

基于多任务标签一致性机制的中文命名实体识别

吕书宁*, 刘健*, 徐金安†, 陈钰枫, 张玉洁

北京交通大学计算机与信息技术学院

北京市海淀区上园村3号

{19120392,jianliu,jaxu,chenyf,yjzhang}@bjtu.edu.cn

摘要

实体边界预测对中文命名实体识别至关重要。现有研究为改善边界识别效果提出的多任务学习方法仅考虑与分词任务结合，缺少多任务标签训练数据，无法学到任务的标签一致性关系。本文提出一种新的基于多任务标签一致性机制的中文命名实体识别方法：将分词和词性信息融入命名实体识别模型，使三种任务联合训练；建立基于标签一致性机制的多任务学习模式，来捕获标签一致性关系及学习多任务表示。全样本和小样本实验表明了方法的有效性。

关键词： 中文命名实体识别；多任务学习；标签一致性机制；BERT模型

Chinese Named Entity Recognition based on Multi-task Label Consistency Mechanism

Shuning Lv¹, Jian Liu¹, Jinan Xu², Yufeng Chen, Yujie Zhang

School of Computer and Information Technology, Beijing Jiaotong University

No.3 Shangyuan Village, Haidian District, Beijing

{19120392,jianliu,jaxu,chenyf,yjzhang}@bjtu.edu.cn

Abstract

Entity boundary prediction is essential for Chinese named entity recognition. The multi-task learning method proposed by the existing research to improve the effect of boundary recognition only considers the combination with word segmentation task and the lack of multi-task label training data, so the label consistency relationship of the task cannot be learned. This paper proposes a new Chinese named entity recognition method based on the multi-task label consistency mechanism. The method integrates word segmentation and part-of-speech information into the named entity recognition model to achieve joint training of three tasks and establishes a multi-task learning mode based on the label consistency mechanism to capture label consistency relationships and learn multi-task representations. The effectiveness of this method is demonstrated in both full sample and small sample experiments.

Keywords: Chinese named entity recognition , Multi-task learning , Label consistency mechanism , BERT

1 引言

©2021 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

*Equal contribution

†Corresponding author

命名实体识别(Named Entity Recognition, NER)要求从自由文本中识别出属于特定类别的片段,最早由第六届语义理解会议(Message Understanding Conference, MUC-6)提出,定义了地点、机构、人物等实体类别。自MUC-6以来,人们对命名实体识别的兴趣不断增加,各种学术研究(如CoNLL03 (Tjong Kim Sang and De Meulder, 2003)、ACE(Doddington et al., 2004)、IREX(Demartini et al., 2009)、TREC Entity Track(Balog et al., 2010)等)都致力于此。从隐马尔可夫模型(HMM) (Bikel et al., 1997)、支持向量机(SVMs)(Isozaki and Kazawa, 2002)和条件随机场(CRFs)(Lafferty et al., 2001)到近些年流行的深度学习神经网络(Lample et al., 2016; Peng and Dredze, 2016; Luo and Yang, 2016),都为命名实体识别任务在通用类领域的应用,以及各种垂直特定领域的深入推广打下了坚实的基础。

实体边界识别的准确度直接影响了命名实体识别的最终效果(Cao et al., 2018)。相较于英文命名实体识别,中文命名实体识别的边界识别效果并不理想,这导致目前中文命名实体识别任务效果不佳。为了解决上述问题,现有一些方法提出利用多任务学习将分词信息融入到命名实体识别任务中,在一定程度上提升了识别效果(Peng and Dredze, 2016; Wang et al., 2013)。然而,现有方法中仍然存在两个主要问题:

- 相比起分词信息, **词性信息**能为命名实体识别提供更具判别性的特征。例如,研究表明命名实体大多都是名词词性(Li et al., 2018)。然而,现有的方法只考虑了融入分词信息,很少考虑词性信息对命名实体识别任务的影响。
- 受标注成本、环境影响,目前同时具备命名实体、分词、词性信息三种标注资源的数据集非常有限,本文意在关注更通用的设定。现有的多任务学习方法针对不同任务的数据集分开进行学习,在这种模式下,无法学习重要的**标签一致性信息**:例如,命名实体的开始边界应该同时对应分词标签的开始边界。

针对上述问题,我们提出一种新的基于多任务学习模式的中文命名实体识别方法:第一,在融入分词信息的基础上进一步融入词性信息,实现三种任务的联合信息共享,提升中文命名实体识别效果。第二,提出标签一致性训练机制,在训练过程中考虑不同标签之间的一致性关系(实体标签的边界与分词和词性标签边界的对齐关系),指导模型更好的学习多任务表示,改善中文命名实体识别效果。在方法层面,我们的模型主要包含三个模块:(1)NER数据集预标注模块,以现有的NER数据集为基础,融入与原始数据相一致的分词、词性标签,用于指导后续多任务一致性学习。(2)多任务共享学习模块,为命名实体识别任务、中文分词(Chinese word segment, CWS)任务与词性标注(Part-of-speech tagging, POS)任务构建共享的网络结构,把多个任务的数据表示嵌入到同一个语义空间中,达到多任务联合学习的目的。(3)联合训练模块,在训练过程中考虑不同任务的解码结果是否一致,实现一致性训练。

为了验证本文方法的有效性,我们分别在新闻和社交媒体两个领域的中文命名实体识别数据集上进行实验,并且考虑了标注数据较为充足的全样本和标注数据较为缺乏的小样本两种环境的实验设置。实验结果表明:(1)全样本实验中,相较于基线模型,本文方法可以提升10.28%的F1值,与现有方法相比提升0.76%以上。(2)小样本实验中,新闻领域模拟小样本数据集取得了F1值为92.98%的实验结果,高于基线系统11.17%,高于现有方法3.05%以上,社交媒体领域真实低资源小样本数据集中F1值达到61.33%,高于基线系统8.84%,与现有方法相比提升2.54%以上。

综上所述,本文的主要贡献包括:

- 提出一种新的中文命名实体识别多任务学习框架,在命名实体识别模型中融合分词和词性信息,实现多任务联合训练,增强中文命名实体识别效果。
- 在以上基础之上,提出基于标签一致性机制的多任务学习模式,可以捕获不同任务解码结果的一致性关系,优化学习多任务表示。
- 在新闻和社交媒体两个领域中验证了本文方法的有效性。同时,实验结果表明,在标注资源匮乏的小样本数据集中本文方法效果更为显著,改善了中文命名实体识别任务可利用已标注数据集资源稀缺、样本量不充足等问题。

2 相关工作

2.1 命名实体识别

现有的命名实体识别方法大致可分为基于统计特征的方法和基于深度学习的方法。在基于统计特征的命名实体识别方法中, [Bikel et al. \(1997\)](#)提出了第一个名为IdentiFinder的基于隐式马尔科夫模型(HMM)的命名实体识别系统, 用于识别和分类名称、日期、时间表达式和数字数量。[Szarvas et al. \(2006\)](#)利用C4.5决策树和AdaBoostM1学习算法开发了多语言命名实体识别系统, 通过不同的特征子集训练几个独立的决策树分类器, 再通过多数投票方案组合它们的决策。[Bikel et al. \(1999\)](#)运用最大熵理论提出了“最大熵命名实体”(MENE), MENE能够利用非常多样化的知识来源来做标记决策, 取得了很好的效果。近年来, 基于深度学习的命名实体识别模型占据了主导地位并取得了先进的成果。与基于特征的方法相比, 深度学习方法有助于自动学习更多隐藏的特征信息。深度学习方法多采用字符级别的输入, [Ma and Hovy \(2016\)](#)利用卷积神经网络(CNN)提取单词的字符级表示, 然后将字符表示向量与词向量融合送入循环神经网络(RNN)上下文编码器提取实体。[Kuru et al. \(2016\)](#)提出的CharNER, 是一种与语言无关的字符级表示, 将句子视为字符序列, 并利用长短期记忆网络(LSTM)进行表示的提取。目前, 基于深度学习的命名实体识别已成为主流方法。

2.2 基于多任务学习的命名实体识别

现有研究已经开始探索命名实体识别任务与其他任务的关联, 并提出基于多任务学习的方法。[Collobert et al. \(2011\)](#)训练了一个window/sentence方法网络来共同执行NER、POS、组块分析(Chunk)和语义角色标注(SRL)任务。这种多任务机制让训练算法发现并学习对所有任务都感兴趣的有用的内部表示。[Rei \(2017\)](#)发现, 通过在训练过程中加入无监督语言建模目标, 可实现序列标注模型性能的改进。[Lin et al. \(2018\)](#)提出了一种基于低资源的多语种多任务架构, 可有效转移不同类型的知识改善主模型。除了考虑命名实体识别连同其他序列标注任务, 多任务学习框架还可以应用到联合提取实体和关系, 或将命名实体识别模型分为实体切分和实体类别预测, 进一步提升效果([Crichton et al., 2017](#); [Wang et al., 2018](#))。

2.3 中文命名实体识别

现有的中文命名实体识别方法包括基于字的命名实体识别、基于词的命名实体识别、基于字词联合的命名实体识别。[Li et al. \(2014\)](#)通过字级别和词级别统计方法的对比, 表明基于字符的命名实体识别方法一般有更好的表现。[Huang et al. \(2015\)](#)利用双向长短期记忆网络(BiLSTM)提取特征, 将拼写、上下文、词嵌入和地名词典四种类型的特征融于命名实体识别任务。部分研究者([Lu et al., 2016](#); [Dong et al., 2016](#))在基于神经网络的命名实体识别模型中采用基于字的命名实体识别方案。[Zhang and Yang \(2018\)](#)提出的Lattice LSTM网络结构效果较好, 其将传统的LSTM单元改进为网格LSTM, 在字模型的基础上显性利用词与词序信息, 避免了分词错误传递的问题。[Zhou et al. \(2017\)](#)将中文命名实体识别视为一项联合识别和分类任务。[Wang et al. \(2018\)](#)提出了一种适用于中文命名实体识别的门控卷积神经网络(GCNN)模型。在多任务学习方面, [Wang et al. \(2013\)](#)通过将分词和命名实体识别联合学习来融合分词信息。[Peng and Dredze \(2015\)](#)与[He and Sun \(2017\)](#)在字级别的命名实体识别方案中又融入了词的信息, 将分词信息作为soft feature来增强识别效果。[Peng and Dredze \(2016\)](#)提出了基于分词任务和中文命名实体识别任务联合训练的模型。

然而, 现有的中文命名实体识别方法一般只考虑了单纯将中文命名实体识别与中文分词两种任务相结合, 并且没有在学习过程中考虑标签的一致性信息, 这可能只会得到次优的结果。针对以上问题, 本文提出一种新的针对中文命名实体识别的多任务学习框架, 融入分词及词性等词汇信息, 同时考虑多任务解码一致性的信息, 实现更为有效的多任务学习过程。

3 方法

图1展示了本文方法的框架, 主要包含以下三个部分, 下面将对每一模块进行具体介绍。

- **NER数据集预标注模块** 利用现有的NER数据集进行预标注处理, 通过数据预标注融入额外的与原始数据具有一致性的分词、词性标签, 在数据层面即实现一定程度的信息共享, 用于指导后续多任务一致性学习。

- **多任务共享学习模块** 基于多任务学习的命名实体识别方法的共享学习模块，由共享表示层及序列解码层构成。多种任务在本模块共享参数与网络结构学习特征，实现多任务学习训练的信息交互共享。
- **联合训练模块** 在基于多任务学习的联合训练中，融入基于标签一致性机制的训练方法，增强基于多任务学习的命名实体识别方法的识别效果。

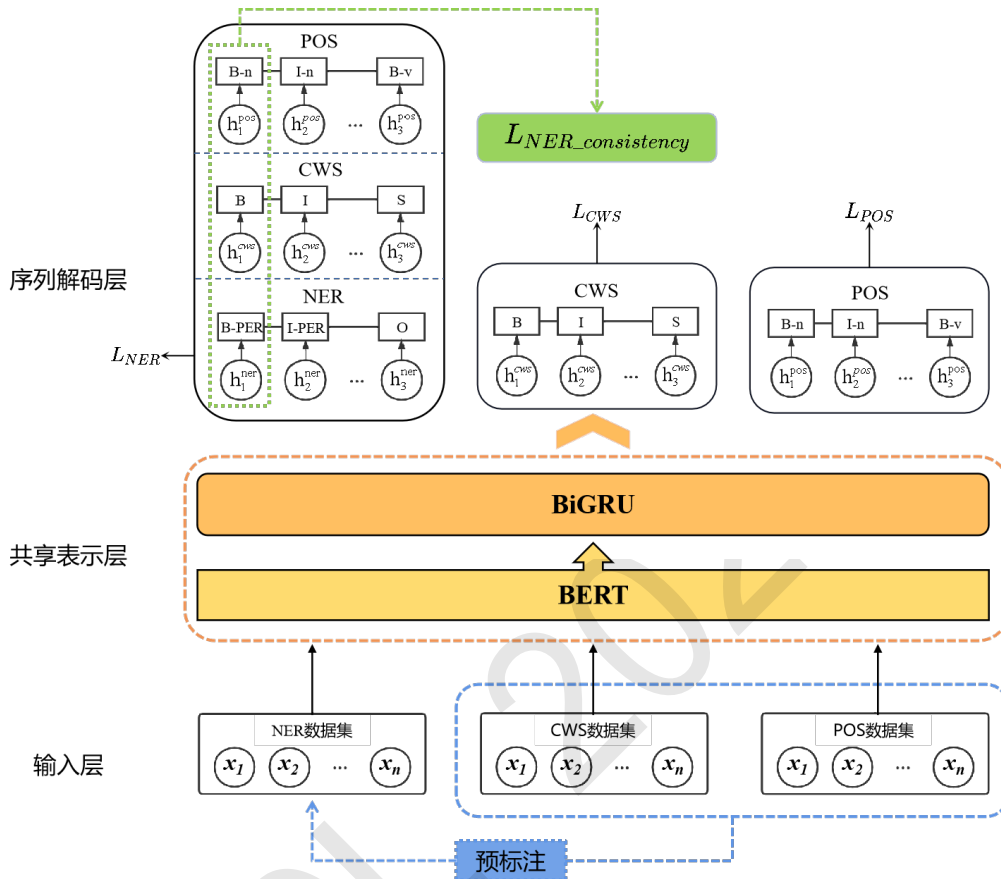


Figure 1: 基于标签一致性机制的多任务学习方法

3.1 NER数据集预标注模块

现有的多任务命名实体识别方法针对不同任务的数据集分别进行学习，不利于学习标签一致性信息。本文首先设置NER数据集预标注模块，在原有的NER数据中融入额外的与命名实体标签相一致的分词、词性标签，便于模型学习标签一致性信息。具体的流程如图2所示。

将命名实体数据集、分词数据集、词性标注数据集分别记为 $N_{NER} = \{n_i^{ner}\}$ 、 $N_{CWS} = \{n_i^{cws}\}$ 、 $N_{POS} = \{n_i^{pos}\}$ ，其中 $n^{ner} = \{V_{NER}, Z_{NER}\}$ 为NER数据集的一条样本， $n^{cws} = \{V_{CWS}, Z_{CWS}\}$ 为CWS数据集的一条样本， $n^{pos} = \{V_{POS}, Z_{POS}\}$ 为POS数据集的一条样本。形式上， $V = \{v_1, v_2, \dots, v_n\}$ ， v_i 为第 i 个字符， $Z = \{z_1, z_2, \dots, z_n\}$ 表示每种任务的标签。我们首先使用 N_{CWS} 、 N_{POS} 训练对应的分词器[‡]与词性标注器[‡]。之后，我们将 N_{NER} 中的样本去掉标签，处理成纯文本格式，使用分词器和词性标注器对其进行基于词粒度的分词与词性标注操作，再基于字符粒度进行四标签的分词标注，提取新获取的分词、词性标签。最后，整合标签并对齐到原始NER数据集上，得到同时对应三种任务标签的多任务学习特殊数据集。在预标注过程中，如遇到歧义性，以命名实体识别的标签为准进行对齐。例如，人名实体 <王小明> 经过分词器处理后可能会得到 <王\小\明> 的切分与 <B B B> 的分词标签，与原始的NER标签 <B-PER I-PER I-PER> 在边界上并不对应。在这种情况下，我们将

[‡]基于BiLSTM(Huang et al., 2015)

分词标签修正为<B I E>,即以实体边界为准进行标注。经过以上处理,我们得到预标注数据集 $D = \{d_i\}$,其中的一条样本 $d = \{S, X_{NER}, X_{CWS}, X_{POS}\}$, $S = \{s_1, s_2, \dots, s_n\}$, s_i 为第 i 个字符。 $X_{NER} = \{x_1^{ner}, x_2^{ner}, \dots, x_n^{ner}\}$, $X_{CWS} = \{x_1^{cws}, x_2^{cws}, \dots, x_n^{cws}\}$, $X_{POS} = \{x_1^{pos}, x_2^{pos}, \dots, x_n^{pos}\}$, x_i^{ner} 、 x_i^{cws} 、 x_i^{pos} 分别是 s_i 的命名实体识别标签、分词标签、词性标注标签。

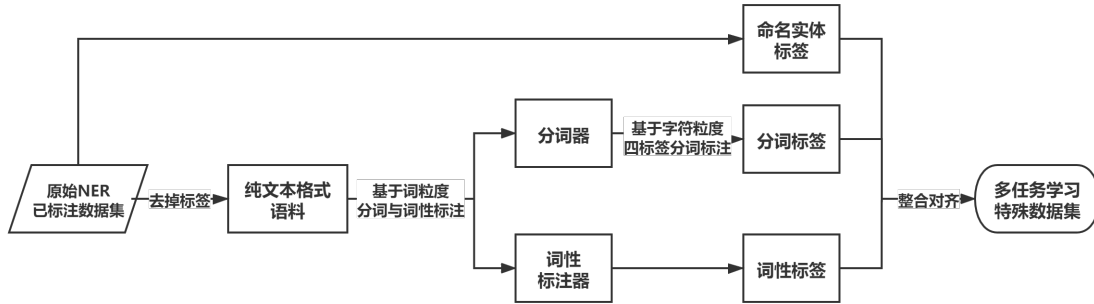


Figure 2: NER数据集预标注流程图

3.2 多任务共享学习模块

在以上基础上,我们构造多任务共享学习模块,针对不同任务联合进行学习。多任务共享学习模块包括共享表示层与序列解码层两个部分。在以下叙述中,将输入序列记为 $A = \{a_1, a_2, \dots, a_n\}$,其中 a_n 表示序列 A 的第 n 个字。 A 表示一条预标注NER数据集 D 的输入,或者原始分词数据集 N_{CWS} 或原始词性标注数据集 N_{POS} 的输入。

3.2.1 共享表示层

BERT层 传统的静态词向量通常无法表征一词多义,本文选用基于BERT(Devlin et al., 2019)的预训练语言模型用于特征学习,以表示上下文相关语义。对于输入序列 $A = \{a_1, a_2, \dots, a_n\}$,将序列统一处理为开始token为特殊分类嵌入符“[CLS]”,结束为特殊分隔符“[SEP]”的形式进行编码。BERT采用多头自注意力(Self-Attention)机制来学习每个字与其他字的依赖关系和上下文语义,然后通过前馈神经网络对Attention计算后的输入进行多个不同线性变换对的投影变换,最终得到序列的全局信息,如式1、2、3所示。

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_n)W^O \quad (2)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (3)$$

经过BERT层编码计算,我们将 A 转换为 $M = \{m_1, m_2, \dots, m_n\}$ 。

BiGRU层 在BERT层基础上,我们增加双向门控循环单元(BiGRU)层(Rana, 2016),进一步学习多任务之间的信息共享。我们将BERT层的输出序列 M 作为BiGRU层的输入,对于时刻 t 的输入,计算式4、5:

$$z_t = \sigma(W_z \cdot [h_{t-1}, m_t]), r_t = \sigma(W_r \cdot [h_{t-1}, m_t]) \quad (4)$$

$$\tilde{h}_t = \tanh(Wm_t + U(h_{t-1} * r_t)), h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (5)$$

z_t 表示更新门, r_t 是重置门, h_t 是当前节点的隐藏层状态, σ 是sigmoid函数, W 和 U 都是用于训练的权重矩阵,*表示按元素相乘, z_t 和 r_t 共同决定隐藏状态的输出。记双向GRU前向、后向两个隐层输出分别为 $H_{front} = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_n\}$ 、 $H_{back} = \{\overleftarrow{h}_1, \overleftarrow{h}_2, \dots, \overleftarrow{h}_n\}$,记最终输出向量为 $H = \{\vec{h}_i, \overleftarrow{h}_i\}$ 。

3.2.2 序列解码层

在共享表示层之上，我们构建了不同的序列解码层，针对不同任务分别进行解码。为了建模标签之间的相邻关系，选用条件随机场(CRF)(Lafferty et al., 2001)作为解码模型。给定共享表示特征 $H = \{\vec{h}_i, \overleftarrow{h}_i\}$ ，CRF定义评估分数计算公式，如式6所示。

$$s(H, y) = \sum_{i, y_i} E_{i, y_i} + \sum_i T_{y_i, y_{i+1}} \quad (6)$$

其中 y 是预测的标签序列， T 是模型参数。针对不同任务，我们分别以式7、8、9为优化目标：

$$P_{ner}(y_{ner}^{gold} | H) = e^{s(H, y_{ner}^{gold})} / \sum_{y \in Y_s} e^{s(H, y_{ner})} \quad (7)$$

$$P_{cws}(y_{cws}^{gold} | H) = e^{s(H, y_{cws}^{gold})} / \sum_{y \in Y_s} e^{s(H, y_{cws})} \quad (8)$$

$$P_{pos}(y_{pos}^{gold} | H) = e^{s(H, y_{pos}^{gold})} / \sum_{y \in Y_s} e^{s(H, y_{pos})} \quad (9)$$

使用 $p_{ner}(y_i | x_i)$ ， $p_{cws}(y_i | x_i)$ ， $p_{pos}(y_i | x_i)$ 表示不同解码模块将 x_i 标注为 y_i 的概率。

3.3 联合训练模块

在多任务学习基础上，本文提出基于标签一致性的模型训练机制，考虑不同任务的解码标签是否一致。具体而言，我们定义了损失函数如式10所示。其中， L_{NER} 、 L_{CWS} 、 L_{POS} 分别为命名实体识别任务、分词任务和词性标注任务的损失函数，分别以各自任务的标准(ground-truth)标签计算损失。

$$L = w_1 * L_{NER} + w_2 * L_{CWS} + w_3 * L_{POS} + w_4 * L_{NER_consistency} \quad (10)$$

$$L_{NER_consistency} = - \sum_{i, \hat{y}_i \neq O} \log(p_{cws}(\hat{y}_i | x_i) + \log p_{pos}(\hat{y}_i | x_i)) \quad (11)$$

$L_{NER_consistency}$ 代表跨任务标签一致性损失，计算方法为式11。其中， \hat{y}_i 代表词 x_i 的原始NER标签针对不同任务的跨任务转换结果。例如，原始NER标签 $Y = \langle O \ B \text{-} PER \ I \text{-} PER \ O \ O \rangle$ 针对分词任务的转化结果为 $\hat{Y} = \langle O \ B \ E \ O \ O \rangle$ 。我们以 \hat{Y} 作为期望输出序列，在分词解码器中计算交叉熵损失(考虑到边界信息更为重要，我们忽略了标签O的损失，只计算标签B与E的损失)。通过优化该项损失，模型可以学习到多任务解码标签的一致性信息，即实体标签的边界与分词和词性标签边界的对齐关系，从而实现多任务信息的有效融合。

为了学习不同损失的权重 w_1 、 w_2 、 w_3 、 w_4 ，本文选用动态加权 $L = \sum_i w_i(t) * L_i$ ，即根据学习速率动态调整参数(Ma and Hovy, 2016)，计算方法为：

$$w_i(t) = \frac{N \exp(r_i(t-1)/T)}{\sum_n \exp(r_n(t-1)/T)}, r_n(t-1) = \frac{L_n(t-1)}{L_n(t-2)} \quad (12)$$

其中， $w_i(t)$ 代表了每个任务 i 的权重， $L_n(t-1)$ 、 $r_n(t-1)$ 分别代表了任务 n 在第 $t-1$ 步时的Loss和训练速度， r 越小表示任务训练的越快，其中 $r_n(t-1) \in (0, +\infty)$ ， N 代表任务的数量， T 是一个常数， $T=1$ 时， w 等同于softmax的结果， T 足够大时， w 趋近1，则各个任务的Loss权重相同。在训练过程的每次迭代中，从给定的任务数据集中抽取一批训练实例，同时训练NER、CWS、POS三个任务来更新参数。

4 实验设置

4.1 数据集和评价指标

为了评估本文方法的有效性，我们在两个领域的命名实体识别数据集上进行了实验，包括新闻领域人民日报标注数据集(北京大学计算语言学研究所, 1998)和社交媒体领域微博标注数据

数据集	训练集				测试集			
	语句数	PER	LOC	ORG	语句数	PER	LOC	ORG
NER(人民日报)	20864	8144	16571	9277	2319	884	1951	984
NER(微博)	1229	938	267	165	270	225	32	44

Table 1: 命名实体识别数据集

数据集	训练集			测试集		
	字数	词数	语句数	字数	词数	语句数
CWS(北大)	1700605	991011	70495	202211	117825	8370
POS(文库)	1651203	1025684	29590	494847	323369	10000

Table 2: 分词及词性标注数据集

集(Peng and Dredze, 2015), 均将其整合成由人名、地名、组织机构名三种实体类型标注的形式。分词任务使用的是北大计算语言所制作的1998年《人民日报》标注语料库, 词性标注使用的是网络文库语料。下面给出各个数据集的详细信息:

本文选用的是BIO标注体系, 在测试过程中, 只有当一个实体的边界和实体的类型完全正确时, 才能判断该实体预测正确。本文采用召回率R、精确率P和F1值来评判模型的性能。各评价指标的计算方法如下: $P = \text{正确识别出的命名实体个数} / \text{识别出的命名实体个数} \times 100\%$; $R = \text{正确识别出的命名实体个数} / \text{标准结果中命名实体个数} \times 100\%$; $F1 = 2 \times P \times R / (P + R) \times 100\%$ 。

4.2 模型实现

本文采用BERT-Base模型, BERT-Base共12层, 隐层为768维, 采用12头模式, 共110M个参数。将BiGRU的隐藏层设为100。在模型训练方面, 采用Adam优化器(Kingma and Ba, 2014)对损失函数进行优化, 将Droupout值设为0.1(Hinton et al., 2012)。我们为不同任务设定了不同的学习率, 如表3所示。

学习率	BERT层	线性层/BiGRU层/CRF层
NER	1e-5	0.00001
CWS	2e-5	0.0001
POS	5e-5	0.0005

Table 3: learning_rate参数设置

5 实验结果

为验证本文方法的有效性, 分别在新闻和社交媒体两个领域的中文命名实体识别数据集上进行实验, 并设置了标注数据较为充足的全样本和标注数据较为缺乏的小样本两种实验环境。

5.1 全样本实验

在新闻领域的全样本实验中, 与以下方法进行对比: 单任务学习方法: 1)GRU-CRF, 该方法联合GRU和CRF进行命名实体识别。2)BiGRU, 该方法使用双向GRU网络, 捕获过去和将来两个方向的信息进行命名实体识别。3)BiGRU-CRF, 该方法结合BiGRU和CRF进行命名实体识别。4)BERT-CRF, 该方法结合预训练语言模型BERT与条件随机场CRF进行命名实体识别。5)BERT-BiGRU-CRF, 该方法结合预训练语言模型, BiGRU和CRF进行命名实体识别。6)Collobert et al. (2011), 该方法采用前馈神经网络, 结合预处理和词缀特征进行命名实体识别。7)Lample et al. (2016), 该方法创新性地BiLSTM与CNN模型结合进行命名实体识别。

别。8)Shen et al. (2017), 该方法将深度学习与主动学习相结合进行命名实体识别。多任务学习方法: 1)Wang et al. (2013), 该方法在联合训练中融合分词信息进行命名实体识别。2)Peng and Dredze (2016), 该方法基于分词和中文命名实体识别联合训练进行命名实体识别。我们将本文提出的基于标签一致性机制的多任务学习方法记为Multi-task(Label consistency)。

Setting	Method	P	R	F1
Single-task learning	GRU-CRF	84.71	83.41	84.01
	BiGRU	84.20	80.20	82.00
	BiGRU-CRF	87.21	83.21	85.09
	BERT-CRF	94.96	85.41	89.84
	BERT-BiGRU-CRF	94.95	87.88	91.22
	Collobert et al. (2011)	88.43	87.68	88.05
	Lample et al. (2016)	90.45	89.72	90.08
	Shen et al. (2017)	91.46	90.18	90.81
Multi-task learning	Wang et al. (2013)	93.95	88.96	91.29
	Peng and Dredze (2016)	92.94	90.17	91.52
	Multi-task(Label consistency)	94.09	90.53	92.28

Table 4: 中文命名实体识别全样本实验结果(单位:%)

实验结果如表4所示。本文基于标签一致性机制的多任务学习方法Multi-task(Label consistency)获得92.28%的F1值, 显著优于其他模型及方法。相比基线模型BERT-CRF、BiGRU-CRF、GRU-CRF、BiGRU分别提升2.44%、7.91%、8.27%、10.28%的F1值; 相比BERT-BiGRU-CRF框架下的单任务训练, 提升1.06%的F1值; 与同样融入词信息的Collobert et al. (2011)方法相比提升4.23%; 与Lample et al. (2016)的混合模型方法相比提升2.2%; 与将深度学习和主动学习相结合的Shen et al. (2017)方法比提升1.47%; 与Wang et al. (2013)和Peng and Dredze (2016)的联合训练方法相比有0.99%和0.76%的提升。可见, 本文方法基于标签一致性机制在数据集阶段即强调实体边界的对齐以及词汇信息的增强, 在多任务学习中再进一步共享编码、网络结构与参数进行中文命名实体识别的有效性。

5.2 小样本实验

考虑到多任务学习可能更适用于标注样本稀缺的环境, 我们设置小样本实验。

Setting	Method	F1	
Simulation	BiGRU	81.81	
	Single-task learning	BiGRU-CRF	84.36
		BERT-BiGRU-CRF	89.93
	Multi-task learning	Peng and Dredze (2015)	89.27
		Multi-task(Label consistency)	92.98
Actual	BiGRU	52.49	
	BiGRU-CRF	53.89	
	Single-task learning	BERT-BiGRU-CRF	57.50
		He and Sun (2017)	54.82
		Yang et al. (2018)	58.79
	Multi-task learning	Peng and Dredze (2015)	56.05
	Multi-task(Label consistency)	61.33	

Table 5: 中文命名实体识别小样本实验结果F1值(单位:%)

1)模拟环境：在新闻领域人民日报数据集中仅取原数据集20%的数据作为训练集，验证模型效果。2)真实环境：选用样本语句数仅有新闻领域5%的社交媒体领域微博数据集验证本文方法的有效性。主要对比方法包括：1)Peng and Dredze (2015)基于中文命名实体识别进行嵌入式联合训练。2)He and Sun (2017)提出基于BiLSTM神经网络的半监督学习模型，将转移概率和深度学习相结合。3)Yang et al. (2018)重建了12个神经序列标记模型，研究构建有效和高效的神经序列标记系统。

表5给出了实验结果。本文方法在低资源小样本设置下提升效果显著：1)在新闻领域的模拟环境中，本文方法可达到92.98%的F1值，相比BERT-BiGRU-CRF模型单任务与Peng and Dredze (2015)多任务方法的F1值分别提升了3.05%、3.71%。同时，高于Multi-task(Label consistency)全样本实验0.83%。2)在社交媒体领域的真实环境中，相较于现有方法具有显著提升。其中，与Yang et al. (2018)、Peng and Dredze (2015)、He and Sun (2017)相比分别提高2.54%、5.28%、6.51%的F1值。以上实验结果表明本文提出的基于多任务学习的命名实体识别方法对标注资源匮乏的情况具有更加显著的效果。

5.3 讨论

5.3.1 不同多任务共享方法

在本文选择的模型框架的基础上，我们进一步对比不同的多任务学习训练模式的影响。本文进行了三种多任务共享方法的实验：1)基于联合损失函数的多任务学习Multi-task(Joint loss)，三种任务使用不同数据集，只联合损失的计算，按动态权重共同作用实现模型的优化。2)基于参数共享的多任务学习Multi-task(Parameter sharing)，在联合损失的基础上，三种任务共享BiGRU层及一个线性层，三种任务增加网络结构与参数的共享，使信息交互共享更加全面充分。3)基于标签一致性机制的多任务学习Multi-task(Label consistency)，基于本文数据预标注模块进行的多任务共享学习，在数据集阶段即实现实体边界的对齐，以及词汇信息的增强，再进一步共享编码、网络结构与参数，最后结合标签一致性机制计算联合损失。

Method	Complete sample	Small sample	
Single-task	91.22	89.93	57.50
Multi-task(Joint loss)	91.90	88.71	59.48
Multi-task(Parameter sharing)	92.06	89.98	60.59
Multi-task(Label consistency)	92.28	92.98	61.33

Table 6: 不同多任务共享方法测试结果F1值(单位:%)

实验结果如表6所示，其中基于联合损失函数的多任务学习Multi-task(Joint loss)中全样本与真实小样本的F1值比其单任务分别提升0.68%、1.98%。基于参数共享的多任务学习Multi-task(Parameter sharing)中全样本及两种小样本比其单任务F1值分别提升0.84%、0.05%、3.09%，且均高于Multi-task(Joint loss)方法。本文基于标签一致性机制的多任务学习Multi-task(Label consistency)，在全样本及两种小样本中分别比其单任务提升1.06%、3.05%、3.83%，比上述两种共享方法F1值涨幅更大，效果提升更显著。综上可见多任务学习共享信息越充分，越能获得更好的表示学习及性能优势。训练效率也是评估的重要指标，基于本文方法的多任务学习模式，共同处理三种任务，共享编码特征及网络结构，大大减少训练参数量，且样本量越小效果越明显，训练速度越快，从而增加资源利用率。

5.3.2 分割实验

基于真实环境低资源小样本微博数据集，进行了分割实验，具体方法设置如下：1)Single-task，基于BERT-BiGRU-CRF框架的命名实体识别单任务实验。2)No POS，在本文方法之中去掉词性信息模块，仅保留实体识别与分词模块。3)No Label consistency，遵循本文方法，但在损失的计算中不加入标签一致性损失 $L_{NER.consistency}$ 参与调整网络。4)Multi-task(Label consistency)，完整的本文方法。

结果如表7所示，本文方法中不加入标签一致性损失或词性信息模块相较于完整方法效果均有下降，分别为0.71%和2.38%，验证了融合词性信息模块与标签一致性损失对命名实体识别结果提升的有效性。

Setting	F1
Single-task	57.50
No POS	58.95
No Label consistency	60.62
Multi-task(Label consistency)	61.33

Table 7: 分割实验结果F1值(单位:%)

5.3.3 辅助任务

最后，我们验证该方法是否可以在提高主任务中文命名实体识别结果的同时，也改善分词与词性标注两个辅助任务的效果。实验结果如表8所示，其中包括单任务实验以及基于联合损失函数的多任务学习Multi-task(Joint loss)、基于参数共享的多任务学习Multi-task(Parameter sharing)、基于标签一致性机制的多任务学习Multi-task(Label consistency)三种不同共享方式的多任务实验。

	Method	NER	CWS	POS
新闻领域	Single-task	91.22	94.85	74.61
	Multi-task(Joint loss)	91.90	95.49	77.60
	Multi-task(Parameter sharing)	92.06	95.63	80.28
	Multi-task(Label consistency)	92.28	93.91	84.59
	Multi-task(Label consistency)(小样本)	92.98	93.62	86.35
社交媒体领域	Single-task	57.50	-	-
	Multi-task(Joint loss)	59.48	91.53	81.13
	Multi-task(Parameter sharing)	60.59	91.63	80.89
	Multi-task(Label consistency)	61.33	90.42	84.36

Table 8: 多任务学习测试结果F1值(单位:%)

从结果可以看出，与单任务相比，Multi-task(Joint loss)为分词和词性标注两个辅助任务带来了0.64%和3%的提升。Multi-task(Parameter sharing)方法下分词与词性标注的结果分别提升0.78%和5.67%。Multi-task(Label consistency)方法下，词性标注的结果提升11.74%。

三种方法在多任务学习共享信息的程度上看，是逐步递进的，分别从不同层次对被共享任务做出增强信息共享效果的贡献。不同程度的信息共享适合适应于不同任务，分词和词性标注作为基础任务在低程度的信息共享方式中即可实现有效提升，命名实体识别作为进一步任务，在越高程度的信息共享中效果提升越显著。

6 结束语

针对中文命名实体识别任务，本文提出了一种新的基于多任务学习模式的中文命名实体识别方法，同时将分词和词性标注信息融入到命名实体识别任务中，实现多任务联合训练。同时，提出标签一致性机制，学习命名实体识别、中文分词和词性标注任务的解码相关性，进一步提升效果。在全样本和小样本环境中的实验结果表明了本文方法的有效性。后续将要研究如何在本文方法中融入更多潜在的词汇信息，深度结合小样本学习的优势，应用到更多资源匮乏的领域进行研究，进一步提高中文命名实体识别的性能。

致谢

本项目受国家重点研发计划项目（2019YFB1405200）和国家自然科学基金面上项目（61976015，61976016，61876198，61370130）资助。

参考文献

- Krisztian Balog, Pavel Serdyukov, and Arjen P. de Vries. 2010. Overview of the trec 2010 entity track. In Ellen M. Voorhees and Lori P. Buckland, editors, *TREC*, volume 500-294 of *NIST Special Publication*. National Institute of Standards and Technology (NIST).
- Daniel M. Bikel, Scott Miller, Richard Schwartz, and Ralph Weischedel. 1997. Nymble: a high-performance learning name-finder. In *Fifth Conference on Applied Natural Language Processing*, pages 194–201, Washington, DC, USA, March. Association for Computational Linguistics.
- Daniel M. Bikel, Richard Schwartz, and Ralph M. Weischedel. 1999. An algorithm that learns what's in a name. *Machine Learning*, 34(1-3):211–231.
- Pengfei Cao, Yubo Chen, Kang Liu, Jun Zhao, and Shengping Liu. 2018. Adversarial transfer learning for Chinese named entity recognition with self-attention mechanism. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 182–192, Brussels, Belgium, October–November. Association for Computational Linguistics.
- Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *J. Mach. Learn. Res.*, 12:2493–2537, November.
- Gamal K. O. Crichton, Sampo Pyysalo, Billy Chiu, and Anna Korhonen. 2017. A neural network multi-task learning approach to biomedical named entity recognition. *BMC Bioinform.*, 18(1):368:1–368:14.
- Gianluca Demartini, Tereza Iofciu, and Arjen P. de Vries. 2009. Overview of the inex 2009 entity ranking track. In Shlomo Geva, Jaap Kamps, and Andrew Trotman, editors, *INEX*, volume 6203 of *Lecture Notes in Computer Science*, pages 254–264. Springer.
- 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, Minneapolis, Minnesota, June. Association for Computational Linguistics.
- George Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, and Ralph Weischedel. 2004. The automatic content extraction (ACE) program – tasks, data, and evaluation. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal, May. European Language Resources Association (ELRA).
- Chuanhai Dong, Jiajun Zhang, Chengqing Zong, Masanori Hattori, and Hui Di. 2016. Character-based lstm-crf with radical-level features for chinese named entity recognition. In Chin-Yew Lin, Nianwen Xue, Dongyan Zhao, Xuanjing Huang, and Yansong Feng, editors, *NLPCC/ICCPOL*, volume 10102 of *Lecture Notes in Computer Science*, pages 239–250. Springer.
- Hangfeng He and Xu Sun. 2017. F-score driven max margin neural network for named entity recognition in Chinese social media. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 713–718, Valencia, Spain, April. Association for Computational Linguistics.
- Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2012. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580.
- Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional lstm-crf models for sequence tagging. cite arxiv:1508.01991.
- Hideki Isozaki and Hideto Kazawa. 2002. Efficient support vector classifiers for named entity recognition. In *COLING 2002: The 19th International Conference on Computational Linguistics*.
- Diederik P. Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. cite arxiv:1412.6980 Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.
- Onur Kuru, Ozan Arkan Can, and Deniz Yuret. 2016. CharNER: Character-level named entity recognition. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 911–921, Osaka, Japan, December. The COLING 2016 Organizing Committee.

- John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01*, pages 282–289, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 260–270, San Diego, California, June. Association for Computational Linguistics.
- Haibo Li, Masato Hagiwara, Qi Li, and Heng Ji. 2014. Comparison of the impact of word segmentation on name tagging for Chinese and Japanese. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 2532–2536, Reykjavik, Iceland, May. European Language Resources Association (ELRA).
- Jing Li, Aixin Sun, Jianglei Han, and Chenliang Li. 2018. A survey on deep learning for named entity recognition. *CoRR*, abs/1812.09449.
- Ying Lin, Shengqi Yang, Veselin Stoyanov, and Heng Ji. 2018. A multi-lingual multi-task architecture for low-resource sequence labeling. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 799–809, Melbourne, Australia, July. Association for Computational Linguistics.
- Yanan Lu, Yue Zhang, and Dong-Hong Ji. 2016. Multi-prototype chinese character embedding. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, H el ene Mazo, Asunci on Moreno, Jan Odiijk, and Stelios Piperidis, editors, *LREC*. European Language Resources Association (ELRA).
- Wencan Luo and Fan Yang. 2016. An empirical study of automatic Chinese word segmentation for spoken language understanding and named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 238–248, San Diego, California, June. Association for Computational Linguistics.
- Xuezhe Ma and Eduard Hovy. 2016. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1064–1074, Berlin, Germany, August. Association for Computational Linguistics.
- Nanyun Peng and Mark Dredze. 2015. Named entity recognition for Chinese social media with jointly trained embeddings. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 548–554, Lisbon, Portugal, September. Association for Computational Linguistics.
- Nanyun Peng and Mark Dredze. 2016. Improving named entity recognition for Chinese social media with word segmentation representation learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 149–155, Berlin, Germany, August. Association for Computational Linguistics.
- Rajib Rana. 2016. Gated recurrent unit (gru) for emotion classification from noisy speech. *CoRR*, abs/1612.07778.
- Marek Rei. 2017. Semi-supervised multitask learning for sequence labeling. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2121–2130, Vancouver, Canada, July. Association for Computational Linguistics.
- Yanyao Shen, Hyokun Yun, Zachary Lipton, Yakov Kronrod, and Animashree Anandkumar. 2017. Deep active learning for named entity recognition. In *Proceedings of the 2nd Workshop on Representation Learning for NLP*, pages 252–256, Vancouver, Canada, August. Association for Computational Linguistics.
- Gy orgy Szarvas, Rich ard Farkas, and Andr as Kocsor. 2006. A multilingual named entity recognition system using boosting and c4.5 decision tree learning algorithms. In Ljupco Todorovski, Nada Lavra c, and Klaus P. Jantke, editors, *Discovery Science*, pages 267–278, Berlin, Heidelberg. Springer Berlin Heidelberg.

- Erik F. Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, pages 142–147.
- Mengqiu Wang, Wanxiang Che, and Christopher D. Manning. 2013. Joint word alignment and bilingual named entity recognition using dual decomposition. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1073–1082, Sofia, Bulgaria, August. Association for Computational Linguistics.
- Xuan Wang, Yu Zhang, Xiang Ren, Yuhao Zhang, Marinka Zitnik, Jingbo Shang, Curtis Langlotz, and Jiawei Han. 2018. Cross-type biomedical named entity recognition with deep multi-task learning. cite arxiv:1801.09851Comment: 7 pages, 4 figures.
- Jie Yang, Shuailong Liang, and Yue Zhang. 2018. Design challenges and misconceptions in neural sequence labeling. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3879–3889, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.
- Yue Zhang and Jie Yang. 2018. Chinese ner using lattice lstm. *CoRR*, abs/1805.02023.
- Peng Zhou, Suncong Zheng, Jiaming Xu, Zhenyu Qi, Hongyun Bao, and Bo Xu. 2017. Joint extraction of multiple relations and entities by using a hybrid neural network. In Maosong Sun, Xiaojie Wang, Baobao Chang, and Deyi Xiong, editors, *CCL*, volume 10565 of *Lecture Notes in Computer Science*, pages 135–146. Springer.
- 北京大学计算语言学研究所. 1998. 《人民日报》数据集. 人民日报社新闻信息中心.