Shanghai Jiao Tong University

CMeEE 命名实体识别大作业

Student name: 柳纪宇　周骏东　陈天翼

Course: 知识表征与推理 – Professor: 陈露

Date: *2023* 年 *5* 月 *28* 日

1 概述

在本次大作业中，我们完成了 CBLUE 数据集上的命名实体识别任务。首先，我们引入了逐层学习率衰减的策略，这种策略可以对顶层和底层应用不同的学习率。其次，我们采用了同义词替换的方法来增强模型对于文本信息的理解能力，提升模型的泛化性能。此外，我们还引入了 Bart 和 Ernie-Health 两种预训练模型，前者在文本生成领域有着比 Bert 更好的表现，后者在医疗语言领域有出色的性能。最后，我们使用全局指针模型来增强我们模型的定位能力，使其在处理命名实体识别任务时能有更准确的输出。

2 方法介绍

2.1 逐层学习率衰减

传统的模型训练对模型内的所有参数应用同一套学习率，而随着大模型的兴起和微调技术的广泛应用，这样的学习率设置方法已经无法满足一些精细的微调需求。在深度神经网络中，底层通常负责学习较为简单和通用的特征，而顶层则负责学习更为复杂和特定任务相关的特征。因此，我们希望对顶层进行更大幅度的更新，以便更快地适应特定的任务；同时对底层的更新幅度更小一些，以确保基础知识不发生太大的更改。

逐层学习率衰减通过设置顶层学习率并以某一乘法衰减率从上到下逐层衰减学习率实现上述目的，效果如图1所示（在该样例图中，乘法衰减率为 0.1）。

2.2 实体同义词替换

本实验使用中文近义词工具包 Synonyms 完成同义词替换。具体操作如下：通过一定的概率，将某个实体的内容输入到 Synonyms 提供的近义词关系网中，选择距离该实体最近的 10 个近义词。然后从中选取与原实体长度相同且距离最近的近义词，用它替换原实体，并保留替换后的实体在文本中的起始位置、终止位置以及实体类别。示意图可参考图2。实体同义词替换的目的在于帮助模型更好地学习同一实体类别的更多信息，从而提高模型在不同数据集中同类型任务的泛化能力。

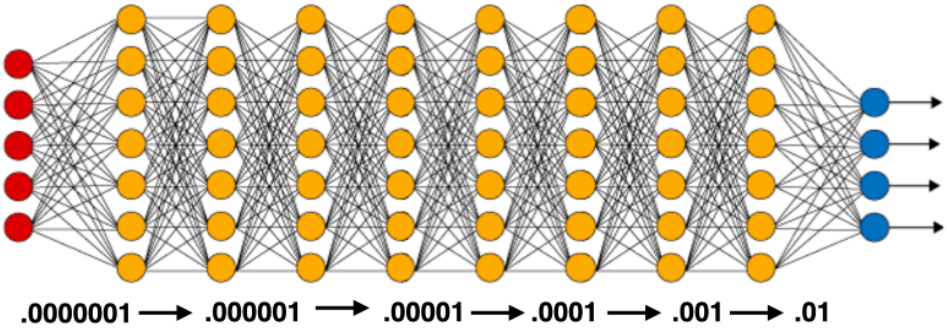

图 1: 逐层学习率衰减示意图

图 2: 数据增强：实体同义词替换

2.3 上下文同义词替换

除了直接替换文本中的实体内容之外，本实验还定义了一种对实体上下文进行近义词替换的数据增强方式。具体实现方法如下：使用 jieba 中文分词工具包对原始文本进行分词处理，然后以一定的概率对某一实体的上下文分词进行类似于前述方式的同义词替换。这种实体上下文替换的示意图可参考图3。通过进行实体上下文替换，可以增强模型在不同语境中理解实体的能力，进而提高模型的稳定性和鲁棒性。

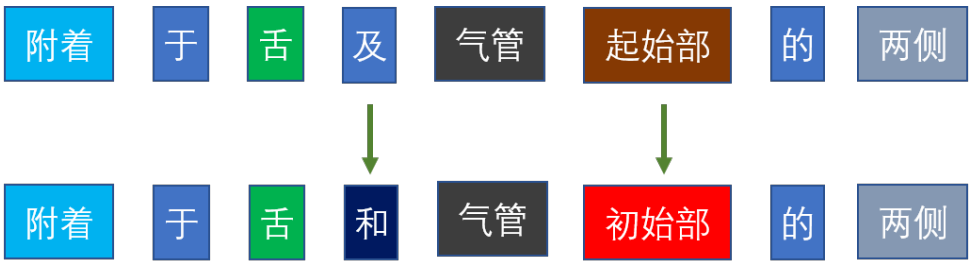

图 3: 数据增强：上下文同义词替换

2.4 Bart 预训练模型

本实验使用 Bart 预训练模型来替代 Bert。Bart 采用序列到序列的架构，将自回归的解码器与双向的编码器相结合，以提高模型准确度。在预训练阶段，Bart 引入了去噪自编码器，将输入文本随机改写，然后模型学习从改写的文本中恢复原始文本，这种方式让模型需要考虑全局的文本信息，从而可以更好地生成句子以及在高维度上提取信

息。由于预训练方法与模型架构的不同，Bart 和 Bert 的擅长领域也有所不同。Bert 在理解型任务上表现更好，而 Bart 在生成句子的合理性和连贯性上有着更大的优势。

2.5 ERNIE-Health 医学专有预训练模型

在应用于下游任务时，领域专有预训练模型适配性更好，往往可以获得超过通用大模型的效果。本实验使用 ERNIE-Health 中文医疗预训练模型替代 Bert, 以进一步提高模型对医疗实体的表征能力和理解精度。ERNIE-Health 利用医疗实体掩码策略成功学习了大量的医疗专业术语等实体知识。同时，通过医疗问答匹配任务，ERNIE-Health 学习了病患病状描述与医生专业治疗方案之间的对应关系，从而进一步增强了对于医疗术语之间逻辑的理解。经过对于 60 多万的医疗专业术语和 4000 多万医疗专业问答数据的学习，ERNIE-Health 对于医疗相关语言的建模和推理能力都有所增强，从而更适合于医疗领域命名实体识别的任务。

2.6 Global Pointer

Global Pointer 提出了对于嵌套和非嵌套实体一种统一的识别框架，通过从全局的视角预测实体起止位置，可以有效避免序列标注方法过度考虑局部边界信息等问题。具体而言，首先采用 Bert 嵌入分词结果，将每一位置嵌入向量的信息输入前馈神经网络，再两两做内积之后，可以获得这一序列的命名实体识别矩阵，从而可以处理得到对应的结果。Global Pointer 的模型结构示意图见图4。本实验中采用了 Effcient Global Pointer, 需要参数更少，推理速度更快，且识别准确率相较 Global Pointer 也有一定提升。

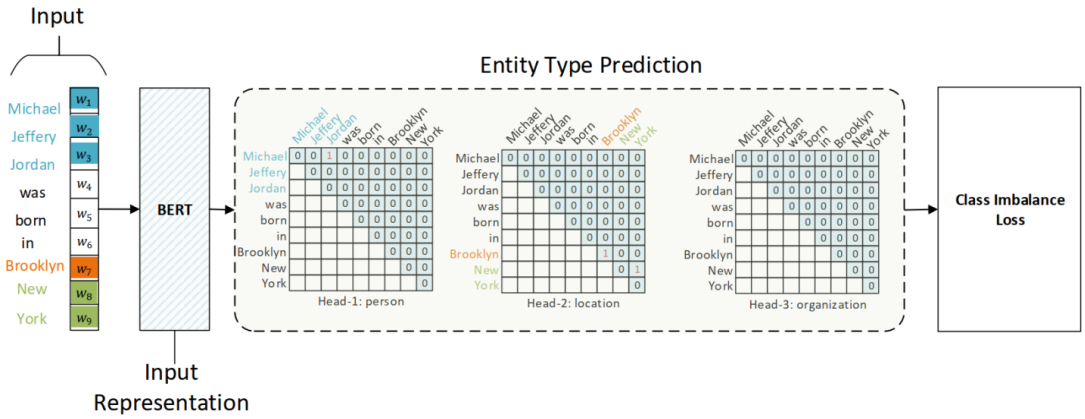

图 4: Global Pointer 模型示意图

模型	分类器	是否嵌套	数据增强	eval loss	f1-score (%)
bert	linear	否	无	0.5407	62.13
bert	crf	否	无	1059.09	62.30
bert	linear	是	无	0.2667	62.36
bert	crf	是	无	549.49	62.99
bert (layerwise)	linear	否	无	0.3658	61.05
bert	linear	否	实体替换	0.6281	62.48
bert	linear	否	上下文替换	0.6193	62.54
bert	linear	否	实体替换 + 上下文替换	0.6137	62.36
bart	linear	否	无	0.6125	62.38
bart	linear	是	无	0.6073	62.95
bart	crf	否	无	554.1	63.06
ernie	linear	否	无	0.2201	63.69
bert	global pointer	否	无	0.7057	65.95

表 1: 各种模型结构和数据处理下测试集结果汇总表

3 结果展示与讨论

3.1 逐层学习率衰减

根据实验结果，我们发现添加了逐层学习率衰减（本实验中的衰减比率为 0.9）的模型尽管在测试误差上优于其他模型，但在最终衡量标准 F1-score 上表现并不理想。我们推测这是因为 bert 大语言模型的层数较多，这使得乘法衰减率的影响在整个训练过程中逐渐叠加，这可能导致顶层的学习率过小，影响了对顶层特征的有效学习和更新，从而影响了模型的性能。此外，更小的学习率可能对模型训练轮数有更高的要求，因此在同样 20 个 epoch 的训练下，应用了逐层学习率衰减的模型的性能表现比原本的模型更差。

3.2 数据增强

本实验验证了三种不同的数据增强方式对模型性能的影响：实体替换、上下文替换、实体替换 + 上下文替换，其中实体替换的比例为 15%，上下文替换的比例为 10%（由于上下文替换会对文本内容产生更多的影响，所以替换比例略小于实体替换）。从实验结果可以看出，这三种增强方式都对模型的预测结果产生了一定程度的积极影响。尤其是上下文替换效果最好，在 F1-score 上相较于 Baseline 结果提升了约 0.4 个百分点。

然而，同时使用两种增强方式并没有达到预期的效果优势，这可能是因为增强数据的比例太大，导致模型在训练过程中过度依赖于增强数据，反而在原数据中拟合得不够

好。

3.3 Bart 预训练模型

从实验结果可以看出，将预训练模型从 bert 替换为 bart，且其它条件不变的情况下，模型的预测正确率有了一定的提升，但除嵌套 linear 分类器外提升幅度较小。虽然实验中采用的 bart 模型相比于 bert 有更大的参数量，但如同2.4中的分析，由于预训练方式不同，bart 更加擅长生成类模型，在此理解与标注的任务上表现不够突出。

3.4 ERNIEHealth 医学专有预训练模型

根据结果可以看出，采用 ERNIEHealth 模型，相比于基于 bert 和 bart 模型的各种方法效果都有较大提升，额外的专有数据相比于大规模无关数据可以更高效模型对于专有领域的表征能力，从而在下游任务中有更好的效果。

3.5 Global Pointer

Global Pointer 识别效果相比之前方法有大幅度提升，Global Pointer 通过更综合考虑序列整体信息，并且可以以一种通用的方式处理嵌套问题，从而取得更好效果。针对于 Global Pointer，我们采用了 Bert Optimizer 取代常用优化器，可以更符合 Global Pointer 下的模型的训练。而采用 Ernie 预训练模型效果反而会下降，我们推测因为分类器中参数量增多，在训练时对于模型参数调节更大，从而扰乱 Ernie 中专有数据训练的作用。

4 贡献排名

柳纪宇: %33
周骏东: %33
陈天翼: %33