文本编码方式，如实体标记法和问答式编码，本文在微调过程中有效提升了模型的关系抽取准确性。此外，为了应对类别不均衡问题，我们提出了加权 Loss 和弱样本数据增强策略，通过调整类别权重和扩展数据集，进一步提高了模型在弱类别上的性能，且在总体精度和鲁棒性上都有显著提升。本研究不仅验证了中文领域特定优化和实体标记方法的重要性，也为中文医疗文本中的关系抽取提供了新的解决方案，具有更高的准确性和更强的泛化能力。

致谢 感谢柯文俊老师半年来在自然语言处理

课程上的教导。

参考文献

- [1] Erik Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the conll-2003 shared task: Language-independent named entity recognition. In Computational Natural Language Learning (CoNLL), pages 142–147.
- [2] Lev Ratinov and Dan Roth. 2009. Design challenges and misconceptions in named entity recognition. In Computational Natural Language Learning (CoNLL), pages 147–155.
- [3] Dmitry Zelenko, Chinatsu Aone, and Anthony Richardella. 2002. Kernel methods for relation extraction. In Empirical Methods in Natural Language Processing (EMNLP), pages 71–78.
- [4] Razvan Bunescu and Raymond Mooney. 2005. A shortest path dependency kernel for relation extraction. In Empirical Methods in Natural Language Processing (EMNLP), pages 724–731.
- [5] Radu Florian, Hany Hassan, Abraham Ittycheriah, Hongyan Jing, Nanda Kambhatla, Xiaoqiang Luo, H Nicolov, and Salim Roukos.

2004. A statistical model for multilingual entity detection and tracking. In North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), pages 1–8.
- [6] GuoDong Zhou, Jian Su, Jie Zhang, and Min Zhang. 2005. Exploring various knowledge in relation extraction. In Association for Computational Linguistics (ACL), pages 427–434.
- [7] Yee Seng Chan and Dan Roth. 2011. Exploiting syntactico-semantic structures for relation extraction. In Association for Computational Linguistics: Human Language Technologies (ACL-HLT), pages 551–560.
- [8] Qi Li and Heng Ji. 2014. Incremental joint extraction of entity mentions and relations. In Association for Computational Linguistics (ACL), pages 402–412.
- [9] Makoto Miwa and Mohit Bansal. 2016. End-to-end relation extraction using LSTMs on sequences and tree structures. In Association for Computational Linguistics (ACL), pages 1105–1116.
- [10] Meishan Zhang, Yue Zhang, and Guohong Fu. 2017a. End-to-end neural relation extraction with global optimization. In Empirical Methods in Natural Language Processing (EMNLP), pages 1730–1740.
- [11] David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. 2019. Entity, relation, and event extraction with contextualized span representations. In Empirical Methods in Natural Language Processing (EMNLP), pages 5788–5793.
- [12] Jue Wang and Wei Lu. 2020. Two are better than one: Joint entity and relation extraction with table-sequence encoders. In Empirical Methods in Natural Language Processing (EMNLP).
- [13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In North American Chapter of the Association for Computational Linguistics (NAACL), pages 4171–4186.
- [14] Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. Scibert: A pretrained language model for scientific text. In Empirical Methods in Natural Language Processing (EMNLP), pages 3606–3611.
- [15] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2020. ALBERT: A lite bert for self-supervised learning of language representations. In International Conference on Learning Representations (ICLR).
- [16] Socher R, Huval B, Manning CD, Ng AY. Semantic compositionality through recursive matrix-vector spaces. In: Proc. of the Joint Conf. on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2012.1201-1211.
- [17] Hashimoto K, Miwa M, Tsuruoka Y, Chikayama T. Simple customization of recursive neural networks for semantic relation classification. In: Proc. of the 2013 Conf. on Empirical Methods in Natural Language Processing (EMNLP). 2013.18-21.
- [18] Cai R, Zhang X, Wang H. Bidirectional recurrent convolutional neural network for relation classification. In: Proc. of the Meeting of the Association for Computational Linguistics. 2016.756-765.
- [19] Zeng D, Liu K, Lai S, Zhou G, Zhao J. Relation classification via convolutional deep neural network. In: Proc. of the 25th Int'l Conf. on Computational Linguistics: Technical Papers (COLING 2014). 2014.2335-2344.
- [20] Xu K, Feng Y, Huang S, Zhao D. Semantic relation classification via convolutional neural networks with simple negative sampling. Computer Science, 2015, 71: 941-949.
- [21] Wang L, Cao Z, Melo GD, Liu Z. Relation classification via multi-level attention CNNs. In: Proc. of the Meeting of the Association for Computational Linguistics. 2016.1298-1307.
- [22] Makoto Miwa and Yutaka Sasaki. 2014. Modeling joint entity and relation extraction with table representation. In Empirical Methods in Natural Language Processing (EMNLP), pages 1858–1869.
- [23] Arzoo Katiyar and Claire Cardie. 2017. Going out on a limb: Joint extraction of entity mentions and relations without dependency trees. In Association for Computational Linguistics (ACL), pages 917–928.
- [24] Changzhi Sun, Yeyun Gong, Yuanbin Wu, Ming Gong, Daxin Jiang, Man Lan, Shiliang Sun, and Nan Duan. 2019. Joint type inference on entities and relations via graph convolutional networks. In Association for Computational Linguistics (ACL), pages 1361–1370.
- [25] Xiaoya Li, Fan Yin, Zijun Sun, Xiayu Li, Arianna Yuan, Duo Chai, Mingxin Zhou, and Jiwei Li. 2019. Entity-relation extraction as multi-turn question answering. In Association for Computational Linguistics (ACL), pages 1340–1350.
- [26] Giannis Bekoulis, Johannes Deleu, Thomas Demeester, and Chris Develder. 2018. Adversarial training for multi-context joint entity and relation extraction. In Empirical Methods in Natural Language Processing (EMNLP), pages 2830–2836.
- [27] Zexuan Zhong and Danqi Chen. 2021. A Frustratingly Easy Approach for Entity and Relation Extraction. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 50–61.
- [28] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186.
- [29] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, Jaewoo Kang. BioBERT: a pre-trained biomedical language representation model for biomedical text mining, Bioinformatics, Volume 36, Issue 4, February 2020, Pages 1234–1240.
- [30] Cui, Yiming, Wanxiang Che, Ting Liu, Bing Qin, Ziqing Yang, Shijin Wang and Guoping Hu. Pre-Training with Whole Word Masking for Chinese BERT. ArXiv abs/1906.08101 (2019): n. pag.
- [31] C. Vasantharajan, K. Z. Tun, H. Thi-Nga, S. Jain, T. Rong and C. E. Siong, "MedBERT: A Pre-trained Language Model for Biomedical Named Entity Recognition," Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Chiang Mai, Thailand, 2022, Pages 1482-1488.

