文章编号:1672-3961(2024)02-0096-07 DOI:10.6040/j.issn.1672-3961.0.2022.353

# 结合自训练模型的命名实体识别方法

肖伟[1,2],郑更生[1,2*],陈钰佳[1,2]

(1.武汉工程大学计算机科学与工程学院、人工智能学院, 湖北 武汉 430205; 2.智能机器人湖北省重点实验室, 湖北 武汉 430205)

**摘要**:针对命名实体识别数据集中存在某些实体类别样本过少,使模型学习该类别特征能力较差,导致整体性能较低的问题,提出结合自训练模型的命名实体识别方法。利用已有的命名实体识别数据集训练一个教师模型,通过改进的文本相似度函数搜寻与原数据集最相似的无标签文本,利用教师模型对无标签文本生成伪标签,并将伪标签与有标签数据集混合重新训练一个学生模型用于下游的命名实体识别任务。试验结果表明,相较基线模型,该方法在公共数据集 MSRA、CONLL03 和法律实体识别数据集上取得更优的性能。

**关键词**:命名实体识别;自训练;文本相似度;自然语言处理;少样本

**中图分类号**:TP391 **文献标志码**:A

**引用格式**:肖伟,郑更生,陈钰佳. 结合自训练模型的命名实体识别方法[J].山东大学学报(工学版),2024,54(2):96-102.

XIAO Wei, ZHENG Gengsheng, CHEN Yujia. Named entity recognition method combined with self-training model [J]. Journal of Shandong University (Engineering Science), 2024, 54(2):96-102.

# Named entity recognition method combined with self-training model

XIAO Wei[1,2], ZHENG Gengsheng[1,2*], CHEN Yujia[1,2]

(1. School of Computer Science & Engineering Artificial Intelligence, Wuhan Institute of Technology, Wuhan 430205, Hubei, China; 2. Hubei Key Laboratory of Intelligent Robot, Wuhan 430205, Hubei, China)

**Abstract**: Aiming to address the issue of insufficient samples for certain entity categories in the named entity recognition dataset, which hampered the model's ability to learn the category's features and resulted in lower overall performance, this study proposed a named entity recognition method that incorporated a self-training model. A teacher model was trained using the available named entity recognition dataset. The improved text similarity function was used to search for unlabeled text that was most similar to the original dataset. The teacher model was utilized to generate pseudo-labels for the unlabeled text. These pseudo-labels were then combined with the labeled dataset to retrain a student model for the downstream named entity recognition task. The experimental results showed that, compared with the baseline model, the method achieved even better performance on the public datasets MSRA, CONLL03, and the legal entity recognition dataset.

**Keywords**: named entity recognition; self-training; text similarity; natural language processing; few-shot

## 0 引言

命名实体识别(named entity recognition, NER)作为自然语言处理任务中的基础任务受到越来越多的关注。它作为许多自然语言处理下游任务的第一步,性能识别的好坏直接影响下游任务,如知识图谱构建、关系抽取、机器翻译等。深度学习是命名实体识别任务的主流方法,而命名实体识别数据集中常常存在某个类别样本较少,导致模型对此类别样本特征学习能力较弱的情况。因此,命名实体识别数据集在只有少量学习样本时会导致样本特征学习能力较差,影响整体模型性能。

为了解决类别样本较少且期望用低成本解决的问题,在图像处理领域首先提出自训练(self-training, ST)[1]方式。自训练是一种半监督的方

**收稿日期**:2022-10-18

**基金项目**:国家自然科学基金青年基金项目(62106179)

**第一作者简介**:肖伟(1998— ),男,安徽枞阳人,硕士研究生,主要研究方向为命名实体识别和关系抽取。E-mail:1984717494@qq.com

*通信作者简介:郑更生(1971— ),男,湖南邵东人,副教授,硕士生导师,博士,主要研究方向为自然语言处理和嵌入式系统。

E-mail:45521534@qq.com

法,利用有标记数据训练一个教师模型,再利用教师模型对大量无标记数据生成伪标记,最后和有标记数据混合起来达到数据增强的效果。由于数据集不同,命名实体识别需要识别的实体类型也不同,选择不同的无标记数据进行自训练将会对模型产生影响。例如,选择的无标记数据与原有标记训练集数据差异过大,自训练的伪标记数据带有较大噪声,反而会降低模型的识别性能。尽管可以通过文本相似度函数、聚类等方法进行选取,但由于数据集间的差异性较大,在选择无标签数据合适性上仍然存在不足。通过对命名实体识别数据集进行分析发现,类别样本过少会导致数据集中某一类别识别性能较差。

针对上述问题,本研究提出结合自训练的命名实体识别模型。在 MSRA、CONLL03 和法律实体识别数据集 3 个公共数据集上进行了验证,试验结果表明本研究提出方法的有效性。本研究的贡献如下。

(1) 结合自训练的方法,通过利用无标记数据集生成伪标记,以增强命名实体识别数据集中因某个实体类别数量较少导致整体性能较弱的问题,模型实现更好地学习实体特征。

(2)在充分分析命名实体识别数据集的基础上,结合数据集的特点改进了文本相似度函数,选择最合适训练集的无标记数据,减少生成的伪标记数据给模型带来的噪声。

(3)提出的自训练方法结合改进的文本相似度函数为解决命名实体识别中某些实体类别数据过少提供一种新思路。

## 1　相关工作

目前,命名实体识别主流的方法是基于神经网络的序列标注学习模型。传统的序列标注模型常用词向量表示进行学习,例如,Word2Vec[2] 和 GloVe[3]。文献[4]利用 Word Embedding 方法将文本中的每个字符转换成对应的 Character Embedding 向量。文献[5]在传统的序列标注模型的基础上,融合双向长短记忆网络和条件随机场的基于字符特征和无监督词表征。文献[6-7]使用卷积神经网络从字符中抽取特征。随着大规模预训练模型双向编码器表示(bidirectional encoder representations from transformers, BERT)面世,命名实体识别任务发展到一个新的阶段[8]。文献[9]提出一种基于 BERT 和多头注意力的中文命名实体识别模型。文献[10]提出一种跨度语义增强的命名实体识别方法。文献[11]在 BERT 的基础上提出一种统一的机器阅读理解框架用于命名实体识别,利用外部知识增强学习特征。文献[12]将机器阅读理解模型应用于中文分词任务,取得了不错的性能。

然而,命名实体识别的序列标注模型是基于监督学习模型,需要大量标记数据。而命名实体识别数据集中常存在类别样本过少,导致模型学习该类别特征能力较弱的问题。文献[13]利用先验知识增强图的少样本学习特征。文献[14]利用图卷积神经网络方法增强训练集上的少样本学习。文献[15]探索了在少样本学习情况下文本分类的应用。文献[16]将标签的实体类别转换成语义表示,与句子部分做数量积操作,生成 NER 标签。

尽管现有的研究在小样本数据集上取得了不错的效果,但还存在以下不足:小样本学习针对数据集整体,而没有关注其中某个类别样本较少的情况,导致影响了模型的整体性能。文献[1]重新探索了自训练和预训练方法在图像处理领域的地位作用。文献[17]将自训练方法用于机器阅读理解,取得了不错的性能。文献[18]通过增添噪声提高自训练方法在图像分类的鲁棒性。文献[19]将自训练方法和远程监督方法结合,应用于命名实体识别任务中,提高了模型的泛化能力。文献[20]提出对于生成的伪标签,根据其重要性程度设置不同权重。

但是,命名实体识别数据集不同的类别数据会对结果造成不同的影响,而以上的方法并没有考虑如何挑选合适类别的无标签数据。

## 2　结合自训练的命名实体识别模型

### 2.1　BERT 预训练模型

BERT 预训练模型基于 Transfomer[21] 进行改进,区别于传统的 word2vec 模型,依赖于自注意力机制,通过拼接多个 Embedding 捕获句子上下文特征。自注意力公式为:

$$z=A(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=f_{\text{softmax}}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}, \tag{1}$$

式中:$z$ 为加权和矩阵;$A$ 为自注意力机制的操作;$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 分别为 Query、Key 和 Value 组成的向量矩阵,由每个输入的词向量与训练后的权重矩阵 $\boldsymbol{W}_Q$、$\boldsymbol{W}_K$ 和 $\boldsymbol{W}_V$ 相乘所得;$d_k$ 为 Query 向量的维度,在式中除以 $\sqrt{d_k}$ 可以使模型在训练过程中梯度平稳下降;$f_{\text{softmax}}$ 用于归一化处理。

BERT 预训练模型由 Token Embedding、

Segment Embedding 和 Position Embedding 3 个 Embedding 拼接而成,很好地融合了字符、句子和位置信息。通过将拼接而成的 Embedding 输入 BERT 预训练模型中进行微调,即可输出所需预测的隐藏层特征。

### 2.2 自训练模型

结合自训练的命名实体识别模型如图 1 所示。本研究将模型分为 4 个步骤。第 1 步,利用有标记数据通过预训练模型 Roberta 训练一个教师模型,其中,使用交叉熵函数去微调模型。第 2 步,通过改进的文本相似度函数,在大量的无标记数据库搜寻与原训练集最相似的数据。无标记数据取自大量的网络数据,例如维基百科。第 3 步,利用教师模型对挑选出的无标记数据进行预测,生成伪标记数据。伪标记数据是指模型的预测值而非真实值。第 4 步,将有标记数据和伪标记混合重新训练出一个学生模型,用于下游的命名实体识别任务。

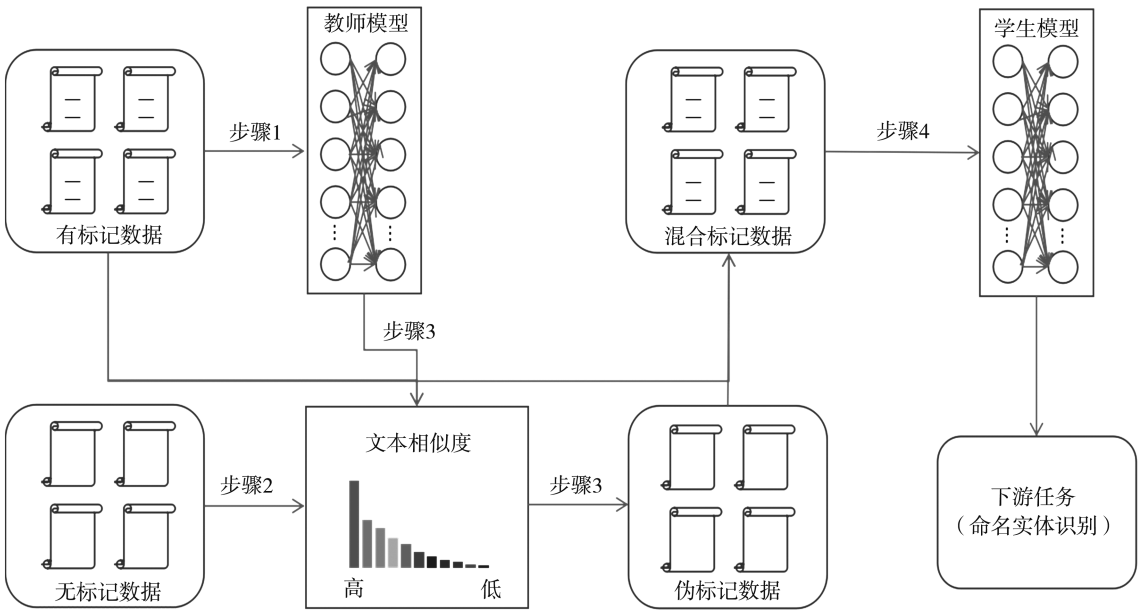


图 1 结合自训练的命名实体识别模型

Fig.1 Combined with self-training named entity recognition model

本研究进行的数据增强技术依赖于无标记数据质量的选择。由于命名实体识别数据集中存在某个实体类别样本少,导致模型对该实体的特征学习能力较差的问题,如何从大量无标记数据库中挑选合适的数据成为本研究的重点。本研究考虑了余弦相似度函数(COSINE)、词频-逆文档频率(term frequency-inverse document frequency, TF-IDF)和 BM25 算法搜寻无标记数据集。

$$\cos\theta(\boldsymbol{x},\boldsymbol{y})=\frac{\boldsymbol{y}_i\sum_{i-1}^{n}\boldsymbol{x}_i}{\sqrt{\sum_{i-1}^{n}\boldsymbol{x}_i^2\sum_{i-1}^{n}\boldsymbol{y}_i^2}},\tag{2}$$

式中 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 为两个文本的向量表示。

$$w(d,t)=t_{\mathrm{f}}(d,t)\times\log\left(\frac{N}{d_{\mathrm{f}}(t)}\right)=t_{\mathrm{f}}(d,t)\times i_{\mathrm{df}}(t),\tag{3}$$

式中,$w(d,t)$ 为词频和逆文档频率的乘积,$t_{\mathrm{f}}(d,t)$ 为词 $t$ 在文档 $d$ 中的出现频率,$d_{\mathrm{f}}(t)$ 为词 $t$ 在文本集合中出现过的文本数目,$i_{\mathrm{df}}(t)$ 为词 $t$ 的逆文档频率,$N$ 为文档总数。

$$S(D,Q_t)=\sum_{i=1}^{n}i_{\mathrm{df}}(q_i)\frac{(k+1)f(q_i,D)}{f(q_i,D)+k(1-b+bL_D/L_{\mathrm{avg}})},\tag{4}$$

式中,$S(\cdot)$ 为 BM25 算法,$D$ 为文档,$Q_t$ 为查询词,$q_i$ 为 $Q_t$ 解析之后的一个语素,$f(q_i, D)$ 为 $q_i$ 在文档中的词项频率,$L_D$ 为文档长度,$L_{\mathrm{avg}}$ 为语料库全部文档的平均长度,$k$ 和 $b$ 为参数。

使用不同的文本相似度函数搜寻所需的无标记数据对最后的自训练有一定影响。例如,余弦相似度函数更多是从方向上区分差异;TF-IDF 从向量空间出发,统计词频值和逆文档频率,但没有考虑到词频上限问题;BM25 源于概率相关模型,与 TF-IDF 相比,设置了一个上限。因此,在考虑文本相似函数优缺点和命名实体识别文本数据的基础上,本研究结合余弦相似度函数和 BM25 算法,获得一个改进的文本相似度函数:

$$S_{\text{im}}=\frac{S(D,Q_t)\cos\theta(\boldsymbol{x},\boldsymbol{y})}{S(D,Q_t)+\cos\theta(\boldsymbol{x},\boldsymbol{y})},\tag{5}$$

所有文档相似度的平均值

$$S_{\text{avg}}=\frac{1}{n}\sum_{i=0}^{n}S_{\text{im}i}。\tag{6}$$

本研究将 BM25 和余弦相似度函数进行结合使用,原因在于:(1)余弦相似度是对两个向量的长度做归一化操作度量两个向量的方向,导致两个向量只要方向一致,在其余不同的情况下都会视为相似。(2)BM25 在 TF-IDF 的基础上进行改进,尽管克服了 TF-IDF 的部分缺点,但仍然无法解决处理词语语义相关性的问题。

因此,基于上述分析,本研究结合的文本相似度函数通过加权平均能够很好综合余弦相似度函数和 BM25 函数的优缺点。

## 3　试验设置及结果分析

### 3.1　数据集

结合自训练的命名实体识别方法,试验所用数据集来自公共数据集 MSRA[22]、法律实体识别数据集[23]和 CONLL03,其中法律数据集存在某些类别样本较少的问题,而 MSRA 和 CONLL03 数据集不存在此问题。通过 3 个典型数据集的试验对比可以看出本研究提出方法的有效性。数据样本的详细信息如表 1 所示。

表 1　数据集样本数量和标签数量统计信息

Table 1　Data set sample number and label number statistics

| 数据集 | 标签 | 训练集数量 | 测试集数量 |
|---|---|---|---|
| MSRA | LOC | 86 849 | 7 291 |
| | ORG | 103 261 | 6 977 |
| | PER | 51 738 | 5 824 |
| 法律数据集 | NASI | 2 587 | 527 |
| | NATS | 289 | 83 |
| | NCGV | 942 | 205 |
| | NCSM | 421 | 78 |
| | NCSP | 194 | 37 |
| | NHCS | 2 961 | 554 |
| | NHVI | 1 285 | 279 |
| | NO | 351 | 70 |
| | NS | 2 587 | 527 |
| CONLL03 | ORG | 12 117 | 2 496 |
| | LOC | 10 391 | 1 925 |
| | PER | 14 277 | 2 773 |
| | MISC | 5 861 | 918 |

### 3.2　评价指标

采用精准率 $P$、召回率 $R$ 和 $F_1$-socre 指标评测试验结果,计算公式为:

$$P=\frac{T_{\text{P}}}{T_{\text{P}}+F_{\text{P}}},\tag{7}$$

$$R=\frac{T_{\text{P}}}{T_{\text{P}}+F_{\text{N}}},\tag{8}$$

$$F_1=\frac{2PR}{P+R},\tag{9}$$

式中,$T_{\text{P}}$ 表示正类预测为正类的数量,$F_{\text{P}}$ 表示负类预测为正类的数量,$F_{\text{N}}$ 表示正类预测为负类的数量。

### 3.3　参数设置

不同超参数设置对模型性能产生较大的影响,本研究超参数设置如表 2 所示,BERT 使用的版本为 Roberta[24]。

表 2　超参数设置

Table 2　Hyperparameter setting

| 初始学习率 | Dropout | Batch_size | Epoch | 优化器 |
|---|---|---|---|---|
| $5\times10^{-5}$ | 0.4 | 32 | 10 | Adam |

### 3.4　试验及分析

为验证结合自训练的命名实体识别方法的有效性,本研究利用基线 BERT 模型与 BERT-ST 模型进行对比试验,试验结果如表 3 所示。

表 3　结合自训练的命名实体识别模型与基线模型的试验结果对比

Table 3　Comparison of experimental results between self-trained named entity recognition model and baseline model　　单位:%

| 数据集 | 模型 | 标签 | $P$ | $R$ | $F_1$ |
|---|---|---|---|---|---|
| MSRA | BERT | LOC | 96.67 | 94.79 | 95.72 |
| | | ORG | 87.14 | 92.95 | 89.95 |
| | | PER | 97.18 | 97.94 | 97.55 |
| | | 整体性能 | 94.40 | 95.08 | 94.74 |
| | BERT-ST | LOC | 96.74 | 95.28 | 96.00 |
| | | ORG | 87.87 | 92.80 | 90.27 |
| | | PER | 98.49 | 97.22 | 97.85 |
| | | 整体性能 | **95.35** | **95.74** | **95.54** |
| 法律数据集 | BERT | NASI | 81.03 | 87.58 | 84.17 |
| | | NATS | 79.14 | 77.46 | 78.29 |
| | | NCGV | 98.29 | 98.05 | 98.17 |
| | | NCSM | 91.03 | 86.84 | 88.89 |
| | | NCSP | 85.54 | 95.95 | 90.45 |
| | | NHCS | 93.98 | 99.07 | 96.45 |
| | | NHVI | 91.78 | 92.69 | 92.23 |
| | | NO | 86.81 | 89.29 | 88.03 |
| | | NS | 79.91 | 87.36 | 83.47 |
| | | NT | 92.42 | 95.87 | 94.12 |
| | | 整体性能 | 88.21 | 92.46 | 90.28 |

表 3(续)

| 数据集 | 模型 | 标签 | $P$ | $R$ | $F_1$ |
|---|---|---|---|---|---|
| 法律数据集 | BERT-ST | NASI | 83.04 | 90.00 | 86.38 |
| | | NATS | 79.47 | 84.51 | 81.91 |
| | | NCGV | 99.25 | 97.32 | 98.28 |
| | | NCSM | 88.34 | 94.74 | 91.43 |
| | | NCSP | 81.32 | 100.00 | 89.70 |
| | | NHCS | 94.46 | 98.69 | 96.53 |
| | | NHVI | 93.95 | 88.93 | 91.37 |
| | | NO | 86.49 | 91.43 | 88.89 |
| | | NS | 80.34 | 89.10 | 84.49 |
| | | NT | 94.63 | 96.86 | 95.73 |
| | | 整体性能 | **89.09** | **93.37** | **91.18** |
| CONLL03 | BERT | LOC | 91.72 | 92.87 | 92.20 |
| | | ORG | 91.89 | 92.93 | 92.41 |
| | | PER | 92.02 | 92.63 | 92.32 |
| | | MISC | 90.12 | 89.72 | 89.92 |
| | | 整体性能 | 92.10 | 91.87 | 91.98 |
| | BERT-ST | LOC | 92.08 | 92.12 | 92.10 |
| | | ORG | 92.12 | 93.32 | 92.71 |
| | | PER | 92.54 | 92.83 | 92.68 |
| | | MISC | 92.44 | 93.62 | 93.02 |
| | | 整体性能 | **93.02** | **92.64** | **92.83** |

由表 3 可以看出,在 BERT 预训练模型的基础上,结合自训练和改进文本相似度函数可以有效提高命名实体识别的性能。在法律数据集中,标签数据较少的 NATS(作案工具)和 NCSM(被盗货币)$F_1$ 分别提高了 3.62%和 2.54%,并且在其他标签类别较多的实体识别性能也略有提高;在 MSRA 数据集中,$F_1$ 总体提高了 0.8%;在 CONLL03 数据集中,$F_1$ 总体提高了 0.85%。这是由于命名实体识别数据集存在实体类别较少或不规则,导致模型对其特征学习能力较差;结合自训练的方法后,通过改进的文本相似度函数增强实体标签数量,在降低伪标记带来噪声的同时增强了模型对实体类别特征的学习能力。

模型结合自训练模型和改进的文本相似度函数,观察其在数据集上平均损失函数变化。MSRA 和 CONLL03 数据标签较为相似,因此只观察其 MSRA 和法律两个数据集平均损失函数变化,损失函数变化如图 2 所示。

由图 2 可知,相较于基线模型,结合自训练和改进的文本相似度函数的命名实体识别模型的损失函数收敛速度明显加快。这是由于通过增强实体标签数量,模型学习到更多实体类别特征,并且利用改进的文本相似度函数自动挑选无标记数据,可以有效降低不合适数据带来的噪声。

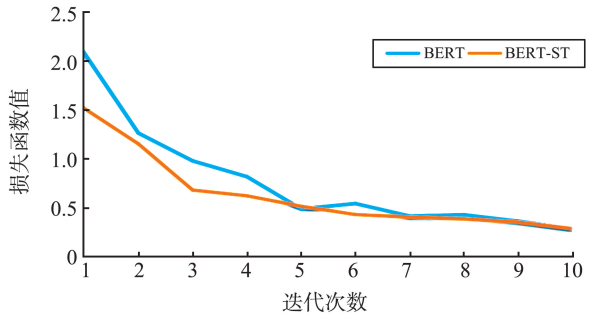


图 2 结合自训练方法的损失函数变化图

Fig.2 Loss function change diagram combined with self-training method

为进一步验证本研究方法的有效性,将 MSRA 和法律数据集的训练集按 10%、20%、40%、80%比例划分,观察在实体类别数量少的情况下,结合了自训练模型的命名实体识别方法性能的变化过程,如表 4 所示,测试集保持不变。

表 4 结合自训练方法在不同比例训练集下性能变化

Table 4 Combined with self-training method, the performance changes under different proportions of training sets

单位:%

| 数据集 | 比例 | 模型 | $P$ | $R$ | $F_1$ |
|---|---|---|---|---|---|
| MSRA | 10 | BERT | 62.34 | 43.08 | 50.95 |
| | | BERT-ST | **75.92** | **73.48** | **74.68** |
| | 20 | BERT | 70.52 | 71.28 | 70.89 |
| | | BERT-ST | **76.93** | **74.56** | **75.72** |
| | 40 | BERT | 83.41 | 86.78 | 85.06 |
| | | BERT-ST | **87.95** | **90.31** | **89.11** |
| | 80 | BERT | 90.12 | 91.54 | 90.82 |
| | | BERT-ST | **94.61** | **95.44** | **95.02** |
| 法律数据集 | 10 | BERT | 67.21 | 45.14 | 54.00 |
| | | BERT-ST | **75.13** | **76.12** | **75.62** |
| | 20 | BERT | 71.28 | 72.19 | 71.73 |
| | | BERT-ST | **77.11** | **76.87** | **76.98** |
| | 40 | BERT | 86.92 | 87.17 | 87.04 |
| | | BERT-ST | **87.20** | **88.74** | **87.96** |
| | 80 | BERT | 88.41 | 89.16 | 88.78 |
| | | BERT-ST | **89.02** | **90.47** | **89.73** |

从表 4 可以看出:在训练集比例很低的情况下,结合自训练的半监督数据增强方法的性能获得较大提升。其中,在训练样本数只有 10%的情况下提升最大;在 MSRA 中 $F_1$ 提升了 23.73%,在法律数据集中 $F_1$ 提升了 21.62%,这是由于在训练集样本特别少的情况下,模型不能充分学习实体类别特征。而通过数据增强的方法,为少样本带来充分数据,并且生成伪标记带来的适度噪声能提高模型的鲁棒性[18]。

最后,为验证改进的文本相似度函数的有效性,本研究与随机选择(random selection, RS)、余弦

相似度函数(COSINE)、TF-IDF 和 BM25 在法律数据集上作试验对比,试验结果如表 5 所示。

表 5　文本相似度函数试验对比

Table 5　Experimental comparison of text similarity function

单位:%

| 方法 | $P$ | $R$ | $F_1$ |
|---|---|---|---|
| RS | 88.62 | 92.74 | 90.63 |
| COSINE | 88.84 | 92.88 | 90.81 |
| TF-IDF | 87.31 | 92.63 | 89.89 |
| BM25 | 88.98 | 93.02 | 90.95 |
| 本研究 | **89.09** | **93.37** | **91.18** |

## 4　结语

针对命名实体识别数据集中个别实体类别样本较少导致模型学习特征能力较差的问题,本研究发现结合自训练模型的方法可以增强命名实体识别在少类别样本特征上的学习能力。通过改进文本相似度函数,从大量无标记数据库中自动挑选出最相似于原训练集的数据,减少不合适数据带来的噪声。试验结果表明,本研究提出的结合自训练的命名实体识别方法能够在多个数据集提高实体类别特征学习的能力。

该方法还有进一步改进空间,在未来的工作中,将在模型中加入去噪算法,改善伪标记给模型带来的影响。

**参考文献:**


[1] ZOPH B, GHIASI G, LIN T Y, et al. Rethinking pre-training and self-training[J]. Advances in Neural Information Processing Systems, 2020, 33: 3833-3845.

[2] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[J]. Advances in Neural Information Processing Systems, 2013, 26: 3113-3119.

[3] PENNINGTON J, SOCHER R, MANNING C D. Glove: global vectors for word representation[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA, USA: ACL, 2014: 1532-1543.

[4] LU J, YE M, TANG Z, et al. A novel method for Chinese named entity recognition based on character vector[C]//Proceedings of the International Conference on Collaborative Computing: Networking, Applications and Worksharing. Berlin, Germany: Springer, 2015: 141-150.

[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. Stroudsburg, PA, USA: ACL, 2016: 260-270.

[6] MA X, HOVY E. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF [C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: ACL, 2016: 1064-1074.

[7] CHIU J P C, NICHOLS E. Named entity recognition with bidirectional LSTM-CNNs [J]. Transactions of the Association for Computational Linguistics, 2016, 4: 357-370.

[8] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of NAACL-HLT 2019. Stroudsburg, PA, USA: ACL, 2019: 4171-4186.

[9] 孙弋, 梁兵涛. 基于 BERT 和多头注意力的中文命名实体识别方法[J]. 重庆邮电大学学报(自然科学版), 2023, 35(1): 110-118.
SUN Yi, LIANG Bingtao. Chinese named entity recognition method based on BERT and multi-head attention[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2023, 35(1): 110-118.

[10] 耿汝山, 陈艳平, 唐瑞雪, 等. 跨度语义增强的命名实体识别方法[J]. 西安交通大学学报, 2022, 56(7): 118-126.
GENG Rushan, CHEN Yanping, TANG Ruixue, et al. Named entity recognition approach with span semantic enhancement[J]. Academic Journal of Xi′an Jiaotong University, 2022, 56(7): 118-126.

[11] LI Xiaoya, FENG Jingrong, MENG Yuxian, et al. A unified MRC framework for named entity recognition [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: ACL, 2020: 5849-5859.

[12] 周裕林, 陈艳平, 黄瑞章, 等. 一种采用机器阅读理解模型的中文分词方法[J]. 西安交通大学学报, 2022, 56(8): 95-103.
ZHOU Yulin, CHEN Yanping, HUANG Ruizhang, et al. Machine reading comprehension for Chinese word segmentation[J]. Academic Journal of Xi′an Jiaotong University, 2022, 56(8): 95-103.

[13] YAO Huaxiu, ZHANG Chuxu, WEI Ying, et al. Graph few-shot learning via knowledge transfer[C]// Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park, CA, USA: AAAI, 2020: 6656-6663.

[14] ZHANG Jianhong, ZHANG Manli, LU Zhiwu, et al. AdarGCN: adaptive aggregation GCN for few-shot

learning[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway, NJ, USA: IEEE, 2021: 3482-3491.

[15] BAO Yujia, WU Menghua, CHANG Shiyu, et al. Few-shot text classification with distributional signatures[C]//Proceedings of ICLR 2020 Conference. California, NJ, USA: ICLR, 2020: 1-20.

[16] CHEN J, LIU Q, LIN H, et al. Few-shot named entity recognition with self-describing networks [EB/OL]. (2022-3-23) [2022-06-23]. https://arxiv.org/abs/2203.12252.

[17] NIU Yilin, JIAO Fangkai, ZHOU Mantong, et al. A self-training method for machine reading comprehension with soft evidence extraction[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: ACL, 2020: 3916-3927.

[18] XIE Q Z, LUONG M T, HOVY E, et al. Self-training with noisy student improves imagenet classification [C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2020: 10687-10698.

[19] MENG Yu, ZHANG Yunyi, HUANG Jiaxin, et al. Distantly-supervised named entity recognition with noise-robust learning and language model augmented self-training [C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: ACL, 2021: 10367-10378.

[20] WANG Y, MUKHERJEE S, LIU X, et al. LiST: lite prompted self-training makes parameter-efficient few-shot learners[C]//Proceedings of the Association for Computational Linguistics. Stroudsburg, PA, USA: ACL, 2022: 2262-2281.

[21] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York, USA: Curran Associates Inc, 2017: 6000-6010.

[22] LEVOW G A. The third international Chinese language processing bakeoff: word segmentation and named entity recognition[C]//Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing. Stroudsburg, PA, USA: ACL, 2006: 108-117.

[23] 最高人民法院司改办, 中国中文信息学会. 中国法律智能技术评测[EB/OL]. (2016-04-14)[2022-08-03]. http://cail.cipsc.org.cn/.

[24] LIU Y H, OTT M, GOYAL N, et al. Roberta: a robustly optimized BERT pretraining approach[EB/OL]. (2019-07-26)[2022-06-23]. https://arxiv.org/abs/1907.11692.

(编辑:李骏)

(上接第 95 页)

[17] JANNATA M S, SALAM R A, SUHENDI A. Study on the near-IR light detection and ranging (LiDAR) potential use as water level sensor[C]//IOP Conference Series: Earth and Environmental Science. Yogyakarta, Indonesia: IOP Publishing, 2021, 704(1):012-040.

[18] XU Nan, MA Yue, ZHANG Wenhao, et al. Monitoring annual changes of lake water levels and volumes over 1984-2018 using landsatimagery and ICESat-2 data[J]. Remote Sensing, 2020, 12(23):4004.

[19] MA Yue, XU Nan, SUN Jinyan, et al. Estimating water levels and volumes of lakes dated back to the 1980s using Landsat imagery and photon-counting LiDAR datasets [J]. Remote Sensing of Environment, 2019, 232(2):111287.

[20] NAWARAJ S, AARON R M, AARON R Y, et al. Groundwater level assessment and prediction in the Nebraska Sand Hills using LiDAR-derived lake water level [J]. Journal of Hydrology, 2021, 600:126582.

[21] PAUL Jonathan, BUYTAERT Wouter, SAH Neeraj. A technical evaluation of LiDAR-based measurement of river water levels[J]. Water Resources Research, 2020, 56(4):26810.

(编辑:郭少华)