

# 基于增强 BiLSTM 的网络文章核心实体识别

周 康,曲卫东,杨艺琛  
(长安大学 信息工程学院,陕西 西安 710064)

**摘 要:**文章核心实体是文章主要描述或担任文章主要角色的实体词。随着互联网的发展与网络传媒平台的兴起,自媒体网络新闻传播严重影响着公众的价值导向,网络文章核心实体识别是情感分析、舆情监测等自然语言处理任务的重要基础步骤。相对于较易识别的传统命名实体,文章核心实体的识别难度更大,其抽取工作不仅需要基于篇章理解的复杂特征信息,还涉及灵活多样的领域表述方式。针对这些问题,提出了一种基于增强双向长短期记忆网络(bidirectional long short-term memory, BiLSTM)的网络文章核心实体识别模型。该模型通过 BiLSTM 来捕捉刻画文章的篇章级特征,在 BiLSTM 模型的基础上利用 Bert 模型提升上下文语义质量,增强原有模型的语义表达能力,同时引入 AdaBoost 集成学习框架来组合多个针对语料数据的不同侧面训练的子模型,从而增强整体的识别效果。文章所做实验验证了该模型的可行性和有效性。

**关键词:**中文命名实体识别;核心实体识别;深度学习;双向长短期记忆网络;AdaBoost

**中图分类号:**TP391

**文献标识码:**A

**文章编号:**1673-629X(2021)01-0007-06

doi:10.3969/j.issn.1673-629X.2021.01.002

## Core Entity Recognition of Web Articles Based on Enhanced BiLSTM

ZHOU Kang, QU Wei-dong, YANG Yi-chen  
(School of Information Engineering, Chang'an University, Xi'an 710064, China)

**Abstract:** The core entity of the article is the entity word that mainly describes or plays the main role of the article. With the development of the Internet and the rise of online media platforms, self-media online news dissemination has seriously affected the public's value orientation. The core entity recognition of web articles is an important basic step for natural language processing tasks such as sentiment analysis and public sentiment monitoring. Compared with the more easily-recognized traditional named entities, the recognition of the core entities of the article is more difficult. Its extraction work not only requires complex feature information based on chapter-level understanding, but also involves flexible and diverse domain representations. Aiming at these problems, we propose a core entity recognition model for web articles based on enhanced BiLSTM (bidirectional long short-term memory). This model uses BiLSTM to capture the text-level features of the article. Based on the BiLSTM model, the Bert model is used to improve the semantic quality of the context and enhance the semantic expression ability of the original model. At the same time, the AdaBoost integrated learning framework is used to combine multiple sub-models trained on different sides of the corpus data to enhance the overall recognition effect. The experiments verify the feasibility and effectiveness of this model.

**Key words:** Chinese named entity recognition; core entity recognition; deep learning; BiLSTM; AdaBoost

## 0 引言

随着互联网的迅速发展,网络文本数量呈现指数级增长,逐渐成为21世纪最有价值的资源之一。以文章为主体进行情感分析、舆情监测等任务也逐渐成为自然语言处理(natural language processing, NLP)领域的热门方向之一。识别文章的核心实体是进行情感分析、舆情监测等工作的基础步骤,要想从互联网海量

的文本信息中获取文章核心实体,依靠传统人工阅读的方式需要消耗大量时间和人力成本,如何迅速准确地获取文章的核心实体成为当下亟待解决的问题。随着机器学习技术在自然语言处理领域的兴起,针对文本信息中语义不明晰等问题,命名实体识别(named entity recognition, NER)的研究逐渐转向了机器学习方向,特别是随着近年来深度神经网络的应用命名实体

收稿日期:2020-02-12

修回日期:2020-06-16

基金项目:长安大学中央高校基本科研业务费专项资金资助(300102248806)

作者简介:周 康(1993-),男,硕士研究生,研究方向为自然语言处理;曲卫东,副教授,硕导,研究方向为 Web 信息检索、文本挖掘、语音信息检索与自然语言处理、大数据及其智能处理。

识别系统的性能得到了大幅提高。Collobert 等人<sup>[1]</sup>率先将深度神经网络应用到命名实体识别,极大地改进了识别的性能。尹继泽等人<sup>[2]</sup>运用注意力机制和长短期记忆网络(long short-term memory, LSTM),通过提取文本全局和局部特征,进一步提高了识别性能。目前双向长短期记忆网络(bidirectional long short-term memory, BiLSTM)结合条件随机场(conditional random field, CRF)已被广泛应用于文本情感分析<sup>[3]</sup>、关键词提取<sup>[4]</sup>、命名实体识别<sup>[5]</sup>等自然语言处理任务中,其中以生物学领域的命名实体识别研究最为成功<sup>[6-9]</sup>。

实体是人、物、地区、机构、团体、企业、行业、某一特定事件等固定存在,且可以作为文章主体的实体词。命名实体识别一般针对特定领域的实体词,由于局部特征明显,模型多使用句子级的特征,模型的识别能力比较强。与一般命名实体识别不同,文章核心实体是文章主要描述或担任文章主要角色的实体词。文章核心实体的识别则需要基于文章篇章级的理解,简单的实体词附近的局部特征信息并不适用于该任务。同时文章核心实体识别面向通用领域,文章复杂多类,核心实体又是篇章语义上的全文核心词,需要结合篇章级的特征信息进行识别。

此外,网络文章通常涉及的范围十分广泛,不同领域文章的表述方式也差别较大,其领域特征信息复杂多样,目前流行的识别方法并不能很好地适应文章核心实体识别任务的要求。为此该文提出一种基于增强 BiLSTM 的文章核心实体识别方法。该方法通过发挥 BiLSTM 对变长序列的处理和记忆能力来捕捉文章篇章级的特征信息<sup>[10]</sup>,同时引入 AdaBoost 集成学习算法,通过调整训练数据分布对让系统对数据集不同侧面分别进行模型化,最后通过组合各细分子类的多分类器全面准确地模型化多领域混合的复杂数据,进而增强模型总体的识别能力。

## 1 增强 BiLSTM 模型

传统的命名实体识别由于实体词附近的局部特征明显,识别时大多只需要句子级的特征,且针对单一领域语料具有很好的效果。而网络文章核心实体的识别由于是基于内容理解进行识别,所以还需要篇章级的结构特征信息,其抽取所要结合的全文语义关系特征并不明显。同时网络文章面向通用领域,内容包含复杂多类的领域特征信息,很难通过单个模型进行全面描述。

针对以上问题,该文将长短不一的文章按照字符映射入 BiLSTM 模型变长的网络结构中,利用 BiLSTM 的记忆单元和双向特性解决篇章级的语义理

解问题。但模型还是难以同时应对复杂的多领域特征信息,为此该文又引入 AdaBoost 集成学习算法以增强原有模型对数据复杂多类的特征进行捕捉刻画,通过调整训练集数据分布,针对数据特征不同侧面训练多个模型,每个模型只关注数据集复杂特征的一个侧面,最终通过组合多个关注数据特征不同侧面的子模型来刻画复杂数据的全局特性,增强模型总体的识别能力。

### 1.1 变长文本和篇章级特征

网络文章的篇幅长短不一致,文章核心实体的数量、位置也不确定,其抽取难度非常大,需要结合篇章级的语义理解,结合不同文本类别的文档特征来识别出文章的核心实体,同时忽略其他语义类别的实体。

基于机器学习的 NER 问题本质上是一个标签序列预测问题,文章核心实体识别亦是如此。LSTM<sup>[11]</sup>继承了递归神经网络(recurrent neural network, RNN)对时序问题的处理能力,通过引入输入门和忘记门控制梯度流动,并能更好地保持长期依赖关系,使得模型能够把握篇章级的特征信息,找到全文的核心实体。BiLSTM 由一个前向 LSTM 和一个后向 LSTM 组成。前向 LSTM 学习序列中各个 token 的上文对映下文信息,后向 LSTM 学习序列中各个 token 的下文对映上文信息,由此构成全文的上下文依赖。BiLSTM 接受文本字符向量序列作为输入,输出文本的隐含状态序列,即预测标签序列。每个时间步  $t$  中,通过前一时刻前向隐藏状态  $h'_{t-1}$  和当前时刻输入  $x_t$  计算得到当前时刻前向隐藏状态  $h'_t$ ,通过前一时刻后向隐藏状态  $h^b_{t-1}$  和当前时刻后向输入  $x_t$  计算得到当前时刻后向隐藏状态  $h^b_t$ ,连接前向后向隐藏状态得到当前时刻隐藏状态  $h_t = [h'_t; h^b_t]$ 。对于输入文本  $S = \{s_1, s_2, \dots\}$ ,输出文本的隐含状态序列  $O = \{o_1, o_2, \dots\}$ 。

BERT (bidirectional encoder representation from transformers)<sup>[12]</sup>是 Google 提出的一种新型语言模型,该模型较完整地保存了文本语义信息,提升了模型的上下文双向特征抽取能力,在各类 NLP 任务上取得了目前最好的结果。

为充分利用文本的篇章级语义信息,该文采用 BERT 预训练模型编码字符级别的语境信息,提升字嵌入的上下文语义质量,增强模型的语义表达能力。同时还在 BiLSTM 后加入 CRF 层使用条件随机场模型对输出进行解码,通过标签间约束条件进一步提升标签序列预测准确性,更好地把握文章全局特性,得到最终的预测标签序列。采用的 BiLSTM 模型结构如图 1 所示。

### 1.2 AdaBoost

由于网络文章核心实体识别是面向通用领域的抽取任务,而且多领域导致文章具有复杂的篇章语义等

特性,单个的 BiLSTM 模型难以同时对多个领域复杂类型的语料进行很好地模型化,使得核心实体识别效

果较差,因此提升模型对复杂多类数据特征的模型化能力在实际应用中很有必要。

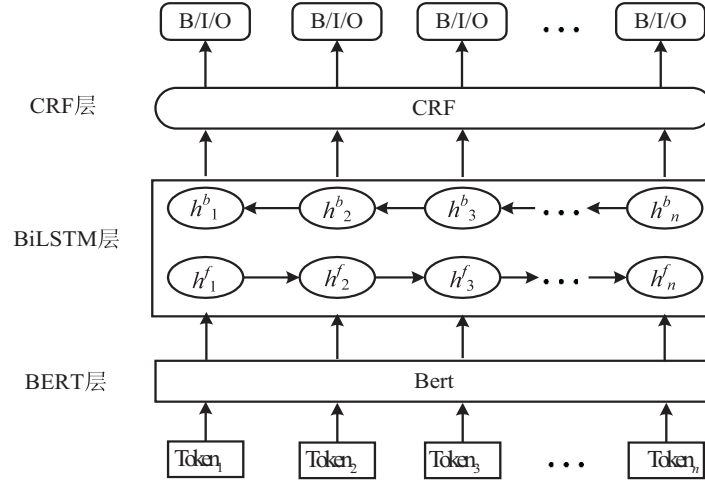


图1 BiLSTM 模型结构

Boosting 是一种重要的集成学习技术,能够有效提升弱分类器的准确率。目前有很多 Boosting 算法,常用的 Boosting 算法是 Freund 和 Schapire 于 1995 年提出的 AdaBoost (adaptive boosting) 算法<sup>[13]</sup>。AdaBoost 是一种把多个简单弱分类器组合成强分类器的方法,最终的输出是由关注数据特征不同侧面的子分类器加权合成计算而得。

常见的 AdaBoost 与弱分类器组合有:使用 AdaBoost 提升决策树准确率<sup>[14]</sup>以及使用 AdaBoost 提升支持向量机(support vector machine, SVM)的准确率<sup>[15-16]</sup>等。新闻事件核心实体识别过程本质上是对预测标签序列进行分类的过程,受此启发,该文提出使用 AdaBoost 算法框架,改善 BiLSTM 模型对复杂数据全局特征的捕获能力,通过将多个关注不同细分子类的子模型组合为更强大的模型,来提高模型整体的分类预测能力。

由于 BiLSTM 模型与决策树、SVM 等分类器不同,需要对原有 AdaBoost 算法与弱分类器结合的方法进行适当修改,以使得 AdaBoost 与序列预测模型结合。AdaBoost 是通过调整训练集中样本权重值来改变每个子模型训练数据的分布,从而获得关注数据特征不同侧面的模型,然后通过各个子模型的正确率来判定各个子模型对最终识别结果的贡献。下面给出该算法的实现细节。

给定一个标记样本集  $X, X = \{x_i, i = 1, 2, \dots, N\}$ ,  $y_i$  是  $x_i$  对应的标签。 $D_k$  为样本分发向量,  $D_k^i$  是  $x_i$  对应的权值,被用于概率加权随机抽取模型训练集,并根据模型错误率来更新。 $D_k$  的总和为 1,初始分布为  $1/N$ 。AdaBoost 通过  $K$  次迭代训练  $K$  个模型并合并它们,在每次迭代  $k = 1, 2, \dots, K$  中,利用分发权值向量  $D_k$  从标记样本集中概率加权随机抽取  $N$  个样本为训练集,使

用此训练集训练 BiLSTM 模型得到一个最优的模型,测试标记样本  $X$  内所有样本,计算模型错误率  $\varepsilon_k =$

$$\sum_{i=1}^N D_k^i |h_k(x_i) - y_i|, \text{进而取得模型权重 } \beta_k = \frac{1}{2} \ln \frac{1 - \varepsilon_k}{\varepsilon_k}。模型的预测效果越好其权重越大。每次迭代$$

后根据训练结果算法还要更新样本集的分布,对于训练失败的样本赋予较大的权重,使得下一轮迭代更关注这些样本,  $D_{k+1}^i = D_k^i \beta_k^{1 - |h_k(x_i) - y_i|}$ 。在  $K$  轮迭代后,生成  $K$  个不同的模型,最终通过预测函数  $H$  组合各个模型  $H(x_i) = \text{sign}(\sum h_k(x_i) * \beta_k)$ 。具体算法如下所示:

算法:AdaBoost 算法。

输入:已标注样本集  $\langle (x_1, y_1), \dots, (x_n, y_n) \rangle$ 、样本权重向量  $D$ 、BiLSTM 模型参数、迭代次数  $K$ 。

初始化:样本权重为  $D_1^i = \frac{1}{N}, i = 1, 2, \dots, N$ 。

(1) for  $k = 1, 2, \dots, K$

(2) 从已标注样本集中依照权重  $D_k$  随机抽取训练集  $T$ 。

(3) 使用训练集  $T$  训练 BiLSTM 模型  $h_k$ 。

(4) 使用模型预测已标注样本集并计算模型  $h_k$  的

错误率  $\varepsilon_k = \sum_{i=1}^N D_k^i |h_k(x_i) - y_i|$ 。

(5) 计算模型  $h_k$  的权重为  $\beta_k = \frac{1}{2} \ln \frac{1 - \varepsilon_k}{\varepsilon_k}$ 。

(6) 更新样本权重  $D_{k+1}^i = D_k^i \beta_k^{1 - |h_k(x_i) - y_i|}$ 。

(7) end for

输出:模型最终输出为各子模型加权预测结果

$$H(x_i) = \text{sign}(\sum h_k(x_i) * \beta_k)。$$

### 1.3 增强 BiLSTM 模型结构

该文提出的经 AdaBoost 增强的 BiLSTM 模型总

体结构如图 2 所示。

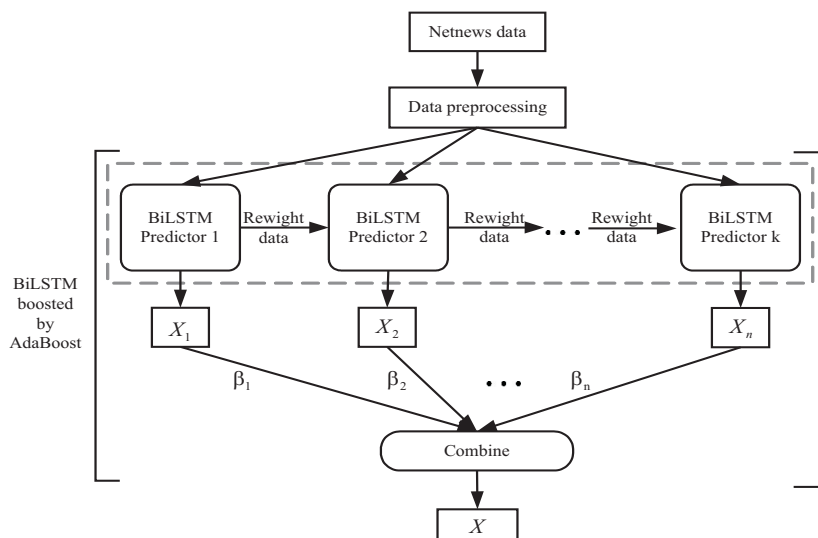


图 2 增强 BiLSTM 模型结构

AdaBoost 算法以迭代的方式训练 BiLSTM 网络，经过  $K$  轮迭代后得到关注数据特征不同侧面的  $K$  个 BiLSTM 模型。在每一轮迭代中，通过对已标注新闻语料进行概率随机抽样形成训练集并送入 BiLSTM 网络进行模型训练，训练结束后根据各模型的错误率生成模型权重  $\beta$  并调整训练数据集的样本概率权重，从而改变下一轮数据分布。数据分布的调整遵照加大错误样本权重减小正确样本权重的原则进行，这就使得下一轮的模型训练更关注于模型未准确捕捉的数据特征的那些样本集合。最后依照模型权重组合  $K$  个 BiLSTM 模型得到增强 BiLSTM 模型来进行网络文章

核心实体的识别。

## 2 实验结果及分析

### 2.1 实验数据集

实验所涉及的语料来自搜狐新闻门户网站爬取的新闻数据，其时间范围为 2019-03-01 至 2019-03-10，共计 5 万篇新闻报道，涉及民生、娱乐、科技、政治等多个领域，文章结构、表述方式复杂多样。采用人工统计法标注新闻数据集，每篇新闻文本依据所报道事件标明 2 到 3 个核心实体，图 3 为语料样例，核心实体为：“3d”、“工业”和“机器视觉”。图 4 为标记样本片段。

```

{"newsId":"4e36d02a","coreEntity":["3d","工业","机器视觉"],"title":"sia智慧工厂展，誉洋以“智”取胜","content":"第十七届上海国际工业自动化及机器人展与上海智能工厂展览会于2019年3月1日圆满落下帷幕。……"}
  
```

图 3 语料样例

```

B I O O O O O O O O O O O O O O O O O O B I I
大 数 据 之 热 度 ， 已 无 需 多 言 。 业 内 众 多 关 于 大 数 据 …
  
```

图 4 标记样本片段

其中，标签策略借鉴命名实体识别的 BIO 系统，B 表示核心实体的首字符，I 表示核心实体首字符以外的其他实体字符，O 表示非核心实体字符。实验中将带有核心实体标签的数据分为训练集和测试集。各数据集集中的文本统计情况如表 1 所示。

表 1 数据集文本分布

	训练集	测试集
文本数	45 000	5 000
字符数	29 250 342	3 249 655

### 2.2 参数设置

在增强 BiLSTM 模型的各子模型训练过程中，首

先要得到句子级字向量序列。模型以句子为单位送入 Bert 模型生成字嵌入向量序列。输入句子最大长度为 512，长度不足的句子以 <PAD> 标签进行补充，从而得到 768 维字向量序列，长度为 514，去除 <CLS>、<SEP> 和 <PAD> 标签对应的向量后得到真实的句子字向量序列。

接下来进入 BiLSTM 层学习文本中全局特征信息。在 BiLSTM 参数学习过程中，隐含层设置为 400 维，BiLSTM 输出隐含层后经过一个全连接网络对其进行线性变换，降低维度，最终送入 CRF 层。输出层的神经元个数为 3，即 3 种分类标签。在训练过程中，



输入的批量文本大小  $\text{batch\_size}$  为 50, 学习率初始值为 0.001。模型使用优化的随机梯度下降算法训练网络参数并动态更新学习率的大小, 其迭代次数设为 100。

在使用 AdaBoost 进行迭代训练模型时, 初始训练集文章权重分布为  $\frac{1}{N}$ ,  $N$  为训练集大小, 在每次迭代后根据模型错误率调整权重分布, 迭代次数  $K$  设为 20。

### 2.3 实验结果分析

CRF++ 和 BiLSTM 分别是统计和深度学习领域比较有代表性的序列标注模型, 为此该文除了使用增强 BiLSTM 进行模型的构建与实验, 还引入了普通 BiLSTM 和 CRF++ 进行对比实验来验证所提出模型的有效性。

实验中核心实体词识别的综合得分 F1-score 计算方法如下:

$$\text{F1-score}(\text{Entities}) = \frac{1}{2} \sum_{0 \leq i \leq n} 2 * \frac{\text{Precision}_i(\text{Entities}) * \text{Recall}_i(\text{Entities})}{\text{Precision}_i(\text{Entities}) + \text{Recall}_i(\text{Entities})} \quad (1)$$

实验使用 TensorFlow 深度学习框架, 借助 GPU 加速模型的训练过程, 表 2 展示了文中方法、普通 BiLSTM 模型和 CRF++ 在测试数据集上核心实体识别的精准度 (Precision)、召回率 (Recall) 和 F1-score 综合得分情况。

表 2 模型性能对比

模型	Precision	Recall	F1-score
CRF++	0.491	0.414	0.449
BiLSTM	0.544	0.396	0.458
增强 BiLSTM	0.630	0.466	0.536

由表 2 中的数据可看出, 在进行文章核心实体识别时, 普通 BiLSTM 模型的识别效果略优于 CRF++, 该文提出的增强 BiLSTM 模型在精准度、召回率和 F1-score 上均优于普通 BiLSTM 模型和 CRF++。说明在应对复杂类型语料时, CRF++ 和普通 BiLSTM 模型无法获得数据的全局特征, 而增强 BiLSTM 模型在普通 BiLSTM 的基础上极大地提高了模型化复杂数据全局特征的能力。增强 BiLSTM 模型通过从不同侧面关注数据特征, 组合多个关注数据集不同侧面的子模型形成最终的增强预测模型。实验结果证明了提出的模型的有效性, 其识别性能比普通 BiLSTM 方法提升了 17.03%, 比 CRF++ 提升了 19.37%。

AdaBoost 算法框架对 BiLSTM 模型的增强效果取决于各个子模型对数据集特征的捕获能力, 为此实验对各个子模型的特征捕获能力进行了实验研究。实

验使用同一测试集对各个子模型进行了测试, 并对各个测试结果中的正确预测样例集合进行了对比分析。采用 Jaccard 相似系数 (Jaccard similarity coefficient) 来衡量各个正确预测样例集合的相似度与差异性, Jaccard 相似系数计算方法如下:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (2)$$

其中,  $A$  表示子模型  $A$  的预测结果中的正确预测样例集合,  $B$  表示子模型  $B$  的预测结果中的正确预测样例集合。Jaccard 相似系数的值域是 0 到 1, 相似系数越大, 两个子模型对测试集的正确预测样例集合相似度越高, 两个子模型的能力越相近, 它们所捕获的特征也就越相似。实验计算了所有子模型对应的正确预测样例集合两两之间的 Jaccard 相似系数, 平均相似系数为 0.483。这说明各个子模型所捕获的数据集的特征不完全相同, 它们对同一测试集展现出不同的预测能力。例如: 子模型 1 对应的正确预测样例集合中大多是关于娱乐新闻的文章, 其核心实体大多包含明星姓名; 而子模型 4 对应的正确预测样例集合中多是关于金融的文章, 其核心实体大多包含公司名称。从而推断子模型 1 和子模型 4 具有不同的能力特性, 它们捕获了不同的数据特征。因此, AdaBoost 可以通过自适应的调整数据分布使得各个子模型的训练数据具有不同偏向性, 从而使各个子模型能够捕获数据集特征的不同侧面, 即数据集的特征中的一部分。

实验还对增强 BiLSTM 模型中 AdaBoost 算法中的迭代次数  $K$  对模型的影响进行了实验研究, 图 5 展示了当  $K$  从 1 到 20 时模型 F1 值的变化趋势。

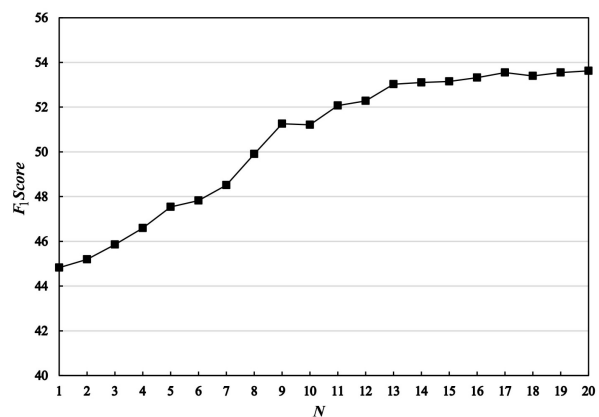


图 5 F1 值变化趋势

由图 5 可以看出, 增强 BiLSTM 模型随着迭代次数的增加, 模型性能逐渐提升, 当迭代次数达到 20 时, 模型性能提升趋势趋于平缓。增强 BiLSTM 通过调整数据分布来关注数据的不同侧面, 每增加一次迭代, 其实是对数据集为模型化特征的一次模型化, 表现为增加错误率高的样本的权重使得下一次训练重点关注该样本。随着迭代次数的增加, 在模型能力范围内未被

模型刻画的特征逐渐变少,模型的增强效果趋势变缓。因此,较于普通 BiLSTM 模型,增强 BiLSTM 模型在面对复杂领域数据时性能较好,但并不能通过无限增加子模型数量来提升模型整体性能,合适的迭代次数  $K$  值应依据数据集特征的复杂程度和模型实际效果而定。

### 3 结束语

介绍了基于增强 BiLSTM 的文章核心实体识别方法:通过 BiLSTM 模型关注文本中字符之间的联系,挖掘文章篇章级的语义特征,使用 AdaBoost 增强模型加强对通用领域复杂数据特征的捕捉能力,从而更全面地刻画数据集全局特征。实验结果表明,该方法在核心实体识别任务上相比传统模型性能有了进一步的提升,从而为情感分析、舆情监测等任务提供更好的支持。增强 BiLSTM 的不足之处是其结构导致模型不能较好地支持并行训练,训练时间成本较高,如何提高模型的并行训练能力将是接下来的研究重点。

#### 参考文献:

- [1] COLLOBERT R, WESTON J. A unified architecture for natural language processing: deep neural networks with multitask learning [C]//25th international conference on machine learning. Helsinki, Finland: ACM, 2008: 160–167.
- [2] GUL Khan Safi Qamas, 尹继泽, 潘丽敏, 等. 基于深度神经网络的命名实体识别方法研究[J]. 信息安全学报, 2017(10): 29–35.
- [3] 李 洋, 董红斌. 基于 CNN 和 BiLSTM 网络特征融合的文本情感分析[J]. 计算机应用, 2018, 38(11): 3075–3080.
- [4] 陈 伟, 吴友政, 陈文亮, 等. 基于 BiLSTM-CRF 的关键词自动抽取[J]. 计算机科学, 2018, 45(6A): 91–96.
- [5] LAMPLE G, BALLESTEROS M, SUBRAMANIAN S, et al. Neural architectures for named entity recognition [C]//Proceedings of the 2016 conference of the North American Chapter of the association for computational linguistics: human language technologies. San Diego: ACL, 2016: 260–270.
- [6] SETTLES B. Biomedical named entity recognition using conditional random fields and rich feature sets [C]//JNLPBA '04: proceedings of the international joint workshop on natural language processing in biomedicine and its applications. Stroudsburg: ACL, 2004: 104–107.
- [7] LEAMAN R, GONZALEZ G. BANNER: an executable survey of advances in biomedical named entity recognition [C]//Pacific symposium on biocomputing. [s. l.]: [s. n.], 2008: 652–663.
- [8] TANG B, CAO H, WANG X, et al. Evaluating word representation features in biomedical named entity recognition tasks [J]. BioMed Research International, 2014, 2014: 240203.
- [9] KAZAMA J, MAKINO T, OHTA Y, et al. Tuning support vector machines for biomedical named entity recognition [C]//Proc of the workshop on natural language processing in the biomedical domain. Stroudsburg: ACL, 2002: 1–8.
- [10] GREENBERG N, BANSAL T, VERGA P, et al. Marginal likelihood training of BiLSTM-CRF for biomedical named entity recognition from disjoint label sets [C]//Proceedings of the 2018 conference on empirical methods in natural language processing. Brussels: ACL, 2018: 2824–2829.
- [11] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735–1780.
- [12] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]//NAACL-HLT 2019: annual conference of the north american chapter of the association for computational linguistics. Minneapolis: ACL, 2019: 4171–4186.
- [13] FREUND Y, SCHAPIRE R E. A decision-theoretic generalization of on-line learning and an application to boosting [C]//Proceedings of the II European conference on computational learning theory. Berlin, Heidelberg: Springer, 1999: 119–139.
- [14] WU Yanli, KE Yutian, CHEN Zhuo, et al. Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping [J]. CATENA, 2020, 187: 104396.
- [15] 梅 比, 汪旭光, 杨仁树. 基于 Adaboost-SVM 组合算法的爆破振动强度预测研究 [J]. 振动与冲击, 2019, 38(18): 231–235.
- [16] 王 燕, 曹建芳, 李艳飞. 融合混合优化组合的大规模场景图像分类算法 [J]. 计算机技术与发展, 2019, 29(9): 86–91.