

基于字符与单词嵌入的航空安全命名实体识别

孙安亮, 时宏伟, 王金策
(四川大学 计算机学院, 四川 成都 610000)

摘要:航空安全命名实体识别是构建航空安全知识图谱中基础且关键的任务,对消除航空隐患,制定有效的纠正措施和宏观政策提供了重要依据。针对航空安全领域包含大量较长的专有名词和名词缩写混合等问题,采用双向长短期记忆模型(BILSTM)、卷积神经网络(CNN)和条件随机场(CRF),构建一种使用字符与词两个粒度的模型,对航空安全事故进行命名实体识别(NER),以提取事故中的实体。采用航空事故报道为实验数据集,利用BILSTM模型自动学习字符粒度的语义特征向量,再结合词粒度的特征向量,通过CNN全局特征,最后通过CRF层对提取到的特征进行序列标注,以提取命名实体。经过实验对比验证,该模型能够有效提取命名实体,F1值相对现有方法提升了2.22%。实验结果表明,增加字符粒度的嵌入并且使用CNN获取全局特征可以有效提高航空安全领域命名实体识别效果。

关键词:命名实体识别;双向长短期记忆网络;卷积神经网络;条件随机场;航空安全

中图分类号:TP391.4

文献标识码:A

文章编号:1673-629X(2022)09-0148-06

doi:10.3969/j.issn.1673-629X.2022.09.023

Named Entity Recognition Based on Character and Word Embedding in Aviation Safety

SUN An-liang, SHI Hong-wei, WANG Jin-ce
(School of Computer, Sichuan University, Chengdu 610000, China)

Abstract:The identification of aviation safety nomenclature is a basic and critical task in the construction of aviation safety knowledge map, which provides an important basis for eliminating aviation hidden dangers and formulating effective corrective measures and macro policies. Aimed at locally sensitive problems such as the mixture of a large number of long proper nouns and noun abbreviations, we use bidirectional long-term short-term memory model (BILSTM), convolutional neural network (CNN) and conditional random field (CRF) to construct a model that uses two granularities of characters and words to perform named entity recognition (NER) for aviation safety accidents to extract the entity in the accident. The aviation accident report is used as the experimental data set, and the BILSTM model is used to learn the semantic feature vector of the character granularity. Combined with the feature vector of the word granularity, the global feature is obtained through CNN, and finally the extracted features are sequenced through the CRF layer to extract the name entity. After experimental comparison and verification, the model can effectively extract named entities and the F1 value is increased by 2.22% compared with the existing methods. Experimental results show that increasing the embedding of character granularity and using CNN to obtain global features can effectively improve the effect of named entity recognition in the aviation safety field.

Key words:named entity recognition; bidirectional long short-term memory; convolutional neural network; conditional random field; aviation safety

0 引言

近年来,随着航空出行数量的不断增加,促使航空安全领域需要更加有效的方法进行评估管理。为保护航空安全,管理部门制定了详细周密的工作流程。但在实际工作中,往往只能注重其中的一个部分,无法全面有效地处理整体情况,这就可能造成某个危险源在

不同系统逐渐扩散。同时航空安全系统会收集产生的大量详细数据,例如时间信息、气象条件信息、飞行器状态信息等。这些信息为消除航空隐患,制定有效的纠正措施和宏观政策提供了重要依据。这些数据中包含大量的知识概念并且具有相关性,但是大部分数据储存在分离的无结构文件中,无法有效地将这些数据

收稿日期:2021-08-30

修回日期:2021-12-30

基金项目:山西省青年科技研究基金(201801D221176);山西能源学院(SY-2018003)

作者简介:孙安亮(1996-),男,硕士研究生,研究方向为航空安全大数据分析、自然语言处理;通讯作者:时宏伟(1966-),男,硕士,教授,国务院政府特殊津贴获得者,研究方向为空管ATM/CNS领域、航空安全大数据分析、智能感知与认知、多源信息融合处理。

进行关联。如何有效使用已经存在的大量先验知识和历史记录,并将这些数据以及知识进行关联,构建系统化的关联事件模型,提取有效的知识来提高航空安全变得十分迫切。知识图谱^[1]可以提供一个较好的解决方案,首先通过信息抽取从异构数据源中自动抽取候选单元,该过程可以将分离存储的数据文件集合并获取实体、关系和属性单元信息;之后通过知识融合将候选单元进行连接合并,该过程可以将无结构文件结构化;最后通过知识加工将整体信息网络化,该过程将整体数据系统化关联。知识图谱在航空安全领域可将复杂的安全因子关系可视化,为有效分析潜在的安全风险提高助力。

命名实体识别是构建知识图谱中基础且关键的任务,实体识别的质量对知识图谱的构建效率和质量影响极大。而航空安全命名实体识别主要特点在于航空安全领域包含大量较长的专有名词和名词缩写混合的情况,例如起降机场、飞机类型和具有专有缩写的组织单位等,加大了命名实体识别的难度。另外在航空安全实体识别中会出现多种描述指向同一实体的情况,例如引擎故障可以描述为左引擎故障,右引擎故障,1号引擎故障,动力减弱,动力丧失等。这导致在航空安全命名实体识别对局部特征更加敏感。这些特殊性使得一般命名实体识别模型在航空安全领域表现效果不佳,需要构建一个适用于航空安全领域的命名实体识别方法,为航空安全领域知识图谱的自动构建提供更好的基础。该文提出使用 BILSTM 模型进行字符级别的特征提取,该方式可以较好地克服航空安全领域中专有名词较长以及名词缩写的局限。同时在联合特征提取使用 CNN 模型而非 LSTM 模型,可以较好地处理航空安全领域数据对于局部特征敏感的情况。

1 相关方法

命名实体识别(Named Entity Recognition, NER)是自然语言处理(Natural Language Processing, NLP)中一项基本性的任务,主要任务是识别句子中表达特殊实体的词语或者短句,比如表示人物、组织、地点等^[2]。同时 NER 也是信息提取、文本理解^[3]、信息检索^[4-5]、自动文本摘要^[6]、问题回答^[7-8]、机器翻译^[9]、知识图谱构建^[10]等任务的基础工具。命名实体识别主要包括两个部分,分别是确定实体边界和识别实体的类别,对于一个实体的正确识别需要确保该实体的边界和类型都正确。航空安全命名实体识别主要目的在于识别航空安全实体,例如航班类型、起降地、事故原因等实体。相较于通用命名实体识别任务,航空安全命名实体识别的难点在于任务实体类别增多、专有名词较多、实体全称与缩写混合表示以及相同实体多

种描述。早期的命名实体识别方法较多基于字典和规则的方法^[2],因为需要专家知识,识别效果较差。之后随着机器学习的发展,命名实体识别逐步发展为使用机器学习方法,比如 Hidden Markov Model (HMM)^[11]和 Conditional Random Field (CRF)^[12]。也有将机器学习与字典规则方法结合取得了较好的效果,但是这些方法较大地依赖人工选择数据特征无法广泛推广^[13]。随着深度学习在各个领域的成功应用,不断吸引着研究者的兴趣。在短短几年的时间里,一系列关于深度学习应用的研究在 NER 领域展开,并且取得了较好的效果。从 Collobert 等^[14]首次将 CNN 应用在命名实体识别任务中,通过 CNN 获取局部特征后通过池化操作获取全局特征,到张聪品等^[15]使用 BILSTM-CRF 模型,使用 BILSTM 高效利用上下文信息,使用 CRF 利用句子级别的标签信息,通过 BILSTM 与 CRF 的结合使得模型整体具有更好的鲁棒性。

航空领域命名实体识别目前仍以 BILSTM-CRF 模型为主,王红等^[16]提出 BILSTM-CRF 用于民航突发事件实体识别,Cheng 等^[17]提出 BILSTM-CRF 用于航空安全命名实体识别任务。但因为航空领域的特殊性,Bao 等^[18]提出 AR-BILSTM-CRF 用于航空设计领域,通过增加注意力机制解决航空设计领域包含较多专有名词的问题,并取得了较好的结果。考虑到航空安全领域包含较多专有名词,并且专有名词和缩写表示的情况,同时受 Kuru 等^[19]工作的启发,字符特征嵌入可以较好地解决单词超出词典范围的问题。该文考虑在航空安全命名实体识别任务中增加字符特征,解决专有名词以及缩写的问题。之后通过 Chen 等^[20]工作的启发,考虑到航空安全领域对于局部特征敏感的情况,在获取字符特征与单词特征后通过 CNN 获取联合特征。最后通过 CRF 进行序列标注,提取命名实体。经过对比实验,提出的模型具有较好的结果,相比 BILSTM-CRF 的表现更佳。

2 模型建立

为了提取出航空安全领域命名实体,适应航空安全领域数据特点,即包含专有名词以及缩写表示,并且对于局部敏感的情况,该文提出了基于字符和单词多粒度嵌入模型。模型结构如图 1 所示。

模型的理论基础是利用 BILSTM 对字符进行特征获取,因为 BILSTM 可以有效利用上下文特征的字符向量,在字符特征提取中使用 BILSTM 可以有效适用航空安全领域专有名词及缩写表示的情况。之后将联合单词嵌入向量放入联合特征获取层中,通过 CNN 卷积得到命名实体的联合特征,因为 CNN 可以更好地感知局部特征关系,更好适用于航空安全领域局部特征

敏感的情况。之后将其放入 CRF 层进行序列标注。该文相较于 BILSTM-CRF 模型的改进在于增加字符特征解决航空安全领域专有名词及缩写表示的问题，

并使用 CNN 获取航空安全命名实体的局部特征, 提高在航空安全领域命名实体识别的效果。

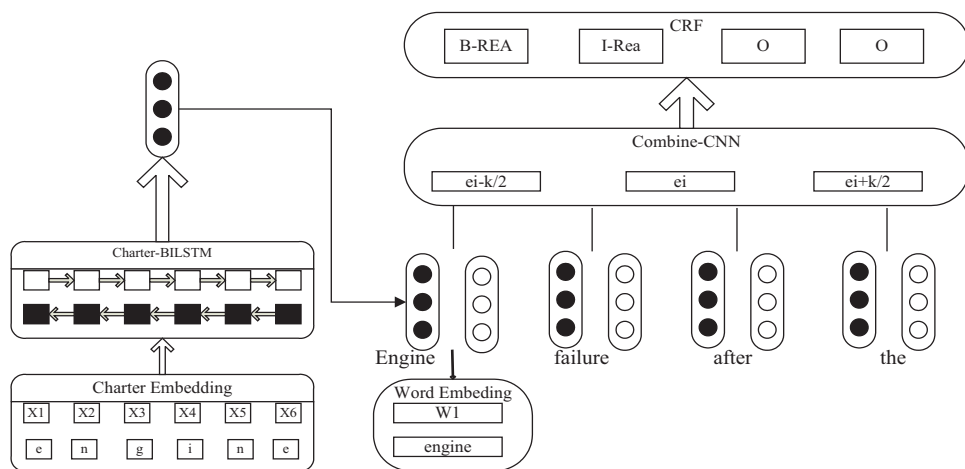


图 1 模型结构

首先定义一下将要使用的字符表达, 设 $s = \{w_1, w_2, \dots, w_n\}$ 表示一个长度为 n 的句子, 句子第 i 个单词记为 w_i , 其向量表示记为 w_i 。 $w_i = \{x_1, x_2, \dots, x_k\}$ 表示字符长度为 k 的单词, 单词中第 k 个字符记为 x_i , 其向量表示记为 x_i 。接下来将详细介绍该模型中的字符层、联合层和 CRF 层的内容。

2.1 字符层

字符层的主要目的在于获取字符特征, 将字符通过字符嵌入得到字符的向量表示, 记为 x_i 。之后使用

字符层模型得到特征表示, 记为 c_i 。字符层使用 BILSTM 作为特征提取方法, 双向 LSTM 由正向 LSTM 和逆向 LSTM 组成, 在结果输出时会两个相反方向的输出结果拼接作为最终的输出结果。LSTM 是一种特殊的循环神经网络 (Recurrent Neural Network, RNN), 不仅仅具有 RNN 处理序列问题的优势, 还通过引入门控机制缓解了 RNN 梯度易爆炸或消失的问题, LSTM 单元结构如图 2 所示。

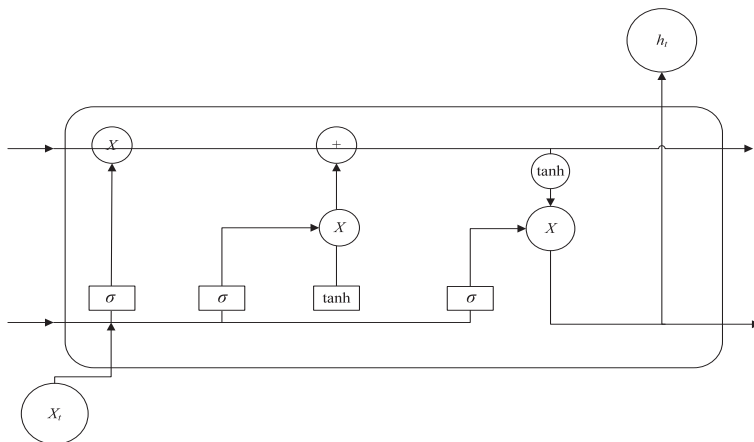


图 2 LSTM 结构

LSTM 中的输入 x_t 代表 t 时刻的输入, 即 t 时刻的字符向量。 h_{t-1} 表示上一时刻的隐层状态, c_{t-1} 表示上一时刻的单元状态, f_t 表示 t 时刻的单元遗忘状态。LSTM 由遗忘门、输入门和输出门组成。

遗忘门的目的在于决定丢弃的信息, 将读取的 h_{t-1} 和 x_t 进行计算得到单元的遗忘状态 f_t , 用于后续的状态更新。计算方式如式(1)所示:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (1)$$

其中, σ 表示激活函数, 而 W_f 和 b_f 则表示遗忘门参数

在模型的训练过程中学习得到。

输入门则是决定加入的新信息, 输入门的实现需要两个步骤, 首先是根据输入信息计算记忆状态 i_t , 得到单元记忆状态信息, 并且使用 \tanh 计算 \tilde{c}_t 作为临时单元状态。之后通过遗忘状态计算当前单元状态更新信息 c_t 。计算方式如式(2)所示:

$$\begin{cases} i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \\ \tilde{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \\ c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \end{cases} \quad (2)$$

其中, W_i, b_i, W_c, b_c 表示输入门计算参数在模型训练过程中学习得到, $*$ 表示向量点乘操作。

输出门则是决定输出的信息, 单元的输出由单元状态决定, 输出门首先计算单元输出状态 o_t , 之后还要更新 t 时刻的单元隐层状态 h_t 。计算方式如式(3)所示:

$$\begin{cases} o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \\ h_t = o_t * \tanh(c_t) \end{cases} \quad (3)$$

其中, W_o, b_o 表示输出门计算参数同样在训练中学习得到, $*$ 表示向量点乘操作。

循环单元计算得到 LSTM 的计算结果, BILSTM 则是将输入的字符向量 x_i 通过两个不同方向的 LSTM, 最后将两个方向的输出 $\vec{c}_i, \overleftarrow{c}_i$ 串联得到字符层的特征向量, 记为 $c_i = \{\vec{c}_i, \overleftarrow{c}_i\}$ 。

根据计算过程分析得到, 字符会通过 BILSTM 更好地获取上下文的信息提取字符特征。对于较长的专有名词更好获取相对依赖关系, 对于缩写混合情况通

过字符特征也可以有效处理, 这样可以很好地解决专有名词缩写混合的问题。当单词内字符通过字符特征提取后, 对应单词通过单词嵌入得到单词的向量表示, 并将字符特征向量 c_i 与单词特征向量 w_i 串联, 得到最终的嵌入特征 $e_i = \{c_i, w_i\}$ 。

2.2 联合层

联合层将字符和单词串联的特征向量 e_i 放入 CNN 卷积神经网络进行特征抽取。字符与单词向量的语境特征存在很强的依赖关系, 对命名实体识别任务具有很大的参考意义。特别对航空安全命名实体识别任务, 因为数据的专业性和缩写的存在, 句子中的依赖关系对于最终标签的预测起着更为重要的作用。同时针对航空安全数据的特点, 以及数据标识类别的特点, 发现航空安全数据的标注任务对于局部依赖关系特别敏感, 对于全局依赖则较为迟缓。基于 CNN 对于局部特征提取的优势, 该文在联合层使用 CNN 结构进行特征提取。联合层模型结构如图 3 所示。

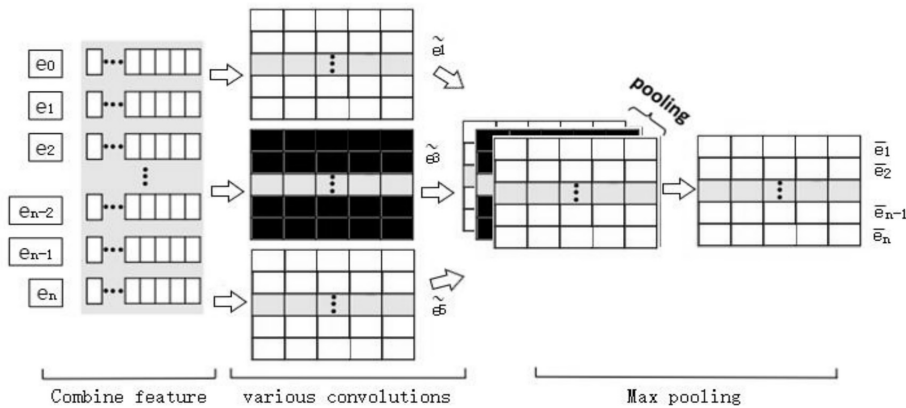


图 3 联合模型结构

联合层主要将获得的联合向量 e_i 通过卷积层进行特征提取, 考虑到不同局部特征的影响, 该文使用多个尺寸卷积进行局部特征提取, 之后通过池化操作获得全局特征。其计算过程如式(4)所示:

$$\begin{cases} \vec{e}_i^k = \text{conv}_k([e_{i-k/2}, \dots, e_i, \dots, e_{i+k/2}]) \\ \overleftarrow{e}_i = \text{pooling}([\tanh(\vec{e}_i^1), \tanh(\vec{e}_i^3), \tanh(\vec{e}_i^5)]) \end{cases} \quad (4)$$

其中, e_i 表示联合特征向量, conv_k 表示进行卷积运算且卷积核为 k , pooling 表示池化操作, \tanh 表示非线性函数计算, \vec{e}_i^1, \vec{e}_i^3 和 \vec{e}_i^5 分别表示卷积核尺寸为 1、3、5。在运算过程中使用 0 进行填充操作以确保每个联合向量都可以得到对应的依赖特征, 将最后得到的结果标记为 \vec{e}_i 。联合层在进行卷积操作之后, 还会将得到的结果 \vec{e}_i 使用非线性 \tanh 函数进行处理, 作为最终预测特征 p_i , 计算方式如式(5)所示:

$$p_i = \tanh(\vec{e}_i) \quad (5)$$

2.3 CRF

对于标签依赖关系的建模在 NER 任务中是很重要的, 因此最后使用 CRF 建模标签之间的依赖关系。

对一个给定的语句 $s = \{w_1, w_2, \dots, w_n\}$ 和它对应的序列标签 $y = \{y_1, y_2, \dots, y_n\}$, 用 $\gamma(s)$ 标记 s 中的所有可能标签序列。CRF 模型首先在 y 给出的所有可能的标签序列上计算条件概率 $p(y|s)$, 计算方式如式(6)所示:

$$p(y|s) = \frac{\prod_{i=1}^n \varnothing(y_{i-1}, y_i, s)}{\sum_{y=\gamma(s)} \prod_{i=1}^n \varnothing(y_{i-1}, y_i, s)} \quad (6)$$

其中, $\varnothing(y_{i-1}, y_i, s) = \exp(f(s_i, y', y))$, f 是一个映射函数, 主要目的是将单词映射到标签, 函数的表达如式(7)所示:

$$f(s_i, y', y) = W_y p_i + b_{y,y} \quad (7)$$

其中, p_i 是联合层输出经过非线性函数得到的最终预

测特征, W_y 是预测权值矩阵, $b_{y',y}$ 是从 y' 到 y 的过渡权重。对于 CRF 模型而言,其损失函数的计算方式如式(8)所示:

$$L = - \sum_s \log p(y | s) \quad (8)$$

最终的目标就是最大化式(6),即 $\arg \max_{y \in \gamma(s)} p(y | s)$,为了便捷和高效,将用 Viterbi 算法进行解码。

3 实验与结果分析

为验证该文提出模型的效果,从 kaggle 上获取从 1908 年全球飞机事故报道,进行处理后形成实验数据进行实验对比。首先将无字符特征嵌入的通用 LSTM-CRF 模型作为基准,以此来了解通用模型在航空安全命名实体识别任务的表现,之后又对比了使用 LSTM 和 CNN 对字符进行特征提取对结果的影响,最后对比使用不同模型在联合层对特征提取的不同影响。

3.1 数据预处理与标注

本次实验使用了从 1908 年到 1973 年之间的航空事故报道,共 3 007 篇,先对数据进行格式处理,并初步清理对 NER 任务的无效数据。例如数据中的时间、登记标号、载客数等。航空事故报道以英语为主,还包括法语、德语等其他语言,因为表达形式的原因,该文没有对语言进行详细分类,即模型本身的训练数据是由多语言集成的。对数据进行初步处理之后,使用程序对数据集中的标准格式部分进行标注处理,用人工标注的方式对其他部分进行标注。对于数据集的标注,该文参考航空事故数据特点与注意事项,并结合其他 NER 数据集的标注方式,将数据的标注分为八个类别,具体类别信息如表 1 所示。

表 1 标准释义

标注类别	说明	标注类别	说明
LOC	地点	CITY	城市
STATE	国家	ORG	组织
CAT	飞行类别	PEO	人物
TYPE	型号	REASON	原因

数据标注按照标准释义表规则进行,并在标注完成后重新检查保证数据标注的质量。数据集的标注策略采用 BIOS 标注策略,其中 B 表示短语的开始部分, I 表示短语的剩余部分, S 表示单个词, O 表示无实体含义的词。数据规模达到 10 万个词,数据统计如表 2 所示。

表 2 数据集统计

数据集	训练数据	测试数据
报告总数	2 125	882
单词总数	70 199	29 962
实体数量	14 230	5 892

将数据集划分为训练集和测试集两个部分,其中训练集由 7 万个词组成,测试集由 3 万个词组成。数据划分的方式在时间有序的情况下进行简单抽样处理,保证训练集与测试集数据的分布均匀。

3.2 实验设置

该模型基于 Pytorch 框架搭建,实验参数基于对比实验进行选择,最终实验参数设置:字符嵌入维度为 30,词嵌入维度为 100,双向 LSTM 将词嵌入维度扩展到 200,卷积核尺寸分别设置为 1、3、5。为防止模型过拟合,将 dropout 设置为 0.6,数据集的 batch_size 设置为 5,训练学习率设置为 0.001,训练优化使用 Adam,模型训练轮次设置为 100。实验采用 F1 值作为模型效果评价指标,NER 任务中常用的评价指标有准确率 P 、召回率 R 和 F1 值,计算公式如式 9 所示:

$$\begin{aligned} P &= \frac{TP}{TP + FP} \\ R &= \frac{TP}{TP + FN} \\ F1 &= \frac{2PR}{P + R} \end{aligned} \quad (9)$$

其中,TP 表示正类预测为正类数,FP 表示负类预测为正类数,FN 表示正类预测为负类数。考虑到 F1 值是准确率和召回率的调和平均数,使用 F1 值作为衡量指标可以更好地兼顾准确率和召回率。

对比实验分为三个部分。第一部分是使用 BILSTM-CRF 作为基线方法,该方法过程为将单词通过嵌入层转化为向量表示后送入 BILSTM 进行特征提取,之后使用 CRF 解码做序列化标注。基线方法没有考虑字符特征,并且在特征提取时使用 LSTM 的上下文感知依赖关系。第二部分是对字符特征使用 LSTM 和 CNN 两个模型,对比不同模型提取字符特征的效果以及增加字符特征对于航空安全命名实体识别的作用。第三部分是对联合特征使用 LSTM 和 CNN 两个模型,对比联合特征提取对命名实体识别的影响,验证对航空安全命名实体识别局部敏感的处理效果。

3.3 结果分析

实验结果对比使用 F1 值作为标准,实验结果如表 3 所示。

表 3 实验结果

模型	F1 值/%
BILSTM-CRF	81.51
CNN-CNN-CRF	81.84
BILSTM-CNN-CRF	83.82
BILSTM-BILSTM-CRF	83.73
CNN-BILSTM-CRF	81.69

首先将基线方法实验结果与对字符特征使用 LSTM 和 CNN 的实验结果进行对比可以发现,字符特

征的加入可以提高命名实体识别任务的效果,采用 LSTM 和 CNN 在 F1 值上分别提升 2.22 和 0.18,结果表明增加字符特征可以提高在航空安全领域命名实体识别的效果,并且使用 LSTM 可以更好地获取上下文依赖关系,在标注结果上可以更好地解决缩写混合的问题。在字符特征采用相同方法进行特征提取,对比联合特征使用 LSTM 和 CNN 的实验结果发现,使用 CNN 获取联合特征效果较好。在字符特征使用 CNN 处理时,联合特征使用 CNN 比 LSTM 效果提高 0.09。在字符特征使用 LSTM 处理时,联合特征使用 CNN 比 LSTM 效果提高 0.15。同时使用 CNN 的收敛速度和训练速度要优于 LSTM。原因主要在于 CNN 模型可以更好地并行运算。

4 结束语

增加字符特征作为命名实体识别模型的特征输入,解决航空安全领域专有名词较长与缩写混合问题,使用 CNN 进行联合特征处理,解决航空安全领域局部敏感问题。并在 1908 年到 1973 年航空安全事故报道上进行了实验。从实验结果来看,航空安全领域命名实体识别任务因为其领域特点,字符特征的加入可以更好地提升航空安全领域命名实体识别的任务效果。在字符与单词特征联合的处理上使用 CNN 的效果更佳,同时 CNN 的训练速度与收敛速度也要优于 LSTM。该方法的主要问题在于数据集标注耗时较长,考虑到航空安全领域数据标注成本较高的问题,后续的研究方向可以结合小样本学习方法减轻对于数据量的依赖情况。同时后续研究发展航空安全领域更细粒度的命名实体识别方法,获取更加详细的实体信息。考虑到注意力机制对于命名实体识别任务的提升效果,后续研究也可以使用注意力机制提高模型的效果。

参考文献:

[1] 刘 峤,李 杨,段 宏,等.知识图谱构建技术综述[J]. 计算机研究与发展,2016,53(3):582-600.

[2] NADEAU D,SEKINE S. A survey of named entity recognition and classification[J]. *Lingvisticae Investigationes*,2007,30(1):3-26.

[3] ZHANG Z,HAN X,LIU Z,et al. ERNIE: enhanced language representation with informative entities [C]//ACL2019. Florence: Association for Computational Linguistics, 2019: 1441-1451.

[4] GUO J,XU G,CHENG X,et al. Named entity recognition in query[C]//ACM SIGIR conference on research and development in information retrieval. NY: Association for Computing Machinery,2009:267-274.

[5] PETKOVA D,CROFT W B,PETKOVA P. Proximity-based

document representation for named entity retrieval [C]//ACM conference on information and knowledge management. NY: Association for Computing Machinery,2007:731-740.

[6] MCDONALD D,CHEN H. Using sentence-selection heuristics to rank text segments in TXTRACTOR [C]//ACM/IEEE-CS joint conference on digital libraries. NY: Association for Computing Machinery,2002:28-35.

[7] 鲍静益,于佳卉,徐 宁,等. 问答系统命名实体识别改进方法研究[J]. 数据采集与处理,2020,35(5):930-941.

[8] MOLLÁ D,VAN ZAAANEN M,SMITH D. Named entity recognition for question answering [C]//Proceedings of the 2006 Australasian language technology workshop. Sydney: Australasian Language Technology Workshop,2006:51-58.

[9] HARTLEY B. Improving machine translation quality with automatic named entity recognition [C]//Proceedings of the 7th international EAMT workshop on MT and other language technology tools, improving mt through other language technology tools: resources and tools for building MT. [s. l.]: Association for Computational Linguistics,2003:1-8.

[10] 徐增林,盛泳潘,贺丽荣,等. 知识图谱技术综述[J]. 电子科技大学学报,2016,45(4):589-606.

[11] 韩 普,姜 杰. HMM 在自然语言处理领域中的应用研究[J]. 计算机技术与发展,2010,20(2):245-248.

[12] 邱 莎,阿 圆,王付艳,等. 基于统计的中文地名自动识别研究[J]. 计算机技术与发展,2011,21(11):35-38.

[13] LI J,SUN A,HAN J,et al. A survey on deep learning for named entity recognition [J]. arXiv:1812.09449v3,2018.

[14] COLLOBERT R,WESTON J,BOTTOU L,et al. Natural language processing (almost) from scratch [J]. *Journal of Machine Learning Research*,2011,12(1):2493-2537.

[15] 张聪品,方 滔,刘昱良. 基于 LSTM-CRF 命名实体识别技术的研究与应用[J]. 计算机技术与发展,2019,29(2):106-108.

[16] 王 红,李浩飞,邸 帅. 民航突发事件实体识别方法研究 [J]. 计算机应用与软件,2020,37(3):166-172.

[17] CHENG Y,JIAO Y,WEI W,et al. Research on construction method of knowledge graph in the civil aviation security field [C]//IEEE international conference on civil aviation safety and information technology. Kongming: IEEE, 2019: 556-559.

[18] BAO Y,AN Y,CHENG Z,et al. Named Entity recognition in aircraft design field based on deep learning [C]//Web information systems and applications. Guangzhou: Springer, 2020:333-340.

[19] KURU O,CAN O A,YURET D. CharNER: character-level named entity recognition [C]//International conference on computational linguistics. Osaka: COLING,2016:911-921.

[20] CHEN H,LIN Z,DING G,et al. GRN: gated relation network to enhance convolutional neural network for named entity recognition [C]//Proceedings of the AAAI conference on artificial intelligence. [s. l.]: AAAI,2019:6236-6243.