

doi:10.3969/j.issn.1673-9833.2022.03.006

基于对抗训练的中文电子病历命名实体识别

孔令巍^{1,2}, 朱艳辉^{1,2}, 张旭^{1,2}, 欧阳康^{1,2}, 黄雅淋¹, 金书川^{1,2}, 沈加锐²

(1. 湖南工业大学 计算机学院, 湖南 株洲 412007;
2. 湖南省智能信息感知及处理技术重点实验室, 湖南 株洲 412007)

摘要: 为提高传统命名实体识别模型在中文电子病历上的准确性, 提出一种在基线模型 BERT-BiLSTM-CRF 中加入对抗训练的方法, 该方法在词嵌入层添加扰动因子从而生成对抗样本, 并利用对抗样本进行迭代训练, 从而优化模型参数。CCKS2021 评测数据集实验结果表明, 加入 FGM 和 PGD 两个对抗训练模型后, 其精准率、召回率以及 F_1 值相比于基线模型均有所提升。并且通过对比实验, 验证了加入对抗训练能够提高模型的预测能力和鲁棒性。

关键词: 中文电子病历; 命名实体识别; 对抗训练; BERT; BiLSTM; CRF; FGM; PGD

中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-9833(2022)03-0036-08

引文格式: 孔令巍, 朱艳辉, 张旭, 等. 基于对抗训练的中文电子病历命名实体识别 [J]. 湖南工业大学学报, 2022, 36(3): 36-43.

Named Entity Recognition of Chinese Electronic Medical Records Based on Adversarial Training

KONG Lingwei^{1,2}, ZHU Yanhui^{1,2}, ZHANG Xu^{1,2}, OUYANG Kang^{1,2},
HUANG Yalin¹, JIN Shuchuan^{1,2}, SHEN Jiarui²

(1. College of Computer Science, Hunan University of Technology, Zhuzhou Hunan 412007, China; 2. Key Laboratory of Intelligent Information Perception and Processing Technology of Hunan Province, Zhuzhou Hunan 412007, China)

Abstract: In view of an improvement of the accuracy of the traditional named entity recognition model in Chinese electronic medical records, a method has thus been proposed with adversarial training added to the baseline model BERT-BiLSTM-CRF. By adopting the proposed method, disturbance factors are added to the word embedding layer for the generation of adversarial samples, which will be used for an iterative training to optimize the model parameters. The experimental results of CCKS2021 evaluation data set show that the accuracy rate, recall rate and F_1 value are improved compared with the baseline model with FGM and PGD confrontation training models added. Based on comparative experiments, it is verified that adding confrontation training can improve the prediction ability and robustness of the model.

Keywords: Chinese electronic medical record; name entity recognition; adversarial training; BERT; BiLSTM; CRF; FGM; PGD

收稿日期: 2021-12-20

基金项目: 湖南省自然科学基金资助项目 (2020JJ6089); 湖南省教育厅科研基金资助重点项目 (19A133)

作者简介: 孔令巍 (1997-), 男, 辽宁盘锦人, 湖南工业大学硕士生, 主要研究方向为自然语言处理和知识工程,
E-mail: klingw@163.com

通信作者: 朱艳辉 (1968-), 女, 湖南湘潭人, 湖南工业大学教授, 硕士生导师, 主要研究方向为自然语言处理与知识工程,
E-mail: swayhzhu@163.com

1 研究综述

近年来, 人工智能的相关技术发展迅速, 现已在不同领域中得到了广泛应用。如在医疗领域中, 电子病历 (electronic medical record, EMR) 的普及化为疾病的诊断与治疗、病历信息的管理等提供了有效支持。电子病历是由记录病人的全部就诊档案所形成的文字、数据、医疗图像、图表等一切资料的总和, 它具有简便性、快捷性、环保性等优点。电子病历不仅能方便医务人员了解患者疾病的发生、检查、诊断、治疗等医疗活动, 还能在疾病预防等方面发挥巨大的作用^[1-2]。但是当前的电子病历文本大部分是以非结构化的形式存储, 因而如何快速从电子病历中提取所需要的信息, 是命名实体识别 (named entity recognition, NER) 技术在电子病历文本分析中的重要应用。

命名实体识别任务, 是指从非结构化的文本中识别出蕴含具体涵义的实体, 例如电子病历中的疾病和诊断、检查、药物以及手术部位等, 并将之划分到预定义的类别中。关于在命名实体识别任务中出现的问题, 研究者们也曾开展过广泛的研究。R. Panchendrarajan 等^[3]提出了一种包含双向 LSTM (bi-directional long short-term memory, BiLSTM) 和双向条件随机场 (bi-directional conditional random fields, Bi-CRF) 的神经网络, 其利用单词和字符级别的信息, 以及相邻标签之间的依赖关系进行命名实体识别, 该模型在多个数据集上被证明是有效的, 但是它存在实体边界检测不清晰的问题。王若佳等^[4]将 Bi-LSTM 模型应用到中文电子病历上, 并在词的不同标注方案下进行了实验对比, 取得了不错的实验效果, 但是其模型存在鲁棒性不高的缺点。李纲等^[5]通过结合 Word2Vec 和外部的词典资源, 对输入的词嵌入进行了更改, 并通过 Bi-LSTM-CRF 模型, 在电子病历的数据集中也取得了较高的 F_1 值, 但是其数据集存在实体类别不均衡、模型鲁棒性较差等缺点。杨文明等^[6]提出了加入独立循环神经网络 (independently recurrent neural network, IndRNN) 的 IndRNN-CRF 模型和加入膨胀卷积 (dilated convolution, DC) 的 IDCNN-BiLSTM-CRF 模型, 并通过在线医疗文本进行了命名实体识别, 发现该模型的整体性能都优于 BiLSTM-CRF 模型的。张旭等^[7]将 SoftLexicon 与 BiLSTM-CRF 相结合, 以引入外部词典资源方法对电子病历进行命名实体识别, 实验结果表明, 相较于 NER 传统方法, 所提出的方法在识别性能和效率上均显著提升。2018 年, 谷歌^[8]发布了一种新的语言

表示模型 BERT (bidirectional encoder representation from transformers), 它将自然语言任务的处理结果推上了更高的阶段。此后, Jia C. 等^[9]提出了一种半监督实体增强的 BERT 预训练方法, 此方法将词典整合到 NER 的预训练中。CCKS2020 评测中, 晏阳天等^[10]通过将 BERT 与字形字音特征相融合, 完成了对电子病历的命名实体识别。杨文明等^[11]通过将 ChiEHRBERT 与多个不同模型进行投票融合, 在医学领域的命名实体识别上取得了不错的成绩。

但在上述研究中, 词嵌入层均存在实体边界检测不清晰的问题, 即位于边界旁侧的样本比远离边界的样本更加容易出现识别错误, 从而影响模型的实体识别性能, 同时, 模型的预测能力以及鲁棒性能均不强。为了解决上述问题, 本研究提出将对抗训练融合到 BERT-BiLSTM-CRF 模型中进行命名实体识别。

对抗训练是新兴起的一门技术, 由于早期在自然语言任务上难以有效生成对抗样本, 所以多数被应用于计算机视觉领域中。近年来, 随着对抗样本相关问题的解决, 对抗训练在自然语言的各个方面都渐有成效。C. Szegedy 等^[12]首次提出对抗样本 (adversarial examples) 的概念, 旨在数据集中添加一些细微的干扰, 从而形成对抗样本。I. J. Goodfellow 等^[13]设计了一种快速生成对抗样本的方法 (fast gradient sign method, FGSM), 该方法简单可行, 并且可以利用该攻击方法产生的对抗样本再次进行对抗训练, 它系统地阐释了对抗样本的存在性、攻击性、防御方法 3 个方面, 该方法之后被广泛应用于各领域中。Zhang H. Z. 等^[14]提出 MHA (master high availability) 算法, 它基于 Metropolis-Hastings 算法的采样法来生成对抗样本。T. Miyato 等^[15]又在 FGSM 的计算扰动部分做了一些修改, 并根据具体的梯度进行标准化, 从而得到了更好的对抗样本, 但存在其实验所花费的时间会大幅度增加的缺点; A. Madry 等^[16]提出了 PGD (projected gradient descent) 模型, 该模型通过多次迭代, 以“小步走, 走多次”的策略找到最优扰动。董哲等^[17]融合了 BERT 和对抗训练, 从而在食品领域进行命名实体识别, 提高了识别实体边界的精准率。

本研究拟将对抗训练融合到 BERT-BiLSTM-CRF 模型中, 并通过对抗训练, 在词嵌入层加入扰动因子, 生成的对抗样本可以增强模型的抗干扰能力, 从而提高模型的鲁棒性和预测能力, 解决了模型中鲁棒性不强的问题。

2 基于对抗训练的中文电子病历实体识别模型

2.1 基于对抗训练的实体识别模型

本研究基于对抗训练的模型结构由 Embedding 层、BiLSTM 层和 CRF 层 3 部分组成，如图 1 所示。中文电子病历数据在进入深度学习模型之前，先将分字后的文本经预训练语言模型 BERT 转换为对应的字向量表示。以图 1 中的“胆囊多发结石”为例，其中每个字都被处理为字向量，然后将对抗训练的扰动因子与字向量相加得到对抗样本，并将对抗样本送到 BiLSTM 神经网络中。经前向传播和反向传播获取序列的特征，随后通过 CRF 层学习序列标签的约束信息，最后得到正确的序列标签。

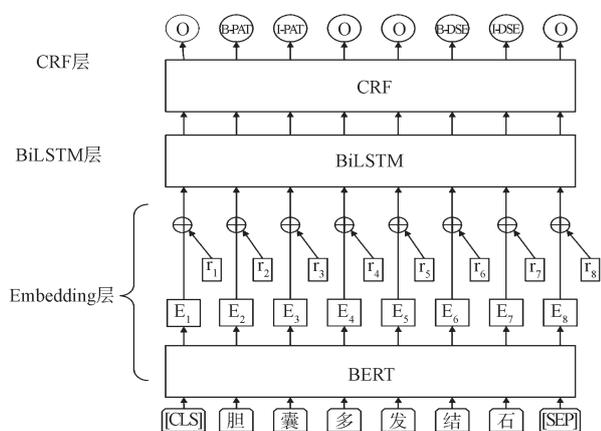


图 1 基于对抗训练的模型结构

Fig. 1 Model structure based on adversarial training

图 1 中，“[CLS] 胆囊多发结石 [SEP]”为输入的文本序列， $E_i (i=1\sim 8)$ 为输入离散的字转换为连续的字向量表示， $r_i (i=1\sim 8)$ 为字向量层的扰动。

2.2 BERT 模型

BERT 是基于深度学习的网络架构，它通过预训练，从大量文本中获取了语义和语法的基础知识，

解决了自然语言处理任务中词与词之间颗粒度不同、指代现象，以及词的理解依赖于上下文等问题。其中，BERT 模型创新性地给出了 MLM (masked language model) 和 NSP (next sentence prediction) 2 个任务，各自捕获词级别和句级别的表达，并进行联合训练。MLM 主要用于训练深度双向语言的表示向量，方法为遮住句子中的某些词汇，让解码器预测此单词的原始词汇。NSP 是指通过预训练一个二分类的语句模块来学习语句之间的关联，具体是让模型学习区分训练语句中的两个输入语句之间是否为连续片段。本研究中建立的 BERT 预训练语言模型的网络结构，如图 2 所示。

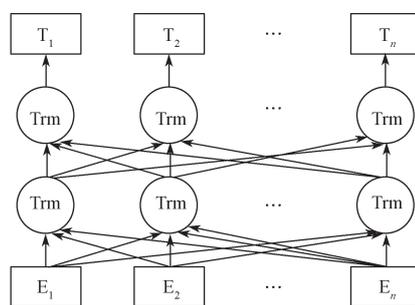


图 2 BERT 预训练语言模型的网络结构图

Fig. 2 Network structure diagram of BERT pre-training language model

BERT 预训练语言的输入是电子病历中的每一个字符，而输出则是每个字符所对应的特征向量。特征向量由字向量、句子的切分向量和位置向量相加得出。模型的输入如图 3 所示，第一个位置的符号 [CLS] 和最后一个位置的符号 [SEP] 分别代表输入序列的开始位置和结束位置。例如输入的文本是“病人患有胆结石”，经标记处理就变成“[CLS] 病人患有胆结石 [SEP]”，这两个特殊字符将在分类和划分句子中起到作用。

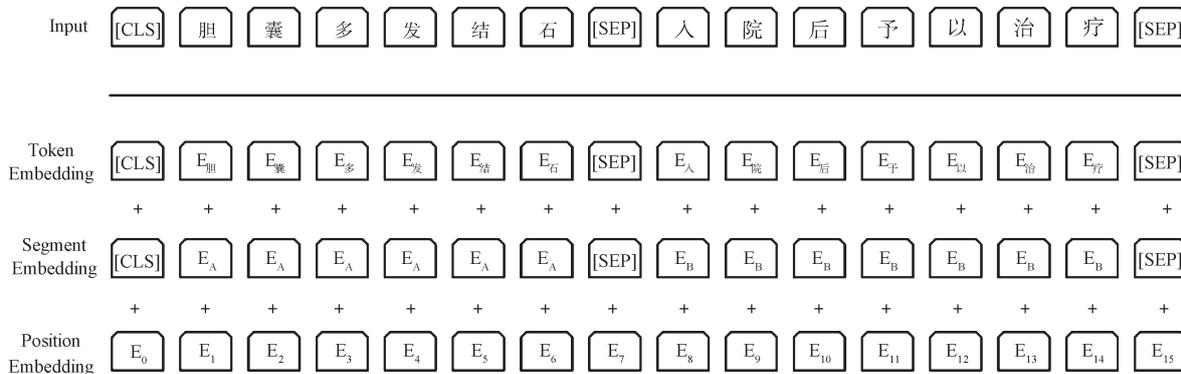


图 3 BERT 模型输入示例

Fig. 3 BERT model input samples

在 BERT 中, 字嵌入层是将每个字转化为 768 维的向量表示, 并且文本在输入到字嵌入层之前, 会进行标记化处理, 即在文本的开头和结尾处插入两个特殊的标记——[CLS] 和 [SEP], 分字后的文本通过字嵌入层转换为对应的向量表示。切分嵌入层主要用来区别两种句子, 即判断两个句子的先后顺序, 前一个句子的标记都用 A 表示, 后一个句子的标记都用 B 表示。位置嵌入层则是用来对序列中的每个标记进行编号, 用以记录每个标记的位置信息, 同时每个编号都对应一个向量。在 BERT 的一条序列语句中, 如果其长度被设置为 512, 那么位置嵌入层的向量表示为 (512, 768), 位置向量的计算公式如式 (1) 和式 (2) 所示。最后, 将这 3 个嵌入层相加, 即可以得到其特征向量。

$$PE(pos, 2i) = \sin(pos/10\ 000^{2i/d_{\text{model}}}), \quad (1)$$

$$PE(pos, 2i+1) = \cos(pos/10\ 000^{2i/d_{\text{model}}}). \quad (2)$$

式 (1) (2) 中: i 为电子病历中字的维度;

pos 为字所在位置;

d_{model} 为编码后的向量维度。

本文选用 BERT 来获取输入向量表示而非传统 Word2VEC 的原因, 在于 BERT 提高了词与词之间的联系性和表达性, 在 Word2VEC 中, 词向量的表达是静态的, 即一个词无论在何种上下文环境中, 它的向量表示都是相同的。而由于 BERT 的向量表示中包含了关于周围词的信息, 在截然不同的上下文环境中, 对这个词向量的表示方式也是截然不同的, 即是动态的。因此, BERT 为进行对抗训练提供了更加全面的词向量表达。

2.3 对抗训练

对抗训练 (adversarial training) 是一种引入噪声的规范化监督学习方法, 用于提高分类器对于样本数量小或者有损坏情况的样本鲁棒性。该方法通过在嵌入层的字向量中添加一些较小的干扰, 而不是对原始输入的样本本身加以干扰, 将获得的对抗样本再反馈给模型。对抗训练也可以认为是在加入扰动后的对抗样本下, 预测出真实标签的概率, 对抗训练的定义可简化为如下公式:

$$\max_{\theta} p(y|\mathbf{x}+\Delta\mathbf{x}, \theta), \quad (3)$$

式中: y 为真实标签;

\mathbf{x} 为原始样本;

$\Delta\mathbf{x}$ 为添加的扰动;

θ 为模型参数;

p 为增加扰动后预测真实标签的概率。

在实验中, 医疗文本“胆囊多发结石, 入院予以

治疗”, 经过 BERT 预训练语言模型生成对应的字向量, 然后根据字向量、字向量对应标签及模型参数计算出扰动值, 将扰动值与字向量相加即可得到对抗样本。其中, 常见的扰动计算方法有两种, 其一为 FGM (fast sign method) 法, 具体思路以输入序列的嵌入向量 $\mathbf{x}=[v_1, v_2, \dots, v_t]$ (式中 v 为字向量, t 为字的位置下标) 为例, 首先复制预训练阶段的词向量字典, 计算出 \mathbf{x} 的梯度, 并且根据梯度作标准化处理得到扰动值 $\Delta\mathbf{x}$, 扰动值的计算公式如式 (4) 所示; 随后将得到的扰动值与 \mathbf{x} 相加, 用新的词向量重新求出其梯度, 并累加到原梯度上, 然后根据此时的梯度对参数进行更新。

$$\Delta\mathbf{x} = \varepsilon \cdot \frac{\mathbf{g}}{\|\mathbf{g}\|_2}, \quad (4)$$

式中: ε 为一个缩放因子;

\mathbf{g} 为损失函数关于 \mathbf{x} 的偏导, 即梯度, 且

$$\mathbf{g} = \nabla_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}, \theta), \quad (5)$$

其中, $L(\mathbf{x}, \mathbf{y}, \theta)$ 为损失函数。

2.3.1 FGM 算法描述

对于数据集中的 \mathbf{x} :

1) 计算 \mathbf{x} 的前向损失、反向传播得到梯度;

2) 通过 embedding 矩阵的梯度算出 $\Delta\mathbf{x}$, 并加在当前 embedding 上, 结果相当于是 $\mathbf{x}+\Delta\mathbf{x}$;

3) 计算 $\mathbf{x}+\Delta\mathbf{x}$ 的前向损失, 反向传播得到对抗的梯度, 累加到 1) 的梯度上;

4) 将 embedding 恢复为 1) 时的值, 并根据 3) 的梯度对参数进行更新;

5) 重复以上过程, 直到模型训练全部完成。

FGM 的思路是梯度上升, 但是由于它的跨步大, 有可能无法找到约束内的最优点; 相较于 FGM 来说, PGD 进行数次迭代, 运用“小步走”的策略, 从而找到最优解。PGD 的扰动值计算公式如式 (6) 和式 (7) 所示。

$$\mathbf{x}_{t+1} = \prod_{\mathbf{x} \in \mathcal{S}} \left(\mathbf{x}_t + \alpha \frac{\mathbf{g}(\mathbf{x}_t)}{\|\mathbf{g}(\mathbf{x}_t)\|_2} \right), \quad (6)$$

$$\mathbf{g}(\mathbf{x}_t) = \nabla_{\mathbf{x}} L(\mathbf{x}_t, \mathbf{y}, \theta). \quad (7)$$

式 (6) (7) 中:

$\mathcal{S} = \{r \in \mathbb{R}^d : \|r\|_2 \leq \varepsilon\}$ 为扰动的约束空间;

α 为步长;

\mathbf{x}_t 、 \mathbf{x}_{t+1} 分别为前一次和后一次的词向量。

2.3.2 PGD 算法描述

1) 对于数据集中的 \mathbf{x} , 通过计算 \mathbf{x} 的前向损失以及反方向传播, 获得梯度并备份;

2) 对于每步 k , 通过 embedding 矩阵的梯度计算出 $\Delta \mathbf{x}$, 并且加到当前的 embedding 上, 就相当于 $\mathbf{x} + \Delta \mathbf{x}$;

3) 如果 k 不是最后一步, 则将梯度归零, 根据 1) 的 $\mathbf{x} + \Delta \mathbf{x}$ 计算前向和后向的梯度;

4) 如果 k 是最后一步, 则恢复 1) 的所有梯度, 计算最后的 $\mathbf{x} + \Delta \mathbf{x}$, 并把所有梯度累加到 1) 上;

5) 将 embedding 恢复为 1) 时的值, 并根据 4) 的梯度对参数进行更新;

6) 重复以上过程, 直到模型训练全部完成。

2.4 BiLSTM 网络

LSTM (long short-term memory) 模型是一种 RNN (recurrent neural network) 模型, 它是对 Simple RNN 的改进, 同时 LSTM 模型通过门控制单元避免了梯度爆炸。相比 RNN 来说, LSTM 对于输入中长期依赖的信息拥有更优秀的表达, 单个 LSTM 神经元及其运行机制如图 4 所示。

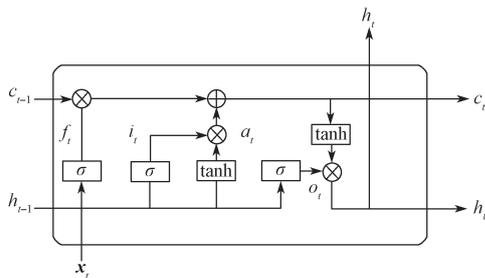


图 4 LSTM 内部结构示意图

Fig. 4 LSTM internal structure diagram

LSTM 模型的遗忘门、输入门、输出门, 以及隐藏状态的计算公式分别如下:

$$f(t) = \sigma(W_f h_{t-1} + U_f x_t + b_f), \quad (8)$$

$$i(t) = \sigma(W_i h_{t-1} + U_i x_t + b_i), \quad (9)$$

$$o(t) = \sigma(W_o h_{t-1} + U_o x_t + b_o), \quad (10)$$

$$a(t) = \tanh(W_a h_{t-1} + U_a x_t + b_a), \quad (11)$$

式 (8) ~ (12) 中: x_t 为 t 时刻的输入;

h_{t-1} 为 $t-1$ 时刻的隐藏层状态值;

W_f 、 W_i 、 W_o 、 W_a 分别为遗忘门、输入门、输出门以及在特征提取过程中 h_{t-1} 的权重系数;

U_f 、 U_i 、 U_o 、 U_a 分别为遗忘门、输入门、输出门以及在特征提取过程中 x_t 的权重系数;

b_f 、 b_i 、 b_o 和 b_a 分别为遗忘门、输入门、输出门以及在特征提取过程中的偏置值;

σ 为激活函数。

由于遗忘门和输入门计算的结果作用于 $c(t-1)$, 构成 t 时刻的细胞状态 $c(t)$ 表示为

$$c(t) = c(t-1) \odot f(t) + i(t) \odot a(t). \quad (12)$$

式中, \odot 为两元素的乘积。

最终 t 时刻的隐藏层状态 $h(t)$, 可由输出门 $o(t)$ 和当前时刻的细胞状态 $c(t)$ 求出:

$$h_t = o_t \odot \tanh(c_t). \quad (13)$$

由于 LSTM 只能保留处理过的信息, 但是在序列标注任务中, 上下文的信息同样重要, 于是提出在原本的模型结构上再加上一层反向的 LSTM, 从而组成 BiLSTM, 如此, 便可以对上下文同时进行信息处理。在本实验中, 经 Embedding 层得到的对抗样本分别以正序和逆序方式被注入到 LSTM 中, 然后将两个输出的特征向量加以拼接, 作为最后的特征向量表达式。

2.5 CRF 层

在本实验中, BiLSTM 层输出的特征向量经由 CRF 层确定最终的输出标签, 即“胆: B-PAT”、“囊: I-PAT”、“多: O”等。CRF 层相较于 BiLSTM 层, 不仅能确保输出标签之间的关系, 而且会在标签之间创造规则, 起到了约束作用。对于每一个序列的输入 \mathbf{x} , 得到了预测标签序列 \mathbf{y} , 定义预测得分函数 S 的表达式如下:

$$S(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n P_{i, y_i} + \sum_{i=0}^n A_{y_i, y_{i+1}}. \quad (14)$$

式中: P_{i, y_i} 为在第 i 个位置 (序列中第 i 个元素) 输出标签为 y_i 的概率;

$A_{y_i, y_{i+1}}$ 为前一元素标签 y_i 到当前元素标签 y_{i+1} 之间的转移概率。

此函数有效弥补了 BiLSTM 的不足, 对标签之间的关系起到约束作用, 如在一个以人名为实体的例子中, I-Person 不能存在于 B-person 前。随后对每个训练样本 \mathbf{X} , 求出代表每个可能性的标注序列 \mathbf{y} 的分数 S , 并且对每个分数进行归一化处理, 公式如式 (15) 所示:

$$P(\mathbf{y} | \mathbf{x}) = \frac{e^{S(\mathbf{x}, \mathbf{y})}}{\sum_{\mathbf{j} \in Y_{\mathbf{x}}} e^{S(\mathbf{x}, \mathbf{j})}}. \quad (15)$$

式中: \mathbf{y} 为正确的标注序列;

$Y_{\mathbf{x}}$ 为所有出现的标签序列。

然后, 利用对数似然法求出它的损失函数:

$$Loss = \log \left(\sum_{\mathbf{j} \in Y_{\mathbf{x}}} e^{S(\mathbf{x}, \mathbf{j})} \right) - S(\mathbf{x}, \mathbf{y}). \quad (16)$$

随后, 用梯度下降法进行网络学习, 更新参数, 直到训练结束。

预测时, 通过训练好的参数求出每个可能的 \mathbf{y} 序列所对应的 S 得分, 本文在此处采用维特比算法, 算

出最高概率标签序列 \mathbf{y}^* 。

$$\mathbf{y}^* = \arg \max_{\mathbf{y} \in Y_X} S(\mathbf{X}, \tilde{\mathbf{y}}) \quad (17)$$

3 实验设计与结果分析

3.1 中文电子病历数据集

为了能更加全面地评估本文中对抗训练模型的效果, 在两个数据集上对模型进行相关实验。其一为 CCKS2021 面向中文电子病历的医疗实体以及事件抽中任务一的数据集, 以下简称为数据集 1; 另一个数据集同样为医疗电子病历数据, 不同的是, 相比于数据集 1, 其中“疾病和诊断”与“解剖部位”两类数据略多于其它 4 类数据, 此数据集各类别的数据分布更加均匀, 以下简称为数据集 2。CCKS (China Conference on Knowledge Graph and Semantic Computing) 评测目的是为了建立检测知识图谱与语义计算相关技术, 以及软件系统的网络平台与信息资源, 而本次 CCKS2021 的实体识别任务是环绕中文电子病历语义化展开的系列评测的一种扩展, 它是在 CCKS2020 评测任务的基石上做出的继续与扩充。数据集 1 具体标注有实体的起始位置和终止位置, 以及预定义类别, 其类别依次为疾病和诊断、检查、检验、手术、药物、解剖部位等 6 种, 具体的类别定义参考表 1。

表 1 CCKS2021 预定义实体类型

Table 1 CCKS2021 predefined entity categories

类别	含义
疾病和诊断	医学界定的疾病, 以及医师在临床上对疾病所作出的诊断
检查	影像检测 (X 线、CT、MR、PETCT 等) + 造影 + 超声波 + 心电图
检验	在实验室进行的物理或化学检查
手术	由医生在患者身体局部实施的切除、缝合等治疗
药物	用于疾病治疗的具体化学物质
解剖部位	指疾病、症状和体征发生的人体解剖学部位

数据的标注方法为 BIO 三位标注法, 即 B-X 代表实体的开头, I-X 代表实体的结尾, O 代表不属于任何类型的非实体。数据集 1 有 1 500 条数据, 数据集 2 与数据集 1 的标注方法以及预定义类别相似, 共有 1 300 条数据。将各数据集中的数据按照 6:2:2 的比例, 划分为训练集、验证集和测试集, 具体的划分情况参见表 2。

表 2 实验数据集的划分

Table 2 Experimental data division

数据集	训练集	验证集	测试集	总数
数据集 1	900	300	300	1 500
数据集 2	780	260	260	1 300

3.2 评价指标

此次实验采用精确率 P (precision)、召回率 R (recall) 和 F_1 值为主要评价指标。精确率又称查准率, 是指实际预测正确的标签数量占全部预测正确标签的比率; 召回率又称查全率, 是指实际预测正确标签占全部正确标签的比率; F_1 值则是精确率与召回率之间的调和平均值。各指标的计算公式如 (18) ~ (20) 所示。

$$P = \frac{T_p}{T_p + F_p} \times 100\%, \quad (18)$$

$$R = \frac{T_p}{T_p + F_n} \times 100\%, \quad (19)$$

$$F_1 = \frac{2PR}{P+R} \times 100\% \quad (20)$$

式 (18) ~ (20) 中:

T_p 为序列中实际预测正确的标签;

F_p 为实体为非正确标签却被预测为正确的标签;

F_n 为实体为正确标签却被预测为非正确的标签。

3.3 实验环境及参数设置

本次实验环境如下: 操作系统为 Ubuntu 16.04 LTS, CPU i7-10750H@2.60 GHz, 内存为 8 GB, GPU NVIDIA Geforce RTX2080Ti, Python3.8, Pytorch1.6.0+cu101。

本次实验的参数设置如下: BiLSTM 隐藏层单元数为 768, Batch_size 为 8, 学习率设置为 0.000 1, Epoch 为 40。

3.4 实验方案

为了验证对抗训练在中文电子病历上命名实体识别的表现, 将 BERT-BiLSTM-CRF、BERT-FGM-BiLSTM-CRF、BERT-PGD-BiLSTM-CRF 模型分别在数据集 1 和数据集 2 上进行实验, 具体实验步骤如表 3 所示。

表 3 实验设计方案

Table 3 Experimental design scheme

实验方法	实验方案
BERT-BiLSTM-CRF	将数据集放在基线模型 BERT-BiLSTM 中进行训练, 得到预测标签
BERT-FGM-BiLSTM-CRF	将对抗训练 FGM 法得到的扰动与 Embedding 层的词向量相加, 得到新的词向量, 送入 BiLSTM
BERT-PGD-BiLSTM-CRF	将对抗训练 PGD 法得到的扰动与 Embedding 层的词向量相加, 得到新的词向量, 送入 BiLSTM

3.5 实验结果与分析

根据上述实验方案, 得到的各实验方法下的数据集识别结果如表 4 所示。

表4 实验结果对比

Table 4 Comparison of experimental results %

实验方法	数据集 1			数据集 2		
	P	R	F_1	P	R	F_1
BERT-BiLSTM-CRF	94.36	91.89	92.87	96.94	94.27	96.22
BERT-FGM-BiLSTM-CRF	95.97	92.49	93.73	97.63	95.86	96.84
BERT-PGD-BiLSTM-CRF	96.21	93.32	93.92	97.89	96.47	97.15

分析表4中的实验数据可以得知,相对于基线模型BERT-BiLSTM-CRF,基于对抗训练的BERT-FGM-BiLSTM-CRF模型和BERT-PGD-BiLSTM-CRF模型,它们在两个数据集上的实体识别效果、识别精度均有不同程度的提升,其中加入FGM的实体识别模型,其 F_1 值在两个数据集上分别提升了约0.86%和0.62%;而加入PGD方法的实体识别模型,其 F_1 值在两个数据集上分别提升了约1.05%和0.93%,由此可见,加入了PGD法的模型的识别效果要略优于加入FGM法的模型。究其原因,很可能是由于这两种对抗训练迭代攻击的次数不同,FGM只进行了一次迭代,而PGD是一种迭代攻击的方法,它进行了多次迭代,并且每次迭代都将扰动投射到规定范围内,从而造成了结果上的差异。

加入了FGM法和PGD法的模型在面对输入数据的微小变动时,依然能够保持高精度的识别效果,而且本文模型不只在特定的数据集上保持良好的识别效果,对于新数据,它依然能够保持敏感性,说明加入对抗训练的模型在面对数据变化时依然能够保持其稳定性及鲁棒性。

综上所述,加入对抗训练的模型能够提升命名实体识别在中文电子病历上的准确性以及模型的稳定性,同时对于实体标签的预测能力也相应提高。

4 结语

为了进一步提高命名实体识别在中文电子病历上的精确率,本文提出了加入FGM和PGD对抗训练方法的命名实体识别模型,该方法在中文电子病历评测任务中达到了良好的成效。但是中文电子病历的命名实体识别尚有较大的改善空间,在后续研究中可从如下方面着重加以完善:

1) 由于中文电子病历中存在大量的专有词汇,导致识别困难,可加入专有医疗词典提升实体识别对于专业医疗词汇的识别率;

2) 加入对抗训练的模型普遍具有需要花费较长时间的特点,后续将研究更有效的方法,以提升模型的识别效率。

参考文献:

- [1] 李飞,朱艳辉,王天吉,等.基于医疗类别的电子病历命名实体识别研究[J].湖南工业大学学报,2018,32(4):61-66.
LI Fei, ZHU Yanhui, WANG Tianji, et al. Research on Electronic Medical Record Named Entity Recognition Based on Medical Categories[J]. Journal of Hunan University of Technology, 2018, 32(4): 61-66.
- [2] 梁文桐,朱艳辉,詹飞,等.基于BERT的医疗电子病历命名实体识别[J].湖南工业大学学报,2020,34(4):54-62.
LIANG Wentong, ZHU Yanhui, ZHAN Fei, et al. Named Entity Recognition of Electronic Medical Records Based on BERT[J]. Journal of Hunan University of Technology, 2020, 34(4): 54-62.
- [3] PANCHENDRARAJAN R, AMARESAN A. Bidirectional LSTM-CRF for Named Entity Recognition[C/OL]// Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation. Hong Kong: [s.n.], 2018: 531-540. <https://www.aclweb.org/anthology/Y18-1061.pdf>.
- [4] 王若佳,魏思仪,王继民. BiLSTM-CRF模型在中文电子病历命名实体识别中的应用研究[J].文献与数据学报,2019,1(2):53-66.
WANG Ruojia, WEI Siyi, WANG Jimin. Applied Research on Named Entity Recognition in Chinese Electronic Medical Record Based on BiLSTM-CRF Model[J]. Journal of Library and Data, 2019, 1(2): 53-66.
- [5] 李纲,潘荣清,毛进,等.整合BiLSTM-CRF网络和词典资源的中文电子病历实体识别[J].现代情报,2020,40(4):3-12,58.
LI Gang, PAN Rongqing, MAO Jin, et al. Entity Recognition of Chinese Electronic Medical Records Based on BiLSTM-CRF Network and Dictionary Resources[J]. Journal of Modern Information, 2020, 40(4): 3-12, 58.
- [6] 杨文明,褚伟杰.在线医疗问答文本的命名实体识别[J].计算机系统应用,2019,28(2):8-14.
YANG Wenming, CHU Weijie. Named Entity Recognition of Online Medical Question Answering Text[J]. Computer Systems & Applications, 2019, 28(2): 8-14.
- [7] 张旭,朱艳辉,梁文桐,等.基于SoftLexicon的医疗实体识别模型[J].湖南工业大学学报,2021,35(5):77-84.
ZHANG Xu, ZHU Yanhui, LIANG Wentong, et al. Medical Entity Recognition Model Based on SoftLexicon[J]. Journal of Hunan University of Technology, 2021, 35(5): 77-84.

- [8] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding[EB/OL]. [2021-12-16]. <https://arxiv.org/abs/1810.04805>.
- [9] JIA C, SHI Y, YANG Q, et al. Entity Enhanced BERT Pre-Training for Chinese NER[C/OL]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). address = "Online". 2020: 6384-6396. <https://www.aclweb.org/anthology/2020.emnlp-main.518.pdf>.
- [10] 晏阳天, 赵新宇, 吴 贤. 基于 BERT 与字形字音特征的医疗命名实体识别 [C/OL]// 中国知识图谱与语义计算大会 (CCKS 2020) 评测任务论文集. 南昌: 中国信息学会语言与知识计算专业委员会, 2020. https://bj.bcebos.com/v1/conference/ccks2020/eval_paper/ccks2020_eval_paper_3_1_2.pdf.
YAN Yangtian, ZHAO Xinyu, WU Xian. Medical Biometrics Recognition Based on BERT and Glyph and Phonetic Features[C/OL] // Proceedings of the Evaluation Tasks at the China Conference on Knowledge Graph and Semantic Computing(CCKS2020). NanChang: Language and Knowledge Computing Committee of Information Society of China, 2020. https://bj.bcebos.com/v1/conference/ccks2020/eval_paper/ccks2020_eval_paper_3_1_2.pdf.
- [11] 杨文明, 毕金良, 邹佳丽, 等. 基于 ChiEHRBert 与多模型融合的医疗命名实体识别 [C/OL]// 中国知识图谱与语义计算大会 (CCKS 2020) 评测任务论文集. 南昌: 中国信息学会语言与知识计算专业委员会, 2020. https://bj.bcebos.com/v1/conference/ccks2020/eval_paper/ccks2020_eval_paper_3_1_3.pdf.
YANG Wenming, BI Jinliang, ZHOU Jiali, et al. Medical Named Entity Recognition Based on ChiEHRBert and Multi-Model Fusion[C/OL] // Proceedings of the Evaluation Tasks at the China Conference on Knowledge Graph and Semantic Computing(CCKS2020). NanChang: Language and Knowledge Computing Committee of Information Society of China, 2020. https://bj.bcebos.com/v1/conference/ccks2020/eval_paper/ccks2020_eval_paper_3_1_3.pdf.
- [12] SZEGEDY C, ZAREMBA W, SUTSKEVER I, et al. Intriguing Properties of Neural Networks[J/OL]. Computer Science, 2013. <https://arxiv.org/abs/1312.6199>.
- [13] GOODFELLOW I J, SHLENS J, SZEGEDY C. Explaining and Harnessing Adversarial Examples[J/OL]. Computer Science, 2014. <https://arxiv.org/abs/1412.6572>.
- [14] ZHANG H Z, ZHOU H, MIAO N, et al. Generating Fluent Adversarial Examples for Natural Languages[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy. Stroudsburg, PA, USA: Association for Computational Linguistics, 2019. <https://arxiv.org/abs/2007.06174>.
- [15] MIYATO T, DAI A M, GOODFELLOW I. Adversarial Training Methods for Semi-Supervised Text Classification[C]//International Conference on Learning Representations. San Juan: Computational and Biological Learning Society, 2016. <https://arxiv.org/abs/1605.07725>.
- [16] MADRY A, MAKELOV A, SCHMIDT L, et al. Towards Deep Learning Models Resistant to Adversarial Attacks[J]. International Conference on Learning Representations, 2018. <https://arxiv.org/abs/1706.06083>.
- [17] 董 哲, 邵若琦, 陈玉梁, 等. 基于 BERT 和对抗训练的食品领域命名实体识别 [J]. 计算机科学, 2021, 48(5): 247-253.
DONG Zhe, SHAO Ruoqi, CHEN Yuliang, et al. Named Entity Recognition in Food Field Based on BERT and Adversarial Training[J]. Computer Science, 2021, 48(5): 247-253.

(责任编辑: 廖友媛)