

# 基于 GRU 的电力调度领域命名实体识别方法<sup>①</sup>



吴超<sup>1,2</sup>, 王汉军<sup>2</sup>

<sup>1</sup>(中国科学院大学, 北京 100049)

<sup>2</sup>(中国科学院 沈阳计算技术研究所, 沈阳 110168)

通讯作者: 吴超, E-mail: 459963992@qq.com

**摘要:** 电力调度领域命名实体识别是电力知识图谱构建步骤中的重要一环, 目前存在基于机器学习和深度学习模型被用于通用领域或是其他专业领域的命名实体识别. 为了解决电力调度领域命名实体识别的问题, 研究 Transformer-BiGRU-CRF 模型, 该模型可以有效的解决电力调度领域中命名实体识别的问题. 通过 Transformer 模型得到语料的字向量, 再通过 BiGRU 和 CRF 进行命名实体识别. 该模型在训练过程中有两种训练方式, 第 1 种方式是只训练 BiGRU 和 CRF 部分的参数; 第 2 种方式是训练包括 Transformer 部分的整个模型的参数. 最后发现, 第 1 种方式达到模型的平稳状态需要的时间更少, 但是第 2 种达到平稳状态准确率会高出接近 5%.

**关键词:** Transformer 模型; 命名实体识别; BiGRU; 电力调度

引用格式: 吴超, 王汉军. 基于 GRU 的电力调度领域命名实体识别方法. 计算机系统应用, 2020, 29(8): 185-191. <http://www.c-s-a.org.cn/1003-3254/7595.html>

## Named Entity Recognition in Electric Power Dispatching Field Based on GRU

WU Chao<sup>1,2</sup>, WANG Han-Jun<sup>2</sup>

<sup>1</sup>(University of Chinese Academy of Sciences, Beijing 100049, China)

<sup>2</sup>(Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)

**Abstract:** Name entity recognition is an important part in the power knowledge map construction in power dispatching field. Currently, machine learning and deep learning models are used to name entity recognition in the general field or other professional fields. In order to solve the named entity recognition in the power dispatching field, the Transformer-BiGRU-CRF model is researched. The character vector of the corpus is obtained through the Transformer model, and the named entity recognition is performed through BiGRU-CRF. There are two training methods in the training process, the first method is only to train the parameters of the BiGRU-CRF part; the second method is to train the whole model parameters including the Transformer part. Finally, it is found that the first approach reaches the stationary state in less time, but the accuracy rate is about 5% higher for the second approach.

**Key words:** Transformer model; Name Entity Recognition (NER); BiGRU; power dispatch

## 1 引言

电力调度是为了使各类电力生产工作有序进行, 保证电网对内可以自身安全稳定运行、对外能够可靠供电所采取的一种有效的管理手段. 具体工作内容是根据各类信息采集设备反馈的数据信息或监控人员报

告的信息, 结合电网在实际运行过程中的电流、电压、负荷、频率等相关参数. 综合考虑工作中的各种情况, 对电网运行状态进行判断, 通过系统或是其他方式自动发布操作指令, 指挥现场操作人员或自动控制系统进行有效调整, 比如调整发电机出力、调整负荷

① 收稿时间: 2020-02-05; 修改时间: 2020-03-03, 2020-03-20; 采用时间: 2020-03-24; csa 在线出版时间: 2020-07-29

分布、投切电容器、电抗器等,使电网能够持续安全稳定的运行<sup>[1,2]</sup>。

电力领域与互联网的结合,使得电力调度向自动化方向发展的速度加快。长久以来形成的电力调度管理系统在自动化的过程中,需要构建电力领域的知识图谱。在知识图谱的构建过程中,命名实体识别(NER)<sup>[3]</sup>是知识图谱构建中的一个重要环节,该任务是从互联网、电力调度运行数据中抽取电力调度领域的命名实体。

命名实体识别是自然语言处理中的一项关键任务,在1995年的MUC会议上被第一次提出,提出的目的是识别文本中特定的名称或者是有意义的短语<sup>[4]</sup>。

早期的命名实体识别很多都是基于规则和统计机器学习的方法<sup>[5,6]</sup>。最初的时候,需要相关的专家制定命名实体过程中所需要的规则以及相关的知识库,然后采用匹配的方式进行命名实体识别。但是在实际的应用过程中,首先是规则的编写非常困难,并且制定的规则只可以用于特定的任务,并不能重用于其他的任务;并且知识库的编写也是一项很大的工程。因此该种方式的效率比较低,并不是一种命名实体识别的一项有效技术。

在机器学习发展的过程中,机器学习模型中的最大熵模型(maximum entropy)<sup>[7]</sup>、支持向量机模型(Support Vector Machine, SVM)、条件随机场模型(Conditional Random Fields, CRF)、隐马尔科夫模型(Hidden Markov Model, HMM)等一系列模型别用于命名实体识别<sup>[8-13]</sup>。机器学习模型的应用在总体上提升了命名实体识别的应用效率和识别的准确率,其中应用较多的模型是隐马尔科夫模型和条件随机场模型,这两个模型在识别效果或是在迁移学习方面取得的效果都好于之前基于规则的方式。但是机器学习模型的局限性也是很明显的,首先根据任务提取很多的特征(特征工程, feature engineering),特征的提取对最后的训练结果有着直接的影响;在训练的过程中也需要大量的时间和大量的资源;而且同规则一样还需要特定的知识库。因此此后发展出的深度学习为命名实体识别提供了更高的效率。

2011年Collobert提出基于窗口的深层神经网络模型,该模型自动学习输入句子的特征,不需要像机器学习那样需要人工提取特征,然后通过BP算法训练参数,性能上超过了前面所提到的算法。但是改模型的局限则是输入句子的长度必须是一样的,而且不能处理

文本中句子的上下文相关信息。因此在解决该问题的基础上,RNN(Recurrent Neural Network, RNN)被用于自然语言处理任务,并且在命名实体识别中也取得不错的效果。RNN基础上发展而来的LSTM神经网络良好的序列记忆能力,成了命名实体识别的主要神经网络<sup>[14-20]</sup>。在此基础上LSTM与CRF所结合形成的LSTM-CRF模型成为命名实体的基础架构之一,然后在此模型上进行扩展或以此模型为基础进行发展的众多模型在命名实体识别上被用于多个平行领域或是垂直领域。

本文所使用的Transformer-BiGRU-CRF模型是以BiLSTM-CRF模型为基础进行改进的模型。因为BiLSTM在训练中,LSTM结构单元的多个门的运算复杂性,训练所需要的时间和资源都比较大,因此希望可以使用一种既可以达到BiLSTM网络所能达到的效果,网络的结构也相对简洁,从而取代BiLSTM模型,通过相关论文与实验的证明BiGRU模型<sup>[21]</sup>可以达到相应的要求。另外,在以前的模型训练过程中,首先需要通过Word2Vec将文本中的句子转换为词向量(word embedding),然后将词向量输入模型进行训练,但是Word2Vec不能保留文本的上下文信息,而且Word2Vec是基于统计的模型,因此在本文的模型中使用Transformer模型<sup>[22]</sup>代替Word2Vec模型部分,将文本句子直接输入Transformer中,去掉分词和词向量部分。整个模型为Transformer-BiGRU-CRF模型,该模型的训练方式使用了2种训练方式。第1种方式是Transformer部分不进行训练,直接使用训练好的模型,只训练BiGRU-CRF部分;第2种方式是将整个模型一起训练,不使用Transformer所训练好的模型,最后得到整个模型参数。

## 2 Transformer-BiGRU-CRF模型

Transformer-BiGRU-CRF模型结构如图1所示,整个模型分为3个部分。首先通过Transformer模型进行语料的预训练得到输入语料的语义表示的字向量,然后将字向量序列输入BiGRU中进行语义编码,最后在CRF层通过分类得到标签所属的类别。

与传统模型相比,使用Transformer取代了Word2Vec得到词向量步骤,不再需要Word2Vec得到词向量之前要先进行分词处理,而是直接将语料输入Transformer模型就可以得到上下文相关的字向量的表示,可以表

征字的多样性, 增强了句子的语义表示; 在 BiGRU 阶段, BiGRU 训练的过程中需要的内存资源相对于 BiLSTM 会少一些. 在训练方式的选择中, 第 1 种是固定 Transformer 模型的参数, 只是训练 BiGRU + CRF 模型部分的参数; 第 2 种是将 Transformer-BiGRU-CRF 做为一个整体进行训练, 训练整个模型的参数. 第 1 种训练方式少了 Transformer 模型部分的参数训练过程, 只需要训练 BiGRU + CRF 模型的参数, 因此可以减少整个模型的训练时间.

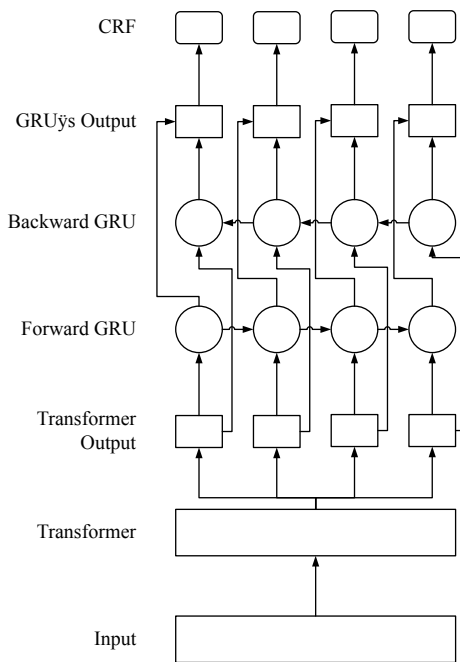


图 1 Transformer-BiGRU-CRF 模型

### 2.1 Transformer 模型结构

Transformer 是一种基于 encoder-decoder 结构的模型, 如图 2 所示.

在 encoder 中, encoder 由 6 个 layer 组成, 即图中的 Nx, 这里为 Nx6, layer 就是图 2 中的左侧部分. 每个 layer 由两个 sub-layer 组成, 它们分别是 multi-head self-attention 和 feed-forward network, 每个 sub-layer 都有 residual connection 和 normalisation, 因此 sub-layer 的输出可以表示为式 (1):

$$Sub\_layer\_output = LayerNorm(x + (SubLayer(x))) \quad (1)$$

在 Decoder 中, 有 encoder 中的 multi-head attention 和 feed-forward network 部分, 除此之外, 还比 encoder 多一个 masked multi-head attention, 最后经过 Linear 和 Softmax 输出  $i$  位置对应的的词语的概率分布.

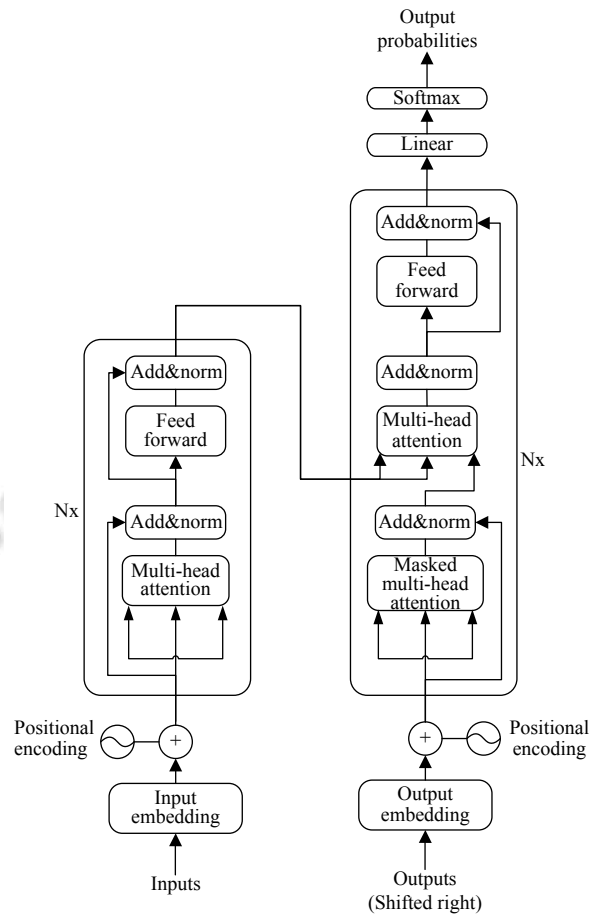


图 2 Transformer 模型结构图

Transformer 在数据预处理部分使用的是 positional encoding, 将 encoding 后的数据与 embedding 数据进行求和, 加入了文本中词语的相对位置信息, 最后学习出一份 positional embedding.

在整个 Transformer 模型中, self attention 机制起到了很大的一个作用, self attention 的表达式为式 (2):

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}} V\right) \quad (2)$$

虽然 self attention 机制没有在模型中简单直接的使用, 但是使用的 multi-head attention 是由多个 self attention 构成的升级版, multi-head attention 的结构图如图 3 所示, multi-head attention 的表达式为式 (3) 和式 (4):

$$MultiHead(Q, K, V) = concat(head_1, \dots, head_n)W^o \quad (3)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (4)$$

Multi-head attention 将同一个词向量从多个维度

进行 self attention, 从不同的角度的得到不同的结果, 最后将不同的结果合并到一个结果中得到一个张量.

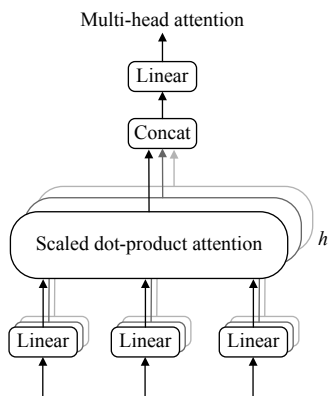


图3 Multi-head attention 结构图

使用 Transformer 代替 Word2Vec. 在 Word2Vec 中是将一个句子里面的每个单词转化为 one-hot vector, 再将 one-hot vector 输入 CBOW 模型或是 skip-gram 模型, 然后进行降维处理就得到句子中相关词的词向量. 但是 Word2Vec 得到的词向量都是长度固定的向量, 在任何语境中都被表达成同一个意思; 不能够进行多维度的解释一个词语. 而 Transformer 模型中, 在 encoder 中对其中的词向量进行 multi-head 分析之后, 有 add&norm 结果的残差处理以及 layer normalization 的归一化处理, 最后通过 position-wise feed-forward networks 部分进行特征提取和降维, 再传入到 decoder 中, 通过对 encoder 的信息和之前历史时刻经过 masked multi-head attention 处理过的 output 信息进行处理, 最后得到 i 时刻的输出, 最后得到学习的 positional embedding. 不仅达到了对句子的序列化处理, 更是得到了不同维度的信息, 解决了 Word2Vec 词向量只能表达同一个意思的问题.

### 2.2 BiGRU 模型结构

输入层的句子通过转化成字面向量, 通过双层的 GRU 神经网络模型进行标记, 如图 4 所示.

在解决自然语言处理中, 通过引入双层循环来处理序列化数据. 在 BiGRU 模型中, 其结构模型分成 3 层, 分别是输入层 (input layer)、隐藏层 (hidden layer)、输出层 (output layer), 其中在隐藏层部分将双层的细胞单元进行前后相连, 使当前节点的信息可以传递到下一个节点, 作为下一个节点输入的一部分, 用此方式达到序列数据节点的“记忆”功能.

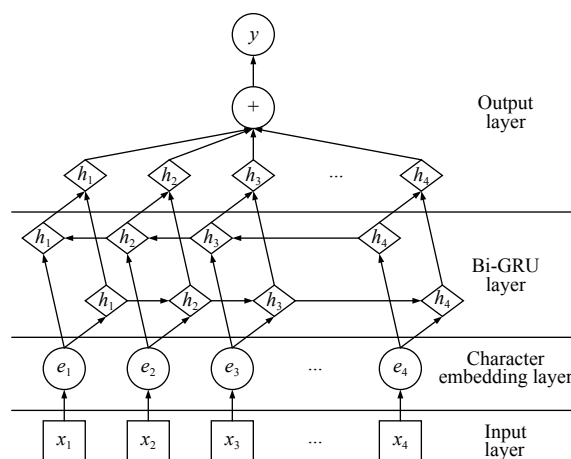


图4 BiGRU 模型

相比于 RNN 在处理任意长度的序列时可能发生的梯度消失问题以及难以学习长期依赖特征, GRU (Gated Recurrent Unit) 首先是一种特殊的 RNN 神经网络, 并且 GRU 是在 LSTM 的基础上进一步发展而来, 因此 GRU 可以同 LSTM 一样解决需要解决长依赖问题的问题.

如图 5 所示,  $x_t$  表示  $t$  时刻的输入向量,  $h_t$  表示隐藏状态, 也是细胞单元的输出向量, 这个向量包含前面  $t$  时刻所有的有效信息.  $z_t$  表示更新门, 控制信息进入下一个状态; 因为  $z_t$  经过 Sigmoid 函数激活, 所以  $z_t$  的值为 0~1 之间.  $r_t$  表示重置门, 控制哪些信息需要保留, 哪些信息会被遗弃, 这两个门共同决定隐藏状态的输出. 在 GRU 细胞中, 其中前向传播计算公式为式 (5)~式 (9)

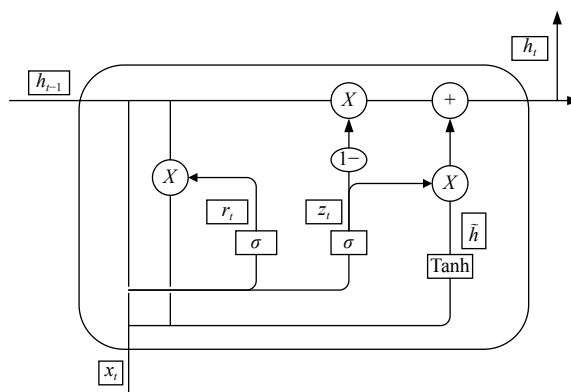


图5 GRU 细胞单元

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{5}$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{6}$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, x_t]) \quad (7)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (8)$$

$$y_t = \sigma(W_o \cdot h_t) \quad (9)$$

其中, [] 表示两个向量相连,  $\sigma$  表示 Sigmoid 函数, \* 表示矩阵的 Hadamard 积。

在模型的训练过程中, 前向传播过程中的公式中需要学习的参数有  $W_r$ 、 $W_z$ 、 $W_{\tilde{h}}$ 、 $W_o$ , 其中公式表示为式 (10)~式 (12):

$$W_r = W_{rx} + W_{rh} \quad (10)$$

$$W_z = W_{zx} + W_{zh} \quad (11)$$

$$W_{\tilde{h}} = W_{\tilde{h}x} + W_{\tilde{h}h} \quad (12)$$

在输出层部分, 设在  $t$  时刻输出层的输入为式 (13), 输出层的输出结果为式 (14), 损失函数为式 (15), 那么某个样本整体在训练过程中的损失函数为式 (16):

$$y_t^i = W_o h \quad (13)$$

$$y_t^o = \sigma(y_t^i) \quad (14)$$

$$E_t = \frac{1}{2} (y_d - y_t^o)^2 \quad (15)$$

$$E = \sum_{t=1}^T E_t \quad (16)$$

根据式 (13)~式 (16) 可以推导出式 (17)~式 (28):

$$\frac{\partial E}{\partial W_o} = \delta_{y,t} h_t \quad (17)$$

$$\frac{\partial E}{\partial W_{zx}} = \delta_{z,t} x_t \quad (18)$$

$$\frac{\partial E}{\partial W_{zh}} = \delta_{z,t} h_{t-1} \quad (19)$$

$$\frac{\partial E}{\partial W_{\tilde{h}x}} = \delta_{\tilde{h},t} x_t \quad (20)$$

$$\frac{\partial E}{\partial W_{\tilde{h}h}} = \delta_{\tilde{h},t} (r_t \cdot h_{t-1}) \quad (21)$$

$$\frac{\partial E}{\partial W_{rx}} = \delta_{r,t} x_t \quad (22)$$

$$\frac{\partial E}{\partial W_{rh}} = \delta_{r,t} h_{t-1} \quad (23)$$

其中, 在式 (17)~式 (23) 中, 有:

$$\delta_{y,t} = (y_d - y_t^o) \cdot \sigma' \quad (24)$$

$$\delta_{z,t} = \delta_{h,t} \cdot (\tilde{h}_t - h_{t-1}) \cdot \sigma' \quad (25)$$

$$\delta_{\tilde{h},t} = \delta_{h,t} \cdot z_t \cdot \Phi' \quad (26)$$

$$\delta_{r,t} = h_{t-1} \cdot (\delta_{\tilde{h},t} W_{\tilde{h}h}) \cdot \delta' \quad (27)$$

其中, 在式 (25)~式 (27) 中的参数  $\delta_{h,t}$  表示为式 (28):

$$\delta_{h,t} = \delta_{y,t} W_o + \delta_{z,t+1} W_{zh} + \delta_{t+1} W_{\tilde{h}h} \cdot r_{t+1} + \delta_{h,t+1} W_{rh} + \delta_{h,t+1} \cdot (1 - z_{t+1}) \quad (28)$$

从训练上看, 以上训练公式可以知道神经网络的结果最后趋于收敛; 从结构上比较, GRU 相比 LSTM 具有更加简单的结构, 门结构只有了更新门和重置门, 因此参数也更少, 可以加快训练的时间. 同时 GRU 具有良好的序列建模能力, 因此本文将 GRU 用于自然语言处理中的命名实体识别。

### 2.3 CRF 模型结构

CRF (Conditional Random Filed) 由 Lafferty 等于 2001 年提出, 结合了最大熵模型和隐马尔可夫模型的特点, 是一种无向图模型, 在分词、词性标注和命名实体识别等序列标注任务中取得了较好的效果. 条件随机场是一个典型的判别式模型, 其联合概率可以写成若干势函数联乘的形式, 其中最常用的是线性链条件随机场. 若让  $x=(x_1, x_2, \dots, x_n)$  表示被观察的输入数据序列,  $y=(y_1, y_2, \dots, y_n)$  表示一个状态序列, 在给定一个输入序列的情况下, 序列标注通常公式化为式 (29),  $t_k(y_{i-1}, y_i, x, i)$  是一个转移函数, 代表在标注序列中, 第  $i-1$  个和第  $i$  个的标注与整个观测序列之间的特征关系;  $s_l(y_i, x, i)$  是一个状态函数, 代表标注序列中第  $i$  个标注于此时相对应的观测序列中的值的特征;  $\lambda_j$  和  $\mu_k$  的值均是从训练数据中进行估计, 较大的负值代表其对应的特征模板可信度低, 而较大的非负值代表其队形的特征事件可信度高, 其中  $Z(x)$  代表归一化因子, 公式为:

$$p(y|x, \lambda, \mu) = \frac{1}{z(x)} \exp\left(\sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_l s_l(y_i, x, i)\right) \quad (29)$$

$$z(x) = \sum_y \exp\left(\sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_l s_l(y_i, x, i)\right) \quad (30)$$

最终的最优化输出序列计算公式为:

$$y^* = \arg \max p(y|x) \quad (31)$$

以往的研究表明, 特征选择在传统的概念抽取中起着重要的作用. NER 的性能在很大程度上取决于不

用意见的领域知识的构建和研究。

### 3 实验分析

#### 3.1 数据

本文数据集来源于某电网公司内的电力调度方面的数据。该数据集包含训练集和测试集两个部分,其中训练集包含 34.2 K 个句子, 1547.8 K 个字; 测试集包含 2.9 K 个句子, 109.3 K 个字。

#### 3.2 标注策略与评价指标

命名实体识别的标注策略有 BIO 模式、BIOE 模式、BIOES 模式。本文采用 BIOES 标注策略, 其中 B (Begin) 表示实体开始, I (Intermediate) 表示中间部分, O (Other) 表示其他与标记无关的字符, E (End) 表示结束, S (Single) 表示单个字符。

命名实体识别的评价指标有准确率 ( $P$ )、召回率 ( $R$ ) 和  $F1$  值。公式中的参数定义如下:  $T_p$  为模型识别正确的实体个数,  $F_p$  为模型识别出的不相关的实体个数,  $F_n$  为相关的实体但是模型没有识别出的个数, 其中公式表示为式 (32)~式 (34):

$$P = \frac{T_p}{T_p + F_p} \times 100\% \quad (32)$$

$$R = \frac{T_p}{T_p + F_n} \times 100\% \quad (33)$$

$$F1 = \frac{2PR}{P+R} \times 100\% \quad (34)$$

#### 3.3 实验过程

Transformer-BiGRU-CRF 模型的训练方式有 2 种。第 1 种方式是将 Transformer 模型的参数进行固定, 只训练 BiGRU 和 CRF 模型的参数; 第 2 种方式是训练 Transformer-BiGRU-CRF 整个模型中的所有参数。实验过程中会使用这两种方式进行试验验证。

在第 1 种训练方式中, Transformer 模型部分不进行训练, 只训练 BiGRU 和 CRF 部分, BiGRU 的输入层根据语料句子的最大长度设定为 20 个输入, 但在实际训练过程中根据句子长度实际变化, 多出的部分直接进初始化为 0, 中间层根据多次训练尝试的结果设定为 40 个单元, 训练的参数规模达到 900 MB 级别, 训练的轮次达到 15 000 次。在第 2 种方式中, transformer 模型部分也跟着训练, 其中共有 6 个 encoder 和 6 个 decoder 共 12 层, 12 个 multi-head attention, 共有 110 MB 的参数数量, 因此加上 transformer 部分的参数, 整个模型的

参数规模达到 1200 MB 级别, 最后的训练轮次达到 17 500 次级别。但两种训练方式都进行了 20 000 轮次的训练, 在训练的过程中记录下每一轮的训练后的准确率, 最后两种方式的准确率如图 6 所示。从图中可以发现, 在使用  $T-1$  方式训练的时候, 开始时准确率比  $T-2$  方式的准确率高, 并且可以较快的将模型训练训练完毕; 使用  $T-2$  训练方式的时候, 模型的准确率提升速度较快, 并且最后的准确率高于  $T-1$  的训练方式。

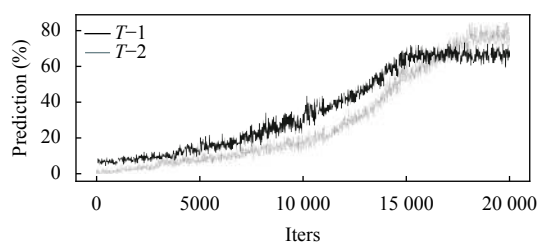


图 6  $T-1$  和  $T-2$  的准确率

#### 3.4 实验结果分析

将该模型与以下几个模型进行对比, 从表 1 中的实验结果可以发现, 基于深度学习的 LSTM 和 GRU 模型在总体效果上都好于基于机器学习的 CRF 模型。在训练的过程中发现, 基于 BiGRU 模型与基于 BiLSTM 模型在准确率上差距不大, 但是在训练过程中发现基于 BiGRU 模型训练所需要的时间比基于 BiLSTM 模型所需要的时间少。最后通过 Transformer 代替 Word2Vec 最后的准确率相比 BiGRU-CRF 模型有着较大的提升。

表 1 各个模型的对比结果 (单位: %)

模型	准确率	召回率	$F1$ 值
CRF	50.31	51.36	50.99
BiLSTM	58.62	56.11	57.34
BiLSTM-CRF	62.73	60.84	61.58
BiGRU-CRF	59.74	60.08	59.91
Transformer-BiGRU-CRF	69.71	70.23	69.97

### 4 结论与展望

本文通过将机器学习模型 CRF 和深度学习 Transformer、BiGRU 进行结合的 Transformer-BiGRU-CRF 模型用于电力调度领域的实体命名识别。为了电力调度专业领域的命名实体识别, 使用 Transformer 的字符级别向量化代替 Word2Vec 的词向量化, 可以联系文本中字符间的含义; BiGRU 相比于 BiLSTM 可以减少模型的训练时间。最后得到的结果相比于用于通

用领域模型的结果,该模型具有更好的效果。但是该模型在电力调度领域没有达到预期中与通用领域接近效果,为了提高该专业领域命名实体识别更好的效果,以后需要在更多的地方做出改进。

### 参考文献

- 1 郭小龙,马丽亚,郭建峰,等. 电力调度的运行保障问题与措施分析. 集成电路应用, 2020, 37(1): 56–57.
- 2 吴永琴. 电力调度监控一体化系统的告警信息优化研究. 通信电源技术, 2019, 36(12): 17–19.
- 3 付瑞吉. 开放域命名实体识别及其层次化类别获取 [博士学位论文]. 哈尔滨: 哈尔滨工业大学, 2014.
- 4 Kaur D, Gupta V. A survey of named entity recognition in English and other Indian languages. *IJCSI International Journal of Computer Science Issues*, 2010, 7(6): 239–245.
- 5 王宁,葛瑞芳,苑春法,等. 中文金融新闻中公司名的识别. 中文信息学报, 2002, 16(2): 1–6. [doi: 10.3969/j.issn.1003-0077.2002.02.001]
- 6 张晓艳,王挺,陈火旺. 命名实体识别研究. 计算机科学, 2005, 32(4): 44–48. [doi: 10.3969/j.issn.1002-137X.2005.04.014]
- 7 Zhou GD, Su J. Named entity recognition using an HMM-based chunk tagger. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, PA, USA. 2002.473–480.
- 8 李贯峰,张鹏. 一个基于农业本体的 Web 知识抽取模型. 江苏农业科学, 2018, 46(4): 201–205.
- 9 刘鹏博,车海燕,陈伟. 知识抽取技术综述. 计算机应用研究, 2010, 27(9): 3222–3226. [doi: 10.3969/j.issn.1001-3695.2010.09.005]
- 10 车海燕,孙吉贵,荆涛,等. 一个基于本体主题的中文知识获取方法. 计算机科学与探索, 2007, 1(2): 206–215. [doi: 10.3778/j.issn.1673-9418.2007.02.009]
- 11 车海燕,冯铁,张家晨,等. 面向中文自然语言文档的自动知识抽取方法. 计算机研究与发展, 2013, 50(4): 834–842. [doi: 10.7544/issn1000-1239.2013.20111324]
- 12 滕青青,吉久明,郑荣廷,等. 基于文献的中文命名实体识别算法适用性分析研究. 情报杂志, 2010, 29(9): 157–161, 169. [doi: 10.3969/j.issn.1002-1965.2010.09.037]
- 13 房辉,汤文兵,桂海霞,等. 基于众包标注的中文微博命名实体识别. 计算机应用与软件, 2019, 36(3): 255–259, 270. [doi: 10.3969/j.issn.1000-386x.2019.03.045]
- 14 陈佳洋,滕冲. 基于强化学习的实体关系联合抽取模型. 计算机应用, 2019, 39(7): 1918–1924.
- 15 徐飞,叶文豪,宋英华. 基于 BiLSTM-CRF 模型的食品安全事件词性自动标注研究. 情报学报, 2018, 37(12): 1204–1211. [doi: 10.3772/j.issn.1000-0135.2018.12.004]
- 16 张应成,杨洋,蒋瑞,等. 基于 BiLSTM-CRF 的商情实体识别模型. 计算机工程, 2019, 45(5): 308–314.
- 17 武惠,吕立,于碧辉. 基于迁移学习和 BiLSTM-CRF 的中文命名实体识别. 小型微型计算机系统, 2019, 40(6): 1142–1147.
- 18 Luo L, Yang ZH, Yang P, *et al.* An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition. *Bioinformatics*, 2018, 34(8): 1381–1388. [doi: 10.1093/bioinformatics/btx761]
- 19 Zhang YJ, Chen ZZ. Recurrent correlation neural network loops. *Proceedings of IJCNN International Joint Conference on Neural Networks Volume II of III*. 1992. 126–131.
- 20 Cho K, Van Merriënboer B, Gülçehre Ç, *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation. *Proceedings of EMNLP 2014*. Doha, Qatar. 2014. 1724–1734.
- 21 王洁,张瑞东,吴晨生. 基于 GRU 的命名实体识别方法. 计算机系统应用, 2018, 27(9): 18–24. [doi: 10.15888/j.cnki.csa.006552]
- 22 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. *Advances in Neural Information Processing Systems*. Long Beach, NY, USA. 2017. 5998–6008.