

中文引用格式:陈述,张超,陈云,等. 基于命名实体识别的水电工程施工安全规范实体识别模型[J]. 中国安全科学学报, 2024, 34(9): 19-26.

英文引用格式:CHEN Shu, ZHANG Chao, CHEN Yun, et al. Model of identifying entities of safety specification for hydropower engineering construction[J]. China Safety Science Journal, 2024, 34(9): 19-26.

基于命名实体识别的水电工程施工安全规范 实体识别模型*

陈述^{1,2}教授, 张超², 陈云^{**1,2}副教授, 张光飞³高级工程师, 李智³正高级工程师
(1 三峡大学 水电工程施工与管理湖北省重点实验室, 湖北 宜昌 443002; 2 三峡大学
水利与环境学院, 湖北 宜昌 443002; 3 中国长江三峡集团有限公司, 湖北 武汉 430010)

中图分类号: X948 文献标志码: A DOI: 10.16265/j.cnki.issn1003-3033.2024.09.0008

基金项目: 国家自然科学基金资助(52479127, 52079073, 52209163)。

【摘要】 为准确识别水电工程施工安全规范实体, 通过预训练模型中双向编码器表征法(BERT)挖掘文本中丰富的语义信息, 利用双向长短期记忆神经网络(BILSTM)提取规范实体语义特征, 依靠条件随机场(CRF)分析实体之间的依赖关系, 构建水电工程施工安全规范的命名实体识别模型; 以《水利水电工程施工安全防护技术规范》(SL714—2015)为例, 计算命名实体识别模型精确率。结果表明: BERT-BILSTM-CRF模型准确率为94.35%, 相比于3种传统方法, 准确率显著提高。研究成果有助于水电工程施工安全规范知识智能管理, 为施工安全隐患智能判别提供支撑。

【关键词】 命名实体识别; 水电工程施工; 安全规范; 双向编码器表征法(BERT); 双向长短期记忆神经网络(BILSTM); 条件随机场(CRF)

Model of identifying entities of safety specification for hydropower engineering construction

CHEN Shu^{1,2}, ZHANG Chao², CHEN Yun^{**1,2}, ZHANG Guangfei³, LI Zhi³

(1 Hubei Key Laboratory of Hydropower Engineering Construction and Management, China Three Gorges University, Yichang Hubei 443002, China; 2 College of Hydraulic & Environmental Engineering, China Three Gorges University, Yichang Hubei 443002, China;
3 China Three Gorges Corporation, Wuhan Hubei 430010, China)

Abstract: To accurately identify the entities of hydropower engineering construction safety specification, the named entity recognition model of hydropower engineering construction safety specification was constructed. The rich semantic information in the text was mined by the BERT. The semantic features of the specification were extracted by using BILSTM. The dependency relationship between entities was analyzed by relying on CRFs. The *Technical Specification for Safety Protection in Construction of Water Conservancy and Hydropower Projects* (SL714-2015) was taken as an example to calculate the named entity recognition model accuracy rate. The results show that the accuracy rate of the BERT-BILSTM-CRF

* 文章编号: 1003-3033(2024)09-0019-08; 收稿日期: 2024-03-15; 修稿日期: 2024-06-20

** 通信作者: 陈云(1993—), 男, 湖北枝江人, 博士, 副教授, 主要从事安全管理研究。E-mail: yunchen@ctgu.edu.cn。

model is 94.21%. Compared with the three traditional methods, the accuracy is significantly improved. The research will effectively assist in the intelligent management of safety regulations knowledge for hydropower engineering construction, and provide important support for the intelligent identification of construction safety hazards.

Keywords: named entity identification; hydropower engineering construction; safety specification; bidirectional encoder representation from transformers (BERT); bi-directional long and short-term memory neural network (BILSTM); conditional random field (CRF)

0 引言

水电工程施工安全隐患多而复杂^[1-2],人工排除隐患耗时费力,且准确性存在主观依赖,基于大数据的智能化检测是必然趋势^[3]。水电工程施工安全规范是安全文本大数据的基础,目前,水电工程施工安全规范是由各种实体词按照中文语法组成的非结构化文本,关键信息难以直接挖掘。通过命名实体识别技术获取水电工程施工安全规范关键实体信息,可有效助力水电工程施工安全规范知识智能管理,为构建水电工程施工安全知识图谱、智能判别施工安全隐患提供重要支撑。

命名实体识别是给定非结构文本后,从文本句子中寻找、识别和分类相关实体,已成为人工智能领域重要的研究方向。1995年,张小衡等^[4]采用人工规则的方法识别与分析中文机构名称,开拓了中文命名实体识别的先河。1999年,BIKEL等^[5]提出基于隐马尔可夫模型(Hidden Markov Models, HMM)的英文命名实体识别方法,通过机器语言规则识别出英文地名、人名。ZHANG Qiqi等^[6]引入条件随机场(Conditional Random Fields, CRF),识别命名实体任务,弥补了HMM模型只能匹配单个实体的缺点,增强了命名实体识别能力。TAEKHYUNG等^[7]以建筑事故文本为模型数据,建立了以命名实体识别为基础的建筑事故案例知识管理系统。李明超等^[8]将命名实体识别中的word2vec技术运用到水电工程领域,提取水电工程领域专业词,然而,工程建设领域由于环境因素复杂,识别文本关键词精度较低,导致匹配实体出现边界模糊的问题。刘婷等^[9]采用预训练模型中的双向编码器表征法(Bidirectional Encoder Representation from Transformers, BERT)生成水利事故文本的动态特征向量,并通过大量预训练提高了文本特征识别的准确率。杨秀璋等^[10]引入情感词典优化特征词的权重,进一步明确了实体的匹配边界。易明等^[11]基于深度学习的文本语义理解和挖掘,使用双向长短期

记忆神经网络(Bi-directional Long Short-Term Memory, BILSTM)提取语义特征,构建实体信息库的自动化分类预测,能有效结合文本上下文信息提取特定领域实体。上述研究大多采用传统的命名实体识别方法,在抽取水电工程施工安全领域实体词过程中容易忽视实体词之间的联系,识别实体词效果不佳,导致缺乏准确性^[12]。

鉴于此,笔者拟采用基于命名实体识别深度学习的方法,通过预训练模型BERT作大规模语料预训练,利用BILSTM模型提取施工安全规范实体的语义特征,依靠CRF模型增加实体之间的约束条件,精准抽取实体,建立水电工程施工安全规范实体识别模型,以期为智能化地安全隐患排查提供参考。

1 命名实体识别模型

研究框架如图1所示。主要研究步骤如下:
①预处理文本数据。收集水电工程施工安全相关规范,去除无用符号、句子隔断,过滤词语,进行分词处理。采用BIO(Begin-Inside-Outside)标注方法实现施工文本实体标注,其中,B表示实体的起始位置,I表示实体的中间与结尾,O表示非实体。并将实体分为施工区域(Place, PLA)、施工设备(Facility, FAC)、施工行为(Action, ACT)、施工人员(People, PEO)等4类。
②预训练文本语库。采用BERT预训练模型,经过隐藏层H,获取水电工程施工安全规范文本对应的词向量,通过大规模文本语料训练,建立水电工程施工安全规范文本的实体特征语库。
③提取语义特征。使用BILSTM双向神经网络,将水电工程施工安全规范分成若干条进行正反向处理输入序列,将2个LSTM的输出拼接起来,深度学习上下文语义特征,提取实体的语义特征,进一步筛选实体特征语库。
④分析实体依赖关系。通过CRF计算实体关系最优解,分析水电工程施工安全规范实体之间的依赖关系,获取实体之间的约束。
⑤输出结果。通过耦合模型BERT-BILSTM-CRF训练,计算命名实体识别准确率。

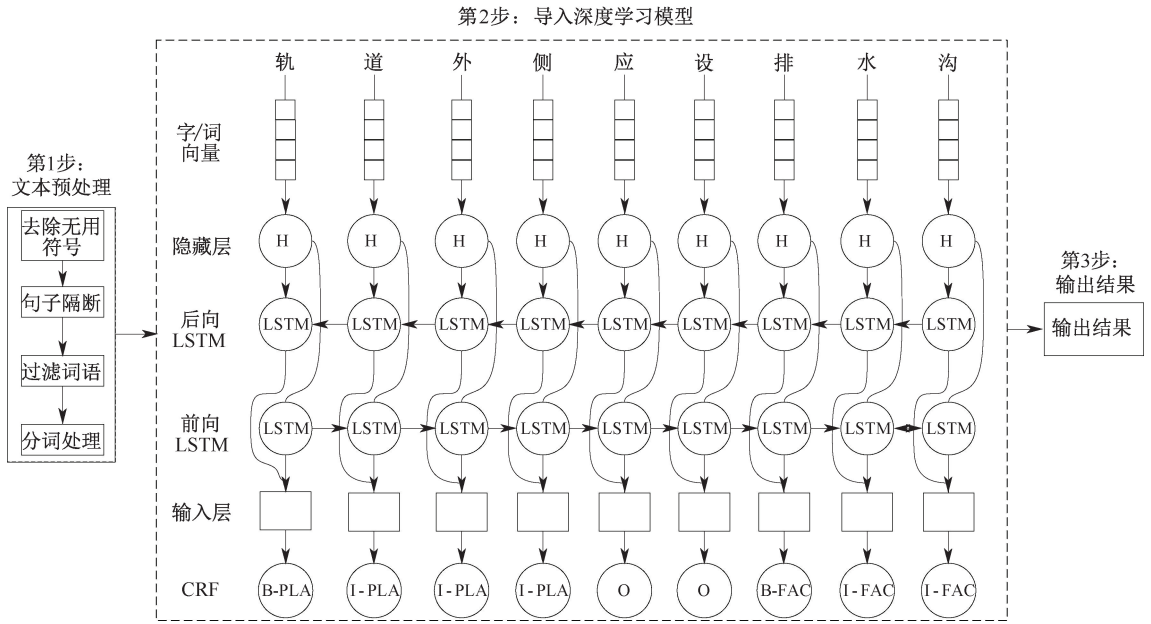


图 1 命名实体识别框架

Fig. 1 Framework of named entity recognition

1.1 数据预处理

收集水电工程施工安全规范文本数据, 建立水电工程施工安全领域的停用词表, 去除各项停用词以及特殊符号, 隔断文本语句, 提取图表信息。采用 BIO 注方法^[13], 对处理后的规范文本数据逐一进行实体标注, 获得带标注的单个字符。

1.2 基于 BERT 的字符向量化

中文词语存在一词多义, 在不同句子表示不同含义^[14]。识别水电工程施工安全规范文本实体需要大量语料库, 才能获取精确的信息表达。为此, 利用大规模无标注语料 BERT 模型训练, 获得丰富语义信息, 建立水电工程施工安全规范的实体特征语库^[15-16]。

以“坠落物”为例进行文本输入, 以 [CLS] 表示命名语句的开始, [SEP] 表示命名识别语句的结束与分隔, 中间部分用句子中的单个字符表示。位置向量从 1 开始, 表示文本中每个字符的位置信息。在 BERT 模块中, 文本输入以单个字符表示。输出文本的字符以 R_1, R_2, \dots, R_N 量化表示, 其中, N 为输出文本字符数量, 经过若干次训练, 建立水电工程施工安全规范的实体特征语库。

1.3 BiLSTM 的语义特征提取

水电工程施工安全规范上下文环境较为复杂, 传统的循环神经网络无法保留与结尾位置相隔较远的上下文信息, 性能受到限制^[17]。使用 BiLSTM 双

向循环神经网络, 每一个时刻都可以综合上下文的信息, 更好地结合上下文语境, 使模型对于语义有更好把握^[18]。

以句子“电梯井、闸门井、门槽、电缆竖井等井口应设置临时防护盖板”为例, BERT 模型训练获取 LSTM 处理的前向、后向隐藏词向量 $\mathbf{h}: \{h_{a1}, h_{a2}, \dots, h_{a7}\}$ 和 $\{h_{b1}, h_{b2}, \dots, h_{b7}\}$, 其中, \mathbf{h}_a 表示前向隐藏向量, \mathbf{h}_b 表示后向隐藏向量。电梯井、闸门井、门槽、电缆竖井、井口、设置、防护盖板在正向 LSTM 和反向 LSTM 中的隐藏向量叠加后得到词向量 $\{h_1, h_2, \dots, h_7\}$, 使 BiLSTM 储存每个字词语义信息。BiLSTM 模型处理过程如图 2 所示。

按以下步骤计算 BiLSTM 双向循环神经网络:

1) 计算经过遗忘门单元的权重 λ_f 和偏差 u_f , 通过 $t-1$ 时刻的隐藏层信息 h_{t-1} 与 t 时刻的当前单元输入规范信息 v_t , 耦合 sigmoid 函数 σ , 获得 LSTM 模型的遗忘门 f_t , 决定应丢弃或保留规范信息:

$$f_t = \sigma(\lambda_f h_{t-1} + \lambda_f v_t + u_f) \quad (1)$$

2) 通过深度学习不断更新记忆门 k_t , 保留更新后的规范信息, 从而进行更迭:

$$k_t = \sigma(\lambda_k h_{t-1} + \lambda_k v_t + u_k) \quad (2)$$

$$v_t = \tanh(\lambda_v h_{t-1} + \lambda_v v_t + u_v) \quad (3)$$

3) 上一单元保留的规范信息 c_{t-1} 和当前单元输入规范信息 v_t 组成当前单元规范信息状态 c_t 输出到下一个单元:

$$c_t = c_{t-1} \cdot f_t + k_t \cdot v_t \quad (4)$$

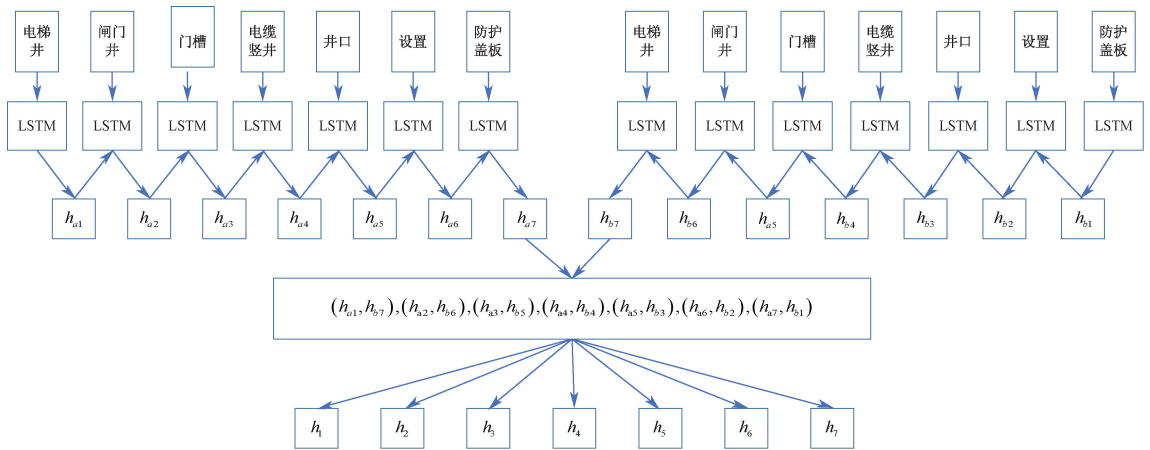


图2 BiLSTM 处理过程

Fig. 2 BiLSTM process

4) 通过计算 t 时刻输出 O_t , 输出用于决定哪些信息可作为当前阶段的任务, 以及各项关联信息:

$$o_t = \sigma(\lambda_o h_{t-1} + \lambda_o v_t + u_o) \quad (5)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (6)$$

5) 将 BiLSTM 层输出的隐藏层互相连接, 组合成一个句子的特征向量来预测标签, 这些标签表示不同类型的实体, 如脚手架、滑坡、门槽等。

1.4 CRF 层的实体依赖关系分析

水电工程施工安全规范实体部分的约束条件往往来源于其他实体^[19]。BiLSTM 只考虑到句子的上下文信息, 而未能考虑实体之间的依赖关系, CRF 可通过学习实体之间的相邻关系, 在命名实体识别过程中关联性限制实体识别, 进而确保实体识别有效。

通过 BiLSTM 模型输出大小为 $n \times m$ 的矩阵 S , 其中, n 为实体词数量, m 为实体类别。 S_{ij} 表示句子中第 i 个字符的第 j 个实体词的概率。对预测序列 $y = (y_1, y_2, \dots, y_n)$ 而言, 得到它的分数函数为:

$$S(x, y) = \sum_{i=1}^n (P_{i, y_i} + W_{y_i, y_{i+1}}) \quad (7)$$

式中: 转移矩阵 W 为 CRF 模型的参数; $W_{y_i, y_{i+1}}$ 为实体词 y_i 转移到实体词 y_{i+1} 的概率; p_{i, y_i} 为第 i 个字符预测为实体词 y_i 的概率。

对于一个句子 x 的预测实体序列 y 产生的概率为:

$$\ln(p(y|x)) = S(y|x) - \ln(\sum \exp(S(x, \tilde{y}))) \quad (8)$$

两头去对数函数, 得到预测实体序列的似然函数:

$$\ln(p(y|x)) = S(y|x) - \ln(\sum \exp(S(x, \tilde{y}))) \quad (9)$$

式中: \tilde{y} 表示命名实体识别中真实的实体序列; $\ln(\sum \exp(S(x, \tilde{y})))$ 表示所有可能的实体序列的得分, 解码后得到最大分数的输出序列:

$$y^* = \operatorname{argmax} \ln(\sum \exp(S(x, \tilde{y}))) \quad (10)$$

通过 CRF 层处理 BiLSTM 层的输出结果, 预测实体序列的概率, 并输出概率较高的实体序列, 从而使标签的输出结果更加准确。

1.5 评价指标

评价模型性能主要从准确率 P 、召回率 R 、加权平均值 F_1 等 3 个方面衡量。准确率 P 指正确识别 (True Precise, TP) 实体数占正确识别和错误识别 (False Precise, FP) 实体个数总和的占比:

$$P = \frac{TP}{TP + FP} \quad (11)$$

召回率 R 为指正确识别出实体个数 TP 占正确识别实体个数 TP 和未能识别 (False Negative, FN) 的实体个数的占比:

$$R = \frac{TP}{TP + FN} \quad (12)$$

加权平均值 F_1 值是结合准确率与召回率进行加权平均, 评价模型整体:

$$F_1 = \frac{2(P \cdot R)}{P + R} \quad (13)$$

2 命名实体识别结果分析

2.1 规范文本数据处理

以《水利水电工程施工安全防护设施技术规范》(SL714—2015)^[20] 为数据文本, 剔除文本中的停用字词、特殊符号等。同时, 为覆盖 SL714—

2015^[20]的所有信息,将实体词划分为4类,包括施工区域、施工设备、施工行为、施工人员。表1为实体对应的定义及示例。针对目标实体类型,将文本切分为单个字符,采用 BIO 标注方法实现施工文本实体标注,并进行实体分类。最终选取 1 067 条安全规范描述开展标注工作, BIO 标注语料 1 049 条规范。

表 1 实体对应的定义及示例

Table 1 Entity definition and examples

实体	定义	示例
PLA	施工区域	作业面、进出口
FAC	施工设备	脚手架、钢爬梯
ACT	施工行为	敷设、拆除
PEO	施工人员	员工、作业人员

2.2 实体识别模型计算环境设置

采用 Python3.6 计算环境,模型的超参数设置如下:总迭代次数选择 70 次是通过测试发现 70 次迭代后 loss 值函数就已经收敛,模型准确性能够得到保证;学习率决定着目标函数能否收敛到局部最小值以及何时收敛到最小值,选择 0.1 的学习率能够确保目标函数在短时间内收敛到局部最小值;Dropout 是作为缓解神经网络过拟合而被提出的一种正则化方法,选择 0.5 时,Dropout 随机生成的网络结构最多,能够有效缓解过拟合现象发生;TensorFlow 版本为 1.14.0;隐含层大小为 256;字词嵌入维度为 100;批处理数为 64。

2.3 计算结果

2.3.1 模型对比

最终构建的实体库包括 PLA、FAC、ACT、PEO 等 4 类实体,共出现 1 433 次,通过实体库筛选出水电工程施工安全相关实体,提取出施工安全关键词。使用 PYTHON 统计规范文本中的实体出现次数,结果见表 2。

表 2 实体次数

Table 2 Entity number

实体类别	次数
PLA	394
FAC	523
ACT	249
PEO	267

BERT-BILSTM-CRF 模型中不同实体计算结果见表 3。通过表 3 可知:BERT-BILSTM-CRF 模型在施工设备和施工人员方面有明确特征边界的实体词的识别准确度能够达到 96% 以上,而在施工区域和

施工行为方面的实体词边界特征较为模糊,导致识别精准度偏低。

表 3 BERT-BILSTM-CRF 模型中不同实体计算结果

Table 3 Calculation results of different entities in BERT-BILSTM-CRF model %

实体类别	P	R	F_1
PLA	91.31	89.55	90.43
FAC	98.12	97.66	97.89
ACT	92.79	91.53	92.16%
PEO	95.18	97.53	96.36

设置 BILSTM、LSTM-CRF、BILSTM-CRF 与 BERT-BILSTM-CRF 等 4 种模型,分别计算 4 种模型的精确率、召回率、 F_1 值,结果见表 4。

表 4 不同实体网络计算结果

Table 4 Calculation results of different entity

模型	networks %		
	P	R	F_1
BILSTM	78.72	77.07	77.89
LSTM-CRF	86.79	84.51	85.65
BILSTM-CRF	89.52	88.16	88.84
BERT-BILSTM-CRF	94.35	94.07	94.21

比较 BILSTM 与 BILSTM-CRF 的试验结果,增加 CRF 模型后, F_1 值提高了 10.95%,主要归因于 CRF 模型能够有效利用相邻实体的关联性,从而能够改善实体识别性能。

比较 LSTM-CRF 和 BILSTM-CRF 这 2 个计算结果,LSTM 用 BILSTM 替换后的 F_1 值高出 3.19%。BILSTM 使用双向长短期记忆网络获取水电工程施工安全规范上下文实体语义特征,从而 BILSTM 在性能上展现出相较于 LSTM 的显著提升。

使用效果更加优异的 BILSTM-CRF,同时耦合 BERT 模型进行预训练,从试验结果看, F_1 值达到 94.21%,同比 BILSTM-CRF 模型, F_1 值提高 5.37%。加入 BERT 模型,充分提取水电工程施工安全规范中各个实体之间关系的特征,预训练后的水电工程施工安全规范能够更准确地表达不同情景中的环境与实体的关联信息,进而增强模型泛化能力,提高实体识别能力。

2.3.2 模型解析

开展试验来验证和评估 BERT-BILSTM-CRF 模型在文本结构化的可靠性。对比 BILSTM 模型、LSTM-CRF 模型、BILSTM-CRF 模型、LSTM-CRF 模型和 BERT-BILSTM-CRF 模型命名实体识别任务。

图 3 为 4 种模型在试验中随着训练步数增加而损失值的变化情况。从图 3 看出,尽管未在解码层

使用 CRF 的 BILSTM 模型也表现出良好的收敛速度,表明 BILSTM 模块具备类似 CRF 分析依赖关系的能力。此外,比较 BILSTM-CRF 模型和 LSTM-CRF 模型的收敛速度,当 2 个模型的解码层都采用 CRF 时,BILSTM-CRF 模型的损失收敛明显快于 LSTM-CRF 模型,表明基于 BILSTM 的模型在损失收敛上更为高效。

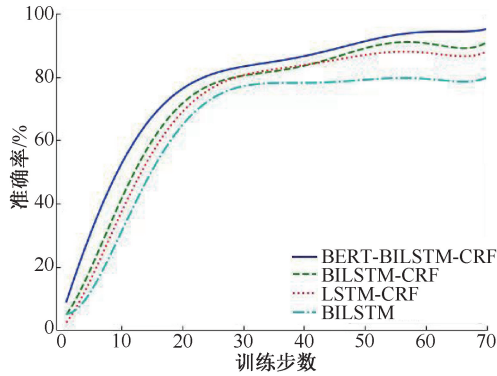


图 3 各标注模型的损耗收敛对比

Fig. 3 Loss convergence comparison of entity recognition models

图 4 为 4 种模型在命名实体识别任务训练集上的精确率对比。从图 4 可以看出,随着训练步数的增加,BERT-BILSTM-CRF 模型的精确率总体上优于其他模型。结合图 3 和图 4 可知:在解码层添加 CRF 模型加快了 BERT-BILSTM-CRF 模型的损失收敛速度,提高了模型的精确率。

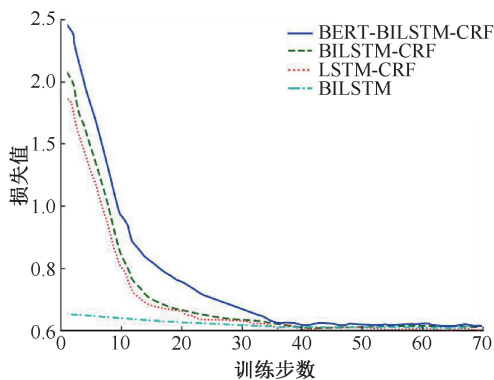


图 4 各标注模型的精确率对比

Fig. 4 Comparison of the accuracy rate of each entity recognition model

2.3.3 模型语句识别对比

通过对同一条规范文本进行命名实体识别,各个模型识别精度有所区别,具体区别见表 5。

通过表 5 可知:BILSTM 模型缺少 CRF 的实体依赖关系分析,导致所有实体都是独立的,无法体现

表 5 模型文本识别示例

Table 5 Example of model text recognition

模型	识别效果
BILSTM	设有消防/安全/通道,油库/内/道路宜布置成环行道,车道宽应不小于 4 m
LSTM-CRF	设有消防/安全通道,油库/内/道路宜布置成环行道,车道宽应不小于 4 m
BILSTM-CRF	设有消防/安全通道,油库内/道路宜布置成环行道,车道宽应不小于 4 m
BERT-BILSTM-CRF	设有/消防安全通道/,油库内/道路/宜布置成环行道,车道宽应/不小于/4 m

关联性;LSTM-CRF 模型无法进行双向深度学习,在处理定位词方面不够精准;BILSTM-CRF 模型缺乏大规模无标注预训练,导致未出现过的词语无法与实体关联;BERT-BILSTM-CRF 模型取得较好的识别效果,在提取实体关联性等方面也有良好性能。

2.3.4 消融试验

对水电工程施工安全规范实体识别模型进行消融试验,以分析各个模型模块对性能的提升效果。表 6 为消融试验的结果。通过逐步移除各个模块并观察其对整体性能的影响,可清楚地了解每个模块在提升模型性能中的作用。首先,去除 CRF 模块,进行预训练和双向循环神经网络迭代,在表中表示为 BERT-BILSTM 模型。可以观察到,在去除 CRF 模块依赖关系的限制条件时, P 、 R 、 F_1 指标均有所下降, P 值下降 6.7%, R 值下降 7.74%, F_1 值下降 7.22%。指标下降原因是由于 CRF 模型对实体识别关联性限制的增强效应。

表 6 消融试验结果

Table 6 Results of ablation experiment

模型	P	R	F_1
BERT	67.53	65.05	66.29
BERT-BILSTM	87.65	86.33	86.99
BERT-BILSTM-CRF	94.35	94.07	94.21

之后,去除水电工程施工安全规范实体识别方法中的双向循环神经网络模型,方法退化为仅剩 BERT 模型进行大规模语料无标注预训练的命名实体识别方法。 P 、 R 、 F_1 指标继续下降, P 值下降 20.12%, R 值下降 21.28%, F_1 值下降 20.7%。单独使用 BERT 模型预训练方法准确率下降明显,说明在大规模语料无标注预训练中,仅采用 BERT 模型难以识别水电工程施工安全领域的实体。消融试

验中,分别对水电工程施工安全规范实体识别方法中的2个核心模块消融,验证每个模块对实体识别效果带来的提升,反映出所提算法在水电工程施工安全领域有良好的适应性。

3 结 论

1) 提出水电工程施工安全规范实体识别 BERT-BiLSTM-CRF 方法,实例计算结果得到模型实

体识别精确率为 94.35%,高于 BiLSTM、LSTM-CRF、BiLSTM-CRF 等 3 种传统命名实体识别方法,证明所提方法在处理水电工程施工安全规范实体识别方面具有优越性。

2) 水电工程施工安全规范知识不仅由实体组成,实体之间也存在实体关联关系,扩大规范知识实体识别范围,提炼实体关系,建立水电工程施工安全的知识图谱,是下一步的研究方向。

参 考 文 献

- [1] 陈述,习俊博,王建平,等. 水电工程施工安全隐患关联规则挖掘[J]. 中国安全科学学报, 2021, 31(8): 75-82.
CHEN Shu, XI Junbo, WANG Jianping, et al. Mining association rules of near-misses of hydropower projects construction[J]. China Safety Science Journal, 2021, 31(8): 75-82.
- [2] 樊启祥,林鹏,魏鹏程,等. 水电工程安全事故发生机制与管理对策[J]. 中国安全科学学报, 2019, 29(1): 144-149.
FAN Qixiang, LIN Peng, WEI Pengcheng, et al. Hydropower engineering safety accident occurrence mechanism and management strategies[J]. China Safety Science Journal, 2019, 29(1): 144-149.
- [3] CHEN Shu, TIAN Ya, JIN Lianghai, et al. Estimating the frequency of exposure to uncertain hazards: impact of wind conditions on concrete dam construction[J]. Journal of Construction Engineering and Management, 2021, 147(2): 1-13.
- [4] 张小衡,王玲玲. 中文机构名称的识别与分析[J]. 中文信息学报, 1997, 11(4): 21-32.
ZHANG Xiaoheng, WANG Lingling. Identification and analysis of Chinese institutional names[J]. Journal of Chinese Information Processing, 1997, 11(4): 21-32.
- [5] BIKEL D M, SCHWARTA R, WEISCHEDEL R M. An algorithm that learns what's in a name[J]. Machine Learning Journal Special Issue on Natural Language Learning, 1999, 34(1/2/3): 211-231.
- [6] ZHANG Qiqi, XUE Cong, ZHOU Peng, et al. Named entity recognition for Chinese construction documents based on conditional random field[J]. Frontiers of Engineering Management, 2022, 10(12): 237-249.
- [7] TAEKHYUNG K, SEOKHO C. Accident case retrieval and analyses: using natural language processing in the construction industry[J]. Journal of Construction Engineering and Management, 2019, 145(3): 1-13.
- [8] 李明超,田丹,沈扬,等. 融入 Attention 机制改进 Word2vec 技术的水利水电工程专业词智能提取与分析方法[J]. 水利学报, 2020, 51(7): 816-826.
LI Mingchao, TIAN Dan, SHEN Yang, et al. An intelligent extraction and analysis approach of professional technical words for hydraulic engineering by improved Word2vec technology with Attention mechanism[J]. Journal of Hydraulic Engineering, 2020, 51(7): 816-826.
- [9] 刘婷,张社荣,王超,等. 水利施工事故文本智能分析的 BERT-BiLSTM 混合模型[J]. 水力发电学报, 2022, 41(7): 1-12.
LIU Ting, ZHANG Sherong, WANG Chao, et al. Text intelligent analysis for hydraulic construction accidents based on BERT-BiLSTM hybrid model[J]. Journal of Hydroelectric Engineering, 2022, 41(7): 1-12.
- [10] 杨秀璋,郭明镇,候红涛,等. 融合情感词典的改进 BiLSTM-CNN + Attention 情感分类算法[J]. 科学技术与工程, 2022, 22(20): 8 761-8 770.
YANG Xiuzhang, GUO Mingzhen, HOU Hongtao, et al. Improved BiLSTM-CNN + Attention sentiment classification algorithm fused with sentiment dictionary[J]. Science Technology and Engineering, 2022, 22(20): 8 761-8 770.
- [11] 易明,李霍然,刘继月. 基于 Glove-BiLSTM 的在线研讨信息分类模型研究[J]. 情报理论与实践, 2022, 45(9): 173-179, 172.
YI Ming, LI Huoran, LIU Jiyue. Research on online discussion information classification model based on Glove-

- BILSTM[J]. *Information Studies: Theory & Application*, 2022, 45(9): 173-179, 172.
- [12] 杨秀璋,彭国军,李子川,等. 基于 Bert 和 BiLSTM-CRF 的 APT 攻击实体识别及对齐研究[J]. *通信学报*, 2022, 43(6): 58-70.
YANG Xiuzhang, PENG Guojun, LI Zichuan, et al. Research on entity recognition and alignment of APT attack based on Bert and BiLSTM-CRF[J]. *Journal on Communications*, 2022, 43(6): 58-70.
- [13] CHEN Shu, XI Junbo, CHEN Yun, et al. Association mining of near misses in hydropower engineering construction based on convolutional neural network text classification[J]. *Computational Intelligence and Neuroscience*, 2022, 2022: 1-16.
- [14] 赵鹏辉,李俊杰,康飞. 河道山地灾害的卷积神经网络快速识别方法[J]. *水利水运工程学报*, 2019(2): 65-70.
ZHAO Penghui, LI Junjie, KANG Fei. Fast recognition method for mountain hazards in river courses based on convolutional neural networks[J]. *Hydro-Science and Engineering*, 2019, 19(2): 65-70.
- [15] 于媛芳,张勇,左皓阳,等. 基于角色信息引导的多轮事件论元抽取[J]. *北京大学学报:自然科学版*, 2023, 59(1): 83-91.
YU Yuanfang, ZHANG Yong, ZUO Haoyang, et al. Multi-turn event argument extraction based on role information guidance[J]. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2023, 59(1): 83-91.
- [16] 乔卫亮,刘阳,周群,等. 基于模糊人工神经网络的安全风险评估模型[J]. *安全与环境学报*, 2021, 21(4): 1405-1411.
QIAO Weiliang, LIU Yang, ZHOU Qun, et al. A risk assessment model based on fuzzy artificial neural network[J]. *Journal of Safety and Environment*, 2021, 21(4): 1405-1411.
- [17] 宋敦江,杨霖,钟少波. 基于 BERT 的灾害三元组信息抽取优化研究[J]. *中国安全科学学报*, 2022, 32(2): 115-120.
SONG Dunjiang, YANG Lin, ZHONG Shaobo. Research on optimization of disaster triplet information extraction based on BERT[J]. *China Safety Science Journal*, 2022, 32(2): 115-120.
- [18] 黄光球,王瑞泽,陆秋琴. 基于 K-CNN-BiLSTM 的关联区域 VOCs 浓度时空关联预测[J]. *安全与环境学报*, 2023, 23(4): 1336-1348.
HUANG Guangqiu, WANG Ruize, LU Qiuqin. Spatial and temporal correlation prediction of VOCs concentration in associated areas based on K-CNN-BiLSTM[J]. *Journal of Safety and Environment*, 2023, 23(4): 1336-1348.
- [19] SUBAKTI A, MURFI H, HARIADI N. The performance of BERT as data representation of text clustering[J]. *Journal of Big Data*, 2022, 9(8): 163-172.
- [20] SL714—2015,水利水电工程施工安全防护设施技术规范[S].
SL714—2015, Technical specification for safety protection facilities for construction of water and hydropower projects[S].

作者简介: 陈述 (1986—),男,湖北英山人,博士,教授,主要从事安全管理研究。E-mail:chenshu@ctgu.edu.cn。