

文章编号:1671-1637(2025)01-0248-15

基于桥检领域-任务迁移的检测报告信息 提取少样本模型

朱彦洁,王瑜晨,熊 文*,蔡春声

(东南大学 交通学院,江苏 南京 211189)

摘 要:为减少桥梁检测关键信息提取方法对大量人工标注样本的依赖,提出了一种适用于少样本场景的桥梁检测关键信息提取模型,由桥检领域预训练语言模型、双向长短时记忆(BiLSTM)网络和条件随机场(CRF)组成;通过使用桥梁领域语料与检测任务数据对原始语言模型进行领域预训练和任务微调,实现从领域知识到任务特征的两阶段迁移,构建出更适应桥梁专业术语和检测报告格式的预训练语言模型;利用 BiLSTM 捕捉桥梁检测报告中的上下文依赖关系,并结合 CRF 对最终信息提取结果进行约束优化;根据行业规范和现有相关研究,重新定义了 8 类桥梁检测报告中通用的关键信息;为验证方法的有效性,分别在仅包含 50 和 100 个句子的少样本数据上进行训练,并在 1 491 个句子的测试集上评估性能。试验结果表明:当训练样本数分别为 50 和 100 个时,本文提出模型的 F_1 值分别达到 0.860 7 和 0.820 2,均显著优于 4 个主流模型,验证了该模型在少样本情况下对桥检报告关键信息的精准提取能力;消融试验进一步证明了领域与任务两阶段迁移学习策略在快速提取少样本数据中领域相关信息和任务显著特征方面的有效性,从而显著提升了模型在少样本场景上的整体性能;提出的少样本场景下桥梁检测信息提取方法可用于构建知识图谱,以评估桥梁的结构状态和预测未来可使用寿命。

关键词:桥梁工程;信息提取;预训练语言模型;桥梁检测报告;深度学习;自然语言处理

中图分类号:U446.3 **文献标志码:**A **DOI:**10.19818/j.cnki.1671-1637.2025.01.018

Few-shot model for extracting inspection report information based on bridge inspection domain-task transfer

ZHU Yan-jie, WANG Yu-chen, XIONG Wen*, CAI Chun-sheng

(School of Transportation, Southeast University, Nanjing 211189, Jiangsu, China)

Abstract: To reduce the reliance on large-scale manually annotated samples for extracting key information of bridge inspection, a bridge inspection key information extraction model applicable to few-shot scenarios was proposed. The method comprised a bridge inspection-specific pre-trained language model, a bi-directional long short-term memory (BiLSTM) network, and a conditional random field (CRF). Through the utilization of bridge-related corpora and inspection

收稿日期:2024-01-30

基金项目:国家自然科学基金项目(52478147,52378135)

作者简介:朱彦洁(1994-),女,江苏宿迁人,东南大学副教授,哲学博士,从事数字桥梁研究。

*通信作者:熊 文(1982-),男,安徽合肥人,东南大学教授,工学博士。

引用格式:朱彦洁,王瑜晨,熊 文,等. 基于桥检领域-任务迁移的检测报告信息提取少样本模型[J]. 交通运输工程学报,2025,25(1):248-262.

Citation: ZHU Yan-jie, WANG Yu-chen, XIONG Wen, et al. Few-shot model for extracting inspection report information based on bridge inspection domain-task transfer[J]. Journal of Traffic and Transportation Engineering, 2025, 25(1): 248-262.

task data for the domain pre-training and task fine-tuning, a two-stage transfer from domain knowledge to task-specific features was implemented, resulting in a pre-trained language model with high generalization for bridge-specific terminology and inspection report formats. Subsequently, the BiLSTM was employed to capture the contextual dependencies within the bridge inspection reports, while the CRF was combined to enforce structured constraints on the final extraction results. According to industry specifications and relevant research, eight key information categories commonly found in bridge inspection reports were redefined. To validate the effectiveness of the proposed approach, few-shot datasets containing only 50 and 100 sentences were utilized for training, respectively, with performance evaluated on a test set containing 1 491 sentences. Experimental results show that, when trained with 50 and 100 samples respectively, the F_1 scores of the proposed model reach 0.860 7 and 0.820 2, respectively, significantly outperforming four mainstream models. This confirms the model's superior capability in accurately extracting key information from bridge inspection reports under few-shot conditions. Moreover, ablation experiments further reveal that the two-stage transfer learning strategy effectively facilitates the rapid extraction of domain-relevant and task-discriminative features from few-shot data, thereby substantially enhancing the model's overall performance in few-shot scenarios. The proposed bridge inspection information extraction method under few-shot scenarios can be used to construct a knowledge graph for evaluating bridge structural conditions and predicting future service life.

Keywords: bridge engineering; information extraction; pre-trained language model; bridge inspection report; deep learning; natural language processing

Funding: National Natural Science Foundation of China (52478147, 52378135)

* **Corresponding author:** XIONG Wen(1982-), male, professor, PhD, wxiong@seu.edu.cn.

0 引言

桥梁检测报告(简称“桥检报告”)蕴含着运营阶段桥梁结构整体与局部状况的重要信息^[1],对桥梁养护决策至关重要。具体而言,桥检报告中蕴含的病害变化、劣化趋势和维修记录等关键信息,是桥梁结构状况评估和剩余使用寿命估计的基础,也是养护维修计划制定的依据^[2-5]。例如,Wang等^[6]融合了检测报告中的各类信息,全面评估了桥梁的整体状况;Xia等^[1]基于定期桥梁检测报告,研究了养护行为对桥梁状况劣化的影响;Yang等^[7]提出了一个知识库问题框架,利用报告中的领域信息为桥梁养护提供支持。然而,历年桥检报告的文本存储形式给数据分析带来了挑战,手动从报告中提取信息的效率低、耗时长、成本高^[8],因此,以自然语言处理方法为核心的信息提取方法成为近年研究热点。

在自然语言处理领域,桥检报告信息提取可以看作是命名实体识别任务,即从检测报告文本中识别出与桥梁构件和病害相关的重要信息。目前的桥检报告信息提取方法主要基于监督学习^[9],且需要

包含至少上千个样本的数据集进行训练^[10-14],才能达到实际任务所需性能。例如,Liu等^[9]使用了1 866个样本,模型的综合性能指标 F_1 值达到了0.907 0;Moon等^[12]标注了1 650个样本,训练后模型的 F_1 值为0.860 0;Li等^[11]提出的方法则使用了13 116个样本才让模型的 F_1 值达到了0.98。现有的研究均使用了成千上万的标注样本供模型训练,才能实现有效的信息提取。

然而,数据标注耗时耗力且受样本标注人员的主观因素影响较大^[15-16]。此外,由于负责桥梁检测的机构不同,行文的特点和标准也不同,使得不同地区/年份的报告在行文上存在较大差异,这均为信息提取带来挑战,因此,难以通过构建1个标准数据集辐射全部桥检报告信息提取任务,只能针对不同的任务需求进行大量的文本数据集构建工作,极大限制了现有方法的应用场景和使用频率以及在实际工程中的推广。目前,亟需开发一个仅使用少量样本就可以实现桥检信息高精度提取的方法,以满足实际工程需求。

预训练语言模型的出现使利用少量样本实现高

精度的信息提取成为可能^[17]。通过使用大量通用语料信息进行预先的模型自监督学习,使模型可通过调整输出网络结构,快速应用于包括实体识别^[18]、文本分类^[19]和对话问答^[20]等众多领域任务。同样,在进行桥检信息提取时,使用预训练语言模型,可以直接加载模型的内部训练参数,输出包含丰富语义信息的文本特征,使信息提取模型可以直接挖掘更深层的文本特征,从而减少标注样本的使用数量,这使得仅使用由少量样本训练的模型实现桥检信息的高精度提取成为可能。2019 年,BERT(Bi-directional Encoder Representation from Transformers)模型在 11 个任务中取得了最优的结果^[21]。但是,由于使用通用语料库训练,使其在桥梁这类专业领域上的泛化能力一般,无法输出包含领域语义特征的文本向量,很难大幅度减少桥检信息提取任务所需的样本量。

综上所述,由于现有桥检信息提取方法对于大样本的需求,加之实际任务需求的高度差异化,使得现有信息提取方法难以有效推广,极大地限制了检测报告中关键信息的利用率。针对此现状,为实现基于少量样本的高精度信息提取,本研究拟使用桥梁领域通用语料和桥梁检测报告文本进行检测领域和信息提取任务上的两阶段迁移学习,以提升预训练语言模型在桥检信息提取任务中的泛化能力,并

通过连接双向长短时记忆(Bi-directional Long Short-Term Memory, BiLSTM)网络和条件随机场(Condition Random Field, CRF)捕捉文本上下文的双向语义依赖,并获取相邻实体之间的依赖关系,从而实现仅使用少量的标注文本进行训练,便可准确地提取桥检报告中的关键信息。

1 桥检关键信息分类方法

桥检信息的高效提取前提是对关键信息的科学分类。表 1 总结了既有研究中桥检报告所需提取的关键信息类别,可以看出当前关键信息类别的定义存在显著差异且无统一标准。其中,Liu 等^[9,22]提出的桥检报告关键信息类别高达 11 种,对于不同地区的桥检信息提取任务而言,若要提取 11 类信息,数据集可能会出现类别不平衡现象。例如,时间信息在大多数检测报告中仅在概况提及,因此,该类信息在报告中的数量很少会导致类别不平衡,而这种不平衡会极大影响模型性能,过细过多的关键信息类别划分并不利于实际应用。Moon 等^[12]仅考虑对构件、病害和原因这 3 类信息进行提取,然而,实际上病害位置和性状等信息对于后续的结构评估十分关键,因此,过少的信息类别很难支撑后续养修决策。此外,由于病害性状的描述形式多变,导致其提取难度较大,进一步为桥检信息的科学划分带来挑战。

表 1 既有研究对关键桥梁检测信息的分类

Table 1 Categories of key bridge inspection information from existing studies

现有研究	年份	信息数量	从桥梁检测报告中提取关键信息的类别
文献[23]	2022	6	桥梁构件、病害、测量值、测量单位、数量描述、严重程度描述
文献[11]	2021	6	桥梁、桥梁构件、构件部分、病害位置、病害、病害描述
文献[10]	2021	5	病害、严重程度定性描述、严重程度定量描述、病害位置、其他
文献[9]、[22]	2021、2017	11	桥梁构件、病害、病害原因、维护措施、维修材料、测量值、测量值单位、数量描述、严重程度描述、日期、其他
文献[12]	2022	3	桥梁构件、病害、原因

为此,本文根据行业规范并总结现有相关研究,对桥梁检测报告中通用的关键信息重新定义。

信息类别 1 和 2:桥梁构件、构件位置。

桥梁构件作为结构组成部分,相同病害出现在不同构件/位置,对结构运营状态影响程度不同,导致养护决策的不同。例如,箱梁腹板和底板均存在裂缝,由于产生裂缝的原因不同,对 2 种病害的维护措施不同。

信息类别 3 和 4:病害、病害特征。

不同病害对于结构状态的影响程度并不同,其重要程度不言而喻。而作为表征病害的病害特征则

会显示更为具体的细节^[24],例如,纵向、横向和网状裂缝的成因不同,需采取的具体养护措施也不同。

信息类别 5~8:病害数量、病害分布、测量类别、测量值。

病害性状信息越丰富具体,制定的养护策略就越详细且有针对性^[25],因此,病害性状的提取十分重要,但以往多数研究均忽略该类信息。其中,病害数量反映构件实际损伤程度;病害分布即病害密度,反映结构整体的损伤情况;测量类别包括裂缝宽度与长度、剥落与锈蚀面积、支座空隙率等;测量值则会量化病害对构件的真实影响。

根据上述分析,最终确定了桥梁构件、构件位置、病害、病害特征、病害数量、病害分布、测量类别、测量值等 8 种类别作为桥梁检测报告待提取的基本关键信息类别。

2 桥检关键信息提取方法

提出的桥检关键信息提取方法如图 1 所示,图中:B 为关键信息开始词的标签;I 为关键信息中间

词的标签。图 1 包含 3 个主要部分:桥检领域预训练语言(BERT-Inspection)模型、BiLSTM 网络、CRF。信息提取具体过程是检测文本通过经桥梁领域和桥检任务迁移的 BERT-Inspection 模型以获得融合领域信息和先验知识的词嵌入,然后经过 BiLSTM 网络以捕捉检测语句的双向语义关系,通过连接 CRF 以获取相邻字符之间的依赖关系,最终输出识别出的桥检关键信息。

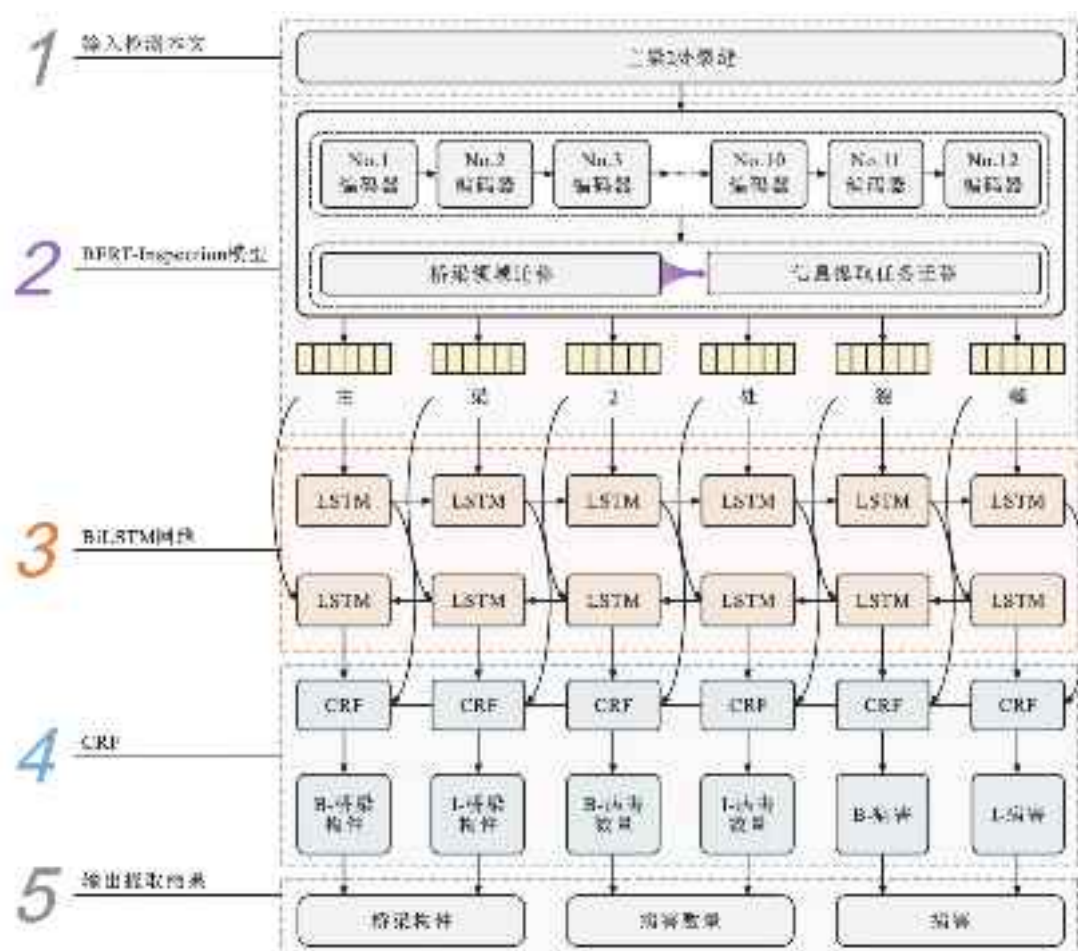


图 1 关键信息提取方法流程

Fig. 1 Workflow of key information extraction method

2.1 BERT-Inspection 模型

为了提高模型的领域和任务的泛化能力,采用两阶段的迁移学习策略以实现仅利用较少标注数据完成关键信息的精准提取。首先,基于已有的 BERT 中文预训练语言模型,利用掩码语言模型和下句预测技术对基础模型进行桥梁领域的迁移训练。然后,采用命名实体识别任务开展检测信息提取的任务微调,最终完成从桥梁领域到检测信息提取任务的两阶段迁移,从而得到更适合桥检报告信息提取任务的预训练语言模型。

BERT-Inspection 模型是由 BERT 中文预训练

语言模型经过领域和任务迁移获得,由 12 个编码器组成。对于该模型而言,为便于识别句子的开头和结尾,模型会自动在文本序列开头添加特殊标记[CLS],并在句子间添加另一个特殊标记[SEP],以确定句子的开始和结束。所有文本信息均进行 3 次嵌入。

(1)词嵌入:将文本序列中所有词编码为向量形式。

(2)句段嵌入:用于表示词所属的具体句子。

(3)位置嵌入:用于表示词所在句中的位置信息。

与 Transformer 模型中位置编码直接使用正余

弦函数嵌入不同,这里的位置嵌入是通过构造参数实现模型自行学习位置信息。

将维度均为 1×768 的 3 组嵌入向量输入到编码层中,并将 3 组向量按元素相加得到最终输入向量 \mathbf{X} ,输入到第一个编码层。在第一个编码层内,输入向量 \mathbf{X} 分别乘以 3 组不同权重矩阵 $\mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v$

$$\begin{cases} \mathbf{Q} = \mathbf{X}\mathbf{W}_q \\ \mathbf{K} = \mathbf{X}\mathbf{W}_k \\ \mathbf{V} = \mathbf{X}\mathbf{W}_v \end{cases} \quad (1)$$

式中: \mathbf{Q} 为查询矩阵; \mathbf{K} 为键矩阵; \mathbf{V} 为值矩阵。

获得 $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 后进行自注意力机制计算

$$B(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = S\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_h}}\right)\mathbf{V} \quad (2)$$

式中: $B(\cdot)$ 为注意力机制运算函数; $S(\cdot)$ 为 Softmax 激活函数; d_h 为嵌入维度, h 为键矩阵 \mathbf{K} 的向量维度。

在这里定义执行第 1 次自注意力机制计算为 h_1 ,执行第 i 次自注意力机制 h_i 为

$$h_i = B(\mathbf{Q}, \mathbf{K}, \mathbf{V})_i \quad (3)$$

式中: i 为多头自注意力机制的并行个数,取 $1 \sim 12$ 。

为了利用不同的自注意力模块获得文本中每个字在不同语义空间下的增强语义向量,使用多层自注意力机制,以计算不同方面注意力信息,并通过将每个头的注意力计算结果进行线性组合,从而得到多头注意结果

$$M(\mathbf{X}) = D(h_1, \dots, h_i, \dots, h_{12})\mathbf{W}_0 \quad (4)$$

式中: $M(\cdot)$ 为多头注意力运算; $D(\cdot)$ 为线性合并运算; \mathbf{W}_0 为线性层的权重矩阵。

随后,将多头注意力结果进行残差和归一化运算

$$\mathbf{Y} = N[\mathbf{X} + M(\mathbf{X})] \quad (5)$$

式中: \mathbf{Y} 为残差和归一化运算结果; $N(\cdot)$ 为归一化运算。

将输出 \mathbf{Y} 连接至包含 2 个全连接层的前馈层,其中第 1 层激活函数为 ReLU,第 2 层为线性激活函数,前馈层输出为

$$F(\mathbf{Y}) = \max\{0, \mathbf{X}\mathbf{W}_1 + \mathbf{b}_1\}\mathbf{W}_2 + \mathbf{b}_2 \quad (6)$$

式中: $F(\cdot)$ 为经过前馈层运算; \mathbf{W}_1 和 \mathbf{W}_2 均为权重矩阵; \mathbf{b}_1 和 \mathbf{b}_2 均为偏置向量。

最后,再次对前馈层输出进行残差和归一化运算,得到第 1 个编码层的输出向量 \mathbf{Z}

$$\mathbf{Z} = N[\mathbf{Y} + F(\mathbf{Y})] \quad (7)$$

第 2 个编码层将以 \mathbf{Z} 作为输入向量,重复执行式(1)~(7),并持续运算至第 12 个编码层,进而获

得最终模型输出。

在桥梁领域迁移训练中,使用掩码语言模型和下一句预测进行领域迁移。(1)掩码语言模型:在原句子中随机选取 15% 进行掩码操作,即选取内容中的 80% 使用特殊字符 [MASK] 进行替换,10% 使用原本句子中文本进行替换,10% 不做任何改变;然后,将掩码后的句子传入模型中预测 [MASK] 掩码位置的正确文本内容,以提升模型对桥梁领域文本上下文相关信息的深度挖掘能力。(2)下一句预测:首先,随机抽取一半正确的句子对和一半错误的句子对作为模型的输入,然后通过判断下句话是否正确来进行模型内容参数的更新和优化,以提高模型对于桥梁领域文本在句意层面的信息理解能力。2 个任务均不需要使用标注信息,仅利用桥梁领域书籍、期刊、规范等文本信息即可进行训练。具体流程如图 2 所示。



图 2 桥梁领域迁移流程

Fig. 2 Transfer workflow of bridge domain

在检测信息任务迁移训练中,使用命名实体识别任务进行微调。首先,根据行业规范收集包含典型信息的样本,包括不同形式桥梁的结构构件、各构件常见损伤病害以及典型的维修方法和维护形式。通过桥梁专业人士对这些样本进行注释,得到微调数据集进行任务迁移。任务迁移的训练过程如图 3 所示。首先,使用经过领域迁移的桥梁领域预训练语言模型,以获得输入检测文本的嵌入向量。随后,将获得的嵌入向量输入到微调层(微调层为全连接网络并采用 Softmax 函数为激活函数)中,得到信息提取任务的预测结果。根据获得的任务迁移数据集的标签进行预训练模型内部参数更新。在这个训练过程中,模型的输入是一个句子,输出则是每个字的标签。最终,可以得到包含各种检测关键信息先验知识的桥检预训练语言模型。

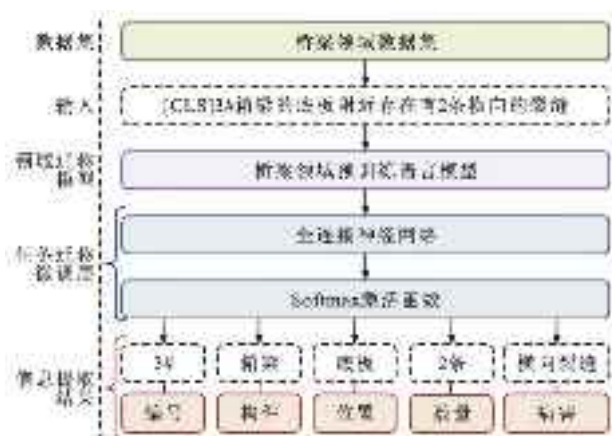


图 3 信息提取任务迁移流程

Fig. 3 Transfer workflow of information extraction task

2.2 BiLSTM 网络

桥检领域预训练语言模型会依据输入文本生成融合了领域信息和先验知识的嵌入向量。为了实现从检测长文本中提取关键信息这一任务,还需一个能够挖掘序列信息的网络。长短期记忆(Long Short-Term Memory, LSTM)网络能够实现序列前后依赖关系的预测和文本长距离依赖关系的捕捉^[26]。然而,LSTM 是一种前向传播结构,只能学习文本序列在正词序中的依赖特征。对于信息提取任务来说,需要同时利用正逆 2 个词序来识别关键信息,因此,使用 BiLSTM 网络用来捕捉这种双向信息。

单向 LSTM 网络由遗忘门、输入门、输出门和存储单元结构组成, h_t 为 LSTM 单元在 t 时间的输出向量,由存储单元和输出门的输出相乘得到。BiLSTM 由关注正词序方向和逆词序方向的 2 个单向 LSTM 组成,并通过合并 2 个单向 LSTM 在同一时间的输出获得最终的神经网络输出。对于 t 时间,对应于正向和逆向信息的实际输出为

$$h_t = (h_{t,1}, h_{t,2}) \quad (8)$$

式中: $h_{t,1}$ 为 t 时间正词序方向 LSTM 输出; $h_{t,2}$ 为 t 时间逆词序方向 LSTM 输出。

为了获得文本序列中每个字与待提取信息类别的具体量化关系,BiLSTM 隐含状态序列的输出连接到一个全连接网络,从而将状态序列从 q 维映射到 r 维,信息类别的得分矩阵计算公式为

$$P = \sum_k P_{k,b_k} \quad (9)$$

式中: P 为 BiLSTM 输出的得分矩阵,其大小为 $q \times r$, r 为序列中词总数, q 为待提取信息类别数; P_{k,b_k} 为 P 中第 k 个信息类别属于标签 b_k 的得分。

BiLSTM 可以通过其双向结构有效捕捉桥检文

本序列中的前(后)依赖关系,这对于理解报告中的复杂结构和语义关系至关重要。同时,由于能够很好处理文本的长距离依赖关系,在长文本中捕捉到远距离的上下文关系,使其更适合处理单句包含较多信息的桥检文本。

2.3 CRF

CRF 层用于序列标注任务最后阶段,以考虑到标签序列之间的依赖关系,从而在全局范围内优化预测的标签序列。确保输出的标签序列在整体上的合理性,提高模型在信息提取任务上的精确度。

由于 BiLSTM 所输出相邻标签间具有依赖关系,而一般 Softmax 函数无法处理这种依赖关系,但 CRF 则可以根据相邻信息类别的标签间关系得到最优预测序列,进而弥补 Softmax 函数缺点。同时,CRF 可以利用标签状态转换概率作为得分,从而引入约束以消除 BiLSTM 可能的输出误差,保证最终信息类别预测结果的有效性。

对于桥梁检测报告文本向量 $T = (t_1, t_2, \dots, t_r)$,将 BiLSTM 得分矩阵 P 和 CRF 转换得分矩阵 A 相加得到信息类别序列集合 $b = \{b_1, b_2, \dots, b_r\}$ 的最终得分

$$C(T, b) = \sum_{k=1}^{r+1} A_{b_{k-1}, b_k} + \sum_{k=1}^r P_{k, b_k} \quad (10)$$

式中: $C(\cdot)$ 为最终矩阵的得分运算; A_{b_{k-1}, b_k} 为 A 中从第 k 个信息类别标签到第 $k-1$ 个信息类别标签的转换评分。

所有可能的信息类别标签在文本序列中出现的概率为

$$p(b | T) = \frac{e^{C(T, b)}}{\sum_{\bar{b} \in B} e^{C(T, \bar{b})}} \quad (11)$$

式中: B 为所有可能的信息类别标签的集合; $p(b | T)$ 为给定输入序列 T 后标签序列 b 出现的概率; \bar{b} 为每种可能的标签序列。

对上式取对数即可得到损失函数 L 为

$$L = -\ln[p(b | T)] \quad (12)$$

对其进行 Viterbi 解码,得到预测的输出序列

$$b^* = \operatorname{argmax}_{\bar{b} \in B} C(T, \bar{b}) \quad (13)$$

式中: b^* 为输入向量 T 上所有可能标签序列中最高的总得分,所对应的标签序列就是模型最终输出的命名实体标注结果。

2.4 模型性能评估指标

在信息提取领域有实体级和标签级 2 类评估方式。标签级计算的是正确标签的数量。然而,实际

信息是由长短不一的标签组成的。这种评价方式很难评估模型在准确提取关键信息方面的真实性能,因此,采用实体级指标来评估所提出模型的综合性能,即一个关键信息全部标签正确才算做提取正确。最终,选择实体级的精准度、召回率和 F_1 值^[27] 3 个代表性指标评估所提出的桥梁检测信息提取方法的性能。

模型的总体精准度记为 P , 总体召回率记为 R , 计算公式分别为

$$P = \sum_{k=1}^N \frac{n_k S_k}{N(S_k + V_k)} \quad (14)$$

$$R = \sum_{k=1}^N \frac{n_k S_k}{N(S_k + v_k)} \quad (15)$$

式中: S_k 为正确提取的第 k 个信息类别个数; V_k 为错误识别为第 k 个信息类别的个数; v_k 为第 k 个信息类别错误识别为其他类别的个数; N 为全部信息类别的数本数量; n_k 为第 k 个信息类别的样本数量。

通常, 精准度变大召回率就会降低, 同样地, 召回率变大精准度就会降低, 因此, 使用 F_1 值以综合评估模型的性能。 F_1 值是精准度和召回率的调和平均值, 是评估不平衡数据的综合评价指标^[28], 提供了二者之间的平衡视角, 计算公式为

$$F_1 = \sum_{k=1}^N \frac{2n_k S_k}{N(2S_k + V_k + v_k)} \quad (16)$$

F_1 值的范围是 $0 \sim 1$, F_1 值越高, 表明模型的性能越好, 即精准度和召回率之间达到了较好的平衡。

3 案例分析与验证

3.1 数据收集和准备

为了验证所提出桥检信息提取方法的有效性和准确性, 收集了来自某省内全部高速公路桥梁的多份定期检测报告, 并形成中文桥梁检测数据集。在数据预处理阶段, 仅保留包含关键信息的文本内容, 并将所有测量单位统一修改为中文表达, 以确保文本的一致性。最终, 将桥检报告以句号为标志分割, 形成包含 1 626 个句子(51 229 个字)的中文桥梁检测数据集, 部分样本信息见表 2。

最终获得包含 11 484 个关键信息的数据集。

对于少样本训练来说, 尚无严格规定限值, 在统计学领域, 一般认为样本量小于 50 个即可以算作少样本。在信息抽取任务中, 由于目标数据集样本量一般较大, 也可以认为小于 100 个即为少样本, 因此, 为了验证模型在少样本数据集集中的性能, 分别随

表 2 训练数据部分样本
Table 2 Partial samples of training data

序号	数据集样本内容
1	伸缩缝存在止水胶带破损病害 1 处, 位于 1#
2	台身存在竖向裂缝 1 条, 裂缝长度为 2.5 m, 裂缝宽度为 0.2 mm, 存在于 Y-0# 台身大里程面
3	桥面铺装存在网向裂缝 32 条, 分布于 2 根构件上, 网向裂缝总长度为 88.66 m
4	小箱梁存在锈胀露筋 1 处, 总面积为 0.03 m ²
5	本次检查共发现翼墙、耳墙病害 2 处, 均为破损
6	整体箱梁存在蜂窝、麻面 10 处, 分布于 9 根构件上, 总面积为 15.31 m ²
7	T 梁表层混凝土存在剥落、掉角现象共 12 处, 总面积为 121.26 m ²
8	混凝土表面蜂窝麻面, 共 4 处, 总面积为 14.74 m ² , 单处面积介于 0.3~9.0 m ² 之间
9	锥坡存在缺陷 1 处, 总面积为 12.0 m ²
10	小箱梁存在横向裂缝 130 条, 分布于 22 根构件上, 横向裂缝总长度为 44.8 m

机选择了 100 和 50 个句子作为训练集, 并随机选择 35 个句子作为验证集, 其余 1 491 个句子为测试集。表 3 列出了不同检测信息类别在训练集、验证集和测试集中的分布情况。

表 3 训练集、验证集、测试集中信息分布数量
Table 3 Number of information distributions in training sets, validation sets, and test sets

关键信息	关键信息数量/个			
	测试集	验证集	训练集	
			样本量为 100 个	样本量为 50 个
桥梁构件	1 801	49	157	84
构件位置	729	14	87	55
病害	3 110	76	205	109
病害特征	888	26	73	42
病害数量	1 646	36	90	38
病害分布	313	8	25	13
测量类别	976	22	74	37
测量值	1 021	21	76	37
总计	10 484	252	787	415

通过网络爬虫和人工收集的方法收集了来自百科、书籍、报告、期刊、规范的 1 万余条信息构成了桥梁领域语料库, 其中包含了梁式桥、拱桥、斜拉桥、悬索桥等多种形式桥梁的信息。此外, 收集和整理了 472 句检测文本信息, 并依据上述标注方法进行了标注, 构成了检测信息提取任务数据集, 以供任务微

调使用。

3.2 超参数设置

试验训练过程使用 AdamW 优化算法,算法内的超参数设置为 0.9 和 0.999,学习率设置为 1×10^{-3} ,每次训练迭代批量大小设置为 32。由于样本量较少,为防止模型发生过拟合,采用 Dropout 机制并设置丢失率设置为 0.5。同时,采用 L2 正则化并设置参数为 0.001。

为防止训练开始时学习率过大易导致网络不稳定/振荡,采用热身率为 0.1 的热身学习率法^[29],利用最初迭代轮次逐渐提高学习率,一旦网络相对稳定,便使用预先设定学习率训练网络,从而实现更快的收敛性能。

领域迁移和任务迁移时模型也均采用 AdamW 优化算法,训练学习率设置均为 5×10^{-5} 。领域迁移的每次训练迭代批量大小设置为 8,任务迁移的每次训练迭代批量大小设置为 16。由于预训练的数据量较大,设置总迭代次数为 200,以减少模型训练耗时。为防止模型过拟合,均采用 L2 正则化,并设置权重衰退系数为 0.01。

4 试验结果和讨论

4.1 有效性分析

图 4 为所提出方法在样本量为 100 和 50 个的数据上进行训练时的损失变化过程,其分别在迭代 200 和 250 轮次左右趋于收敛,均在约 300 轮迭代后达到最佳性能。

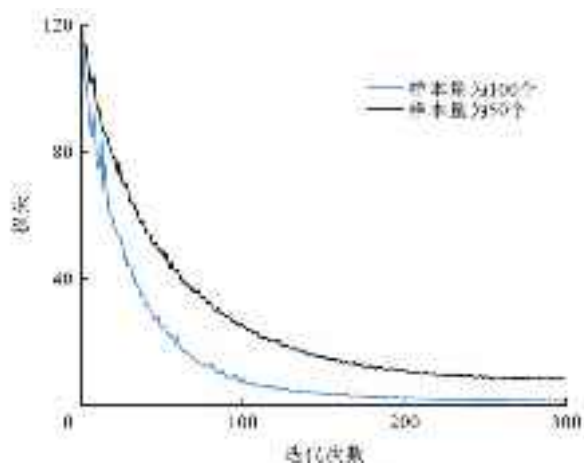


图 4 训练和验证过程的网络损失

Fig. 4 Network loss during training and validation processes

表 4 列出了使用样本量为 100 和 50 个的文本进行训练后模型的性能。使用样本数量为 100 个的数据集训练后的模型性能较好,精准度、召回率和 F_1 值分别可以达到 0.835 1、0.889 1 和 0.860 7,这表明所

提出模型仅使用少量数据也能实现对检测文本高效精准的关键信息提取,说明模型具有一定的泛化能力。将训练数据量减半至 50 个时, F_1 值仅减少了 0.040 5,表明该模型鲁棒性较高,可有效缓解信息提取任务中因样本集数量受限而导致的性能下降问题。

表 4 模型在不同少样本数据下的性能

Table 4 Model performance under different few-shot datasets

样本量/个	精准度	召回率	F_1 值
50	0.798 5	0.844 8	0.820 2
100	0.835 1	0.889 1	0.860 7

在 2 种少样本场景下,模型召回率都大于精准度,表明模型能够识别出更多正确的关键信息。本文认为这比高精度更有价值,因为在信息提取过程中一旦遗漏了某些关键缺陷或特征,就会导致结构风险评估的不全面,因此,提取更多正确的关键信息对于后续工程应用至关重要。这也进一步表明所提出的模型适用于提取桥梁检测关键信息这一实际工程任务。

此外,对模型收敛时间进行分析,以探究模型效率。本次运算电脑采用英特尔 i5-12600KF CPU 和英伟达 GeForce RTX 3080 GPU。当样本量为 100 个时,模型的训练时间大约为 8 980.65 s(约 2 h)。而在样本量减少至 50 个时,模型的收敛速度却相对减慢,导致训练时间增至 15 242.3 s(约 4 h)。但相较于标注成千上万个样本所需的大量时间与人力成本,所提出的方法在提升训练效率方面仍显示出明显的优越性与实用性。

4.2 不同桥检信息的提取性能分析

表 5 列出模型在 2 个少样本数据集下不同关键信息类别的识别性能:随着样本量减少到 50 个,模型在识别构件位置类关键信息时,性能存在较大下降, F_1 值下降了约 20.75%;对于桥梁构件类关键信息,模型性能下降了约 7.66%;然而,对于其余关键信息,模型的性能下降均在 5% 以下,其中病害分布和病害特征 2 类关键信息更是变化不足 1%。这些结果说明模型在少样本数据上的泛化能力较强,即使减少一半的训练数据量,仍能保证对多数关键信息类别的高效准确地提取。

由于样本量少,使得模型提取关键信息特征的难度增大,这要求模型具备更强的特征挖掘能力。图 5 为提出模型在测试集上的标签级混淆矩阵。非关键信息的标签 O 数量远大于其余标签,但模型对该类信息仍能达到 91% 的准确率,表明提出模型能够较好区分关键信息和非关键信息,从而利用

表 5 提出模型对于不同检测信息的提取性能
Table 5 Extraction performance of proposed model for various types of inspection information

关键信息类别	F ₁ 值		变化率/%
	样本量为 50 个	样本量为 100 个	
桥梁构件	0.278 0	0.788 2	7.66
构件位置	0.492 8	0.621 8	20.75
病害	0.839 9	0.875 8	4.10
病害特征	0.768 1	0.768 7	0.08
病害数量	0.953 0	0.970 2	1.77
病害分布	0.981 0	0.981 0	0.00
测量类别	0.932 9	0.975 7	4.39
测量值	0.920 6	0.949 0	2.99

显著差异特征使蕴含在文本中的信息被准确提取。

根据图 5,模型对病害数量、病害分布、测量类别、测量值等具有明显任务特征的信息具备较好的提取效果。例如,测试集上总计有 317 个字为 B-病害分布,388 个字为 I-病害分布,模型识别的准确率

分别高达 97.5%和 95.4%。表明任务迁移时使用的具有明显检测信息特征的数据提高了模型信息挖掘与关键特征提取能力。同时,针对桥梁构件、病害、病害特征等包含大量桥梁领域术语的信息,模型仍能够实现准确的识别。例如,测试集上总计 3 280 个字为 B-病害,3 872 个字为 I-病害,模型识别准确率分别为 87.2%和 88.8%。表明该模型具备对桥梁领域相关信息的快速提取能力,反映了领域迁移对于少样本情况下关键信息提取任务的重要性。

根据图 5 发现识别错误主要发生在构件位置这类信息,且常发生在构件位置的 B 和 I 标签以及桥梁构件的相关标签上。本文认为这是由于训练样本的数量限制和该类信息较强的复杂性导致的。此外,由于该类信息常伴随桥梁构件出现,且它们在实际文本中有着相似的上下文情景。上述诸多原因共同导致了该类信息易被错误识别。

4.3 与基线模型的对比分析

为了验证所提模型在少样本数据中的优越性,考

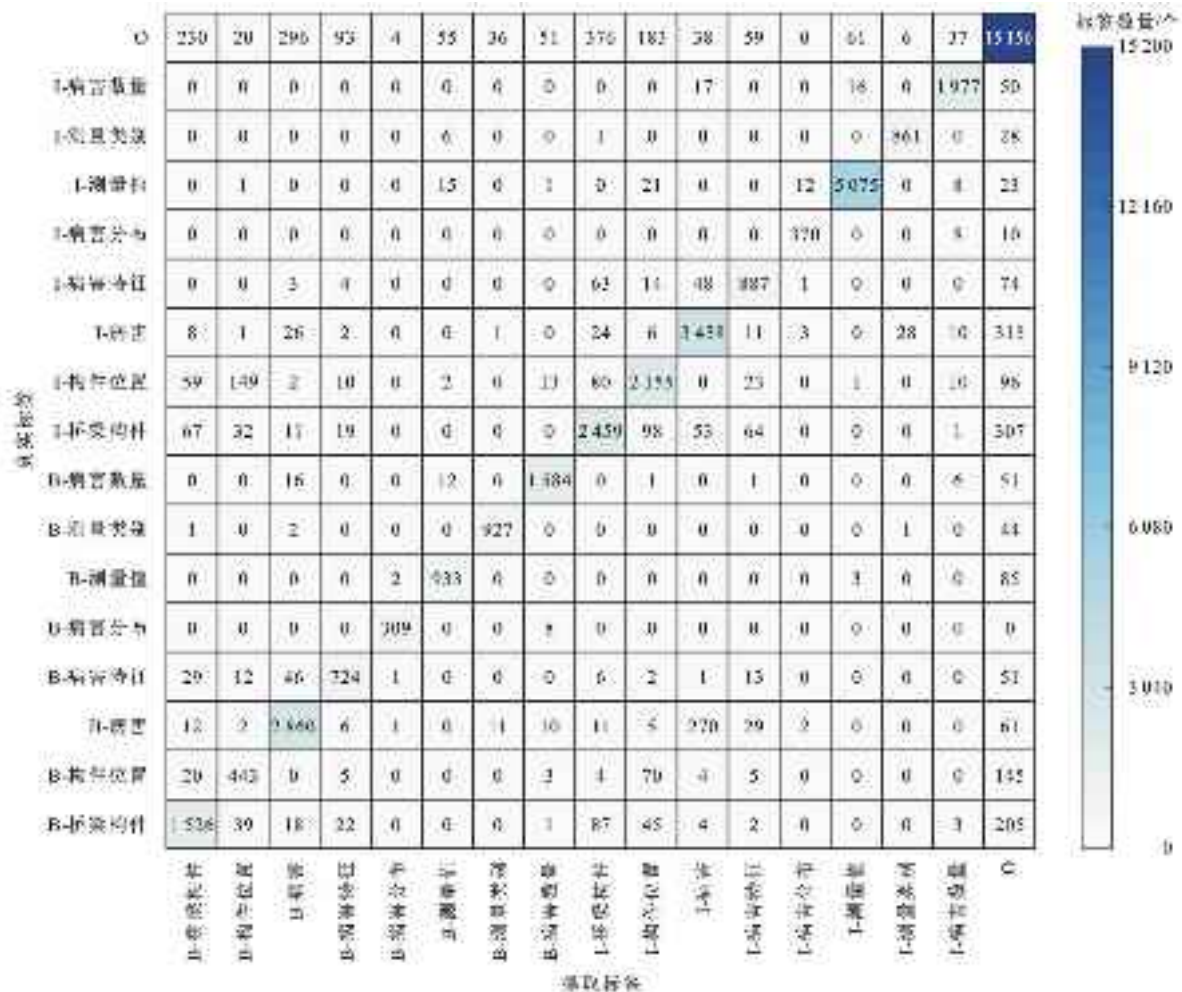


图 5 提出模型在测试集上的混淆矩阵
Fig. 5 Confusion matrix of proposed model on test set

虑将其与已有研究的 Word2Vec-CNN、Word2Vec-BiLSTM、Word2Vec-BiLSTM-CRF 和目前流行的 BERT-BiLSTM-CRF 模型进行对比分析。设定所有模型的维度大小均为 128,以确保对比的公平性。

(1) Word2Vec-CNN。Word2Vec 模型根据具有相似语境的词语具有相似的含义这一原理,将文本转换到高维向量空间,从而实现依据上下文用法就可以捕捉字间的语义关系。该模型获得的文本嵌入是通过浅层神经网络学习而得,这使其在捕捉语法和语义关系方面高效而强大。卷积神经网络(Convolutional Neural Network, CNN)作为特征提取网络,通过卷积层有效捕捉局部和位置上下文信息,因此, Wang 等^[13]采用 Word2Vec-CNN 作为基本模型,从建筑安全法规中提取信息。

(2) Word2Vec-BiLSTM。由于文本是顺序数据, CNN 很难捕捉到这种序列特性。善于处理这种数据的循环神经网络可以很好捕捉文本中依赖关系,但它们在处理长距离依赖关系时效果不佳。BiLSTM 是对其的一种改进,它通过正、逆 2 个方向来解决这一问题,从而增强了模型理解上下文能力,提高信息识别准确性,因此, Moon 等^[12]采用 Word2Vec-BiLSTM 模型来识别检测报告中的涉及桥梁病害的相关信息。

(3) Word2Vec-BiLSTM-CRF。Word2Vec-BiLSTM 模型通常使用 Softmax 函数进行信息类别预测,较难

捕捉序列标签间相互依赖关系,因此, Zhang 等^[14,10]在该模型中加入了 CRF 层,通过挖掘序列中相邻标签间序列依赖关系以提高模型性能,从而确保实体预测更加连贯,上下文更加一致。

(4) BERT-BiLSTM-CRF。结合了 BERT 深度上下文嵌入和 BiLSTM 序列数据处理能力,能有效捕捉上下文的细微差别和序列中依赖关系。在此基础上,增加 CRF 层,有助于通过考虑标签依赖关系来修正识别结果,从而实现更准确的信息识别,因此, BERT-BiLSTM-CRF 模型代表了目前最为流行的方法^[30]。

表 6 和图 6 展示了在样本量分别为 100 和 50 个的数据集上本文提出模型与 4 个基线模型的性能情况,表中: K 为 P 与 R 的比值。在样本量为 100 和 50 个的数据集上,提出模型表现最好,将基线模型中表现最好的 BERT-BiLSTM-CRF 模型的综合性能指标 F_1 值分别提高了 5.14% 和 11.12%。现有研究方法的模型中 Word2Vec-BiLSTM 模型表现最好,但是 F_1 值仍在 0.8 以下,该性能很难满足实际的工程需求。同时,提出模型在样本量为 100 个的数据集上高于该模型 9.14%;在样本量为 50 个的数据集上高于现有研究方法中表现最好的 Word2Vec-BiLSTM-CRF 模型 14.22%。这些结果进一步说明了提出模型能够有效缓解少样本数据对桥检信息提取模型性能的不利影响。

表 6 提出模型与基线模型的性能对比

Table 6 Performance comparison between proposed model and baseline models

模型名称	精准度			召回率			F_1 值			K
	样本量为 50 个	样本量为 100 个	变化率/ %	样本量为 50 个	样本量为 100 个	变化率/ %	样本量为 50 个	样本量为 100 个	变化率/ %	
Word2Vec-CNN	0.425 0	0.468 1	9.21	0.488 6	0.520 0	6.03	0.449 5	0.487 2	7.75	0.82
Word2Vec-BiLSTM	0.703 9	0.770 8	8.67	0.754 7	0.808 9	6.70	0.693 5	0.788 6	12.06	0.75
Word2Vec-BiLSTM-CRF	0.685 8	0.737 5	7.02	0.761 5	0.809 2	5.89	0.718 1	0.764 1	6.01	0.95
BERT-BiLSTM-CRF	0.803 1	0.810 7	0.94	0.682 9	0.827 8	17.51	0.738 1	0.818 6	9.83	7.05
提出模型	0.798 5	0.835 1	4.38	0.844 8	0.889 1	4.98	0.820 2	0.860 7	4.71	1.17

为分析模型的鲁棒性,对比样本量发生变化时各评价指标的变化率。根据表 6 和图 6 可知,当样本量从 100 个减少到 50 个时,提出模型的召回率和 F_1 值的变化率最低,全部指标的变化率均小于 5%,表明提出模型在少样本检测信息提取任务上具有较强的鲁棒性,模型性能较为稳定。然而,基线模型中表现最佳的 BERT-BiLSTM-CRF 模型,数据量在 100 个时,模型性能超过 0.8,但当样本量下降至 50 个时,虽然模型精准度几乎未发生变化,但是

F_1 值的降低了 9.83%,召回率更是降低了 17.51%。这说明 BERT-BiLSTM-CRF 模型在少样本数据集上的鲁棒性较差,较大的样本量变化会引起模型多项指标的剧烈变动。此外,根据前面的分析,召回率的稳定对于检测信息提取任务而言更为关键,因此,该模型不适用于这类任务。现有研究的模型中表现最佳的 Word2Vec-BiLSTM 模型,当数量样本量发生变化时,模型的 F_1 值发生最高的变化率,达到了 12.06%。在实际工程应用时,由于工程

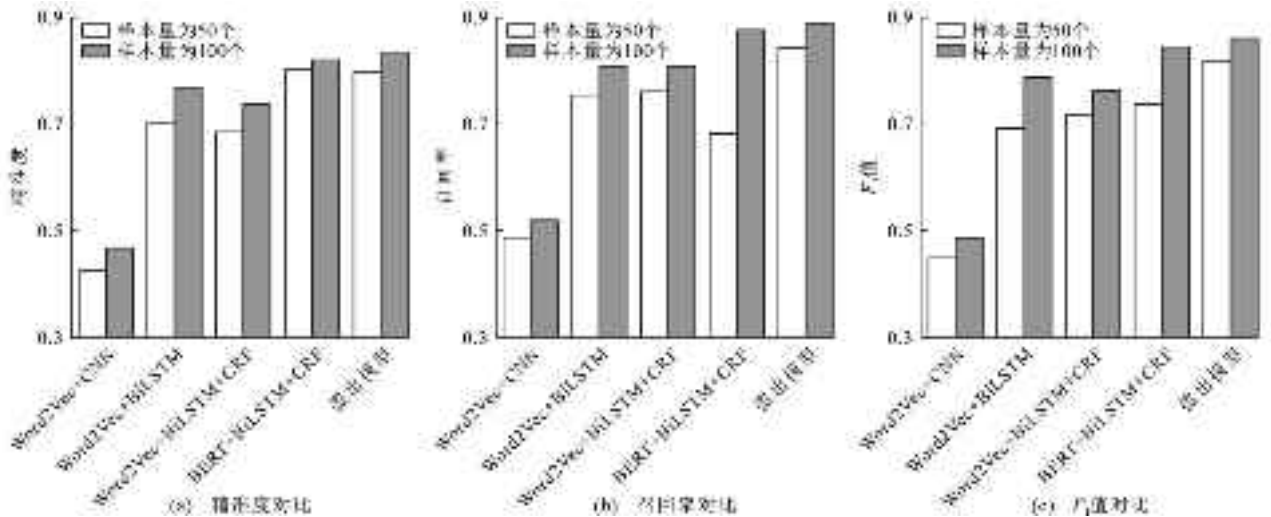


图6 提出模型与基线模型的性能对比

Fig. 6 Performance comparison between proposed model and baseline models

条件的随机性,无法保证可用标注数据的数量是稳定的,这种随样本量变化模型性能发生较大变动的情况,在实际工程中存在较大挑战。

对于桥检信息提取任务而言,模型的召回率相比精准度更为重要。对于多数模型而言,在相同数据量的情况下,模型的召回率都高于精准度,这对于实际任务而言是有价值的。但是,实际工程中数据量是无法一直保持不变,当数据量变化时能否稳定的保持这种特性对于该任务而言也是重要的,这种特性体现了模型的实际应用功能,本文定义为功能性。为此,引入功能性指标 K 来衡量这种特性

$$K = \frac{P}{R} = \frac{R_{100} - P_{100}}{R_{50} - P_{50}} \quad (17)$$

式中: R_{100} 和 P_{100} 分别为样本量为 100 时的模型召回率和精准度; R_{50} 和 P_{50} 分别为样本为 50 时模型召回率和精准度。

根据表 6 的计算结果,提出模型和 Word2Vec-BiLSTM-CRF 模型的比值接近于 1,说明这 2 个模型功能性上相较于其他模型更优越。而基线模型中性能指标最佳的 BERT-BiLSTM-CRF 模型的比值高达 7.05,说明即使是基线模型中性能最好的模型其功能性依然较差,并不适合应用于桥检关键信息的提取任务中,因此,开发一种专门针对于少样本的桥检信息提取模型是必要的。

根据上述分析,提出的模型在识别性能、鲁棒性、功能性上均优于多数基线模型,这进一步证明了其在少样本的桥检信息提取任务上的有效性和优越性。

4.4 消融试验

采用移除领域、任务迁移、BiLSTM 层和 CRF 层,分别探究各组件对模型整体的作用和贡献。消

融试验的基本思想是系统地移除模型的某些部分,观察这些变化对性能的影响,以更好地理解各组件的作用。

首先,对领域迁移开展消融试验,根据表 7 与图 7,领域和任务迁移模型 (BERT-Inspection-BiLSTM-CRF) 相比仅任务迁移模型 (BERT-Inspection (no domain)-BiLSTM-CRF) 表现更优,其中 F_1 值分别提高了 2.30% 和 4.63%,且在识别桥梁构件、构件位置和病害等具有较强桥梁领域特征的信息时有明显的性能提升,如构件位置这类信息性能提升了近 26.5%。表明领域迁移对于桥检信息提取任务而言可提高模型对于桥梁专业信息的快速挖掘能力,能有效提高专业领域信息提取性能。

对任务迁移进行消融试验,如图 7 所示,经过任务迁移的 BERT-Inspection (no domain)-BiLSTM-CRF 模型相比 BERT-BiLSTM-CRF 原始模型,在样本量分别为 100 和 50 个的情况下,模型 F_1 值分别提高了 2.72% 和 6.21%,表明任务迁移能有效提高少样本下桥检信息的提取性能。此外,根据表 8,原始模型无法对于病害分布正确提取,而经过任务迁移后的模型则能高效地识别该类信息,并将 F_1 值维持在 0.98 以上。同时,经过任务迁移后的模型对于病害特征、病害数量、测量类别和测量值这 4 类信息也有一定性能提升,表明任务迁移对于这类具有明显任务特性的信息具有较好的识别效果,能够提高模型快速挖掘该类信息特征的能力。

对 BiLSTM 层进行消融试验,由表 7 可知,相比于 BERT-Inspection-CRF 模型, BERT-Inspection-BiLSTM-CRF 在样本量为 100 和 50 个的情况下,模型 F_1 值分别提高了 2.50% 和 9.92%;由表 8 可知,

表 7 消融试验中模型的性能

Table 7 Model performances in ablation experiment

模型名称	精准度			召回率			F ₁ 值		
	样本量为 50 个	样本量为 100 个	变化率/ %	样本量为 50 个	样本量为 100 个	变化率/ %	样本量为 50 个	样本量为 100 个	变化率/ %
BERT-Inspection-BiLSTM	0.800 9	0.849 4	5.72	0.839 2	0.861 9	2.63	0.818 6	0.855 2	4.28
BERT-Inspection-CRF	0.715 8	0.802 0	10.70	0.782 7	0.881 3	11.20	0.746 2	0.839 8	11.10
BERT-BiLSTM-CRF	0.803 1	0.810 7	0.94	0.682 9	0.827 8	17.51	0.738 1	0.818 6	9.83
BERT-Inspection (no domain)-BiLSTM-CRF	0.780 2	0.821 8	5.06	0.791 6	0.862 2	8.19	0.783 9	0.840 9	6.78
BERT-Inspection-BiLSTM-CRF	0.798 5	0.835 1	4.38	0.844 8	0.889 1	4.98	0.820 2	0.860 7	4.71

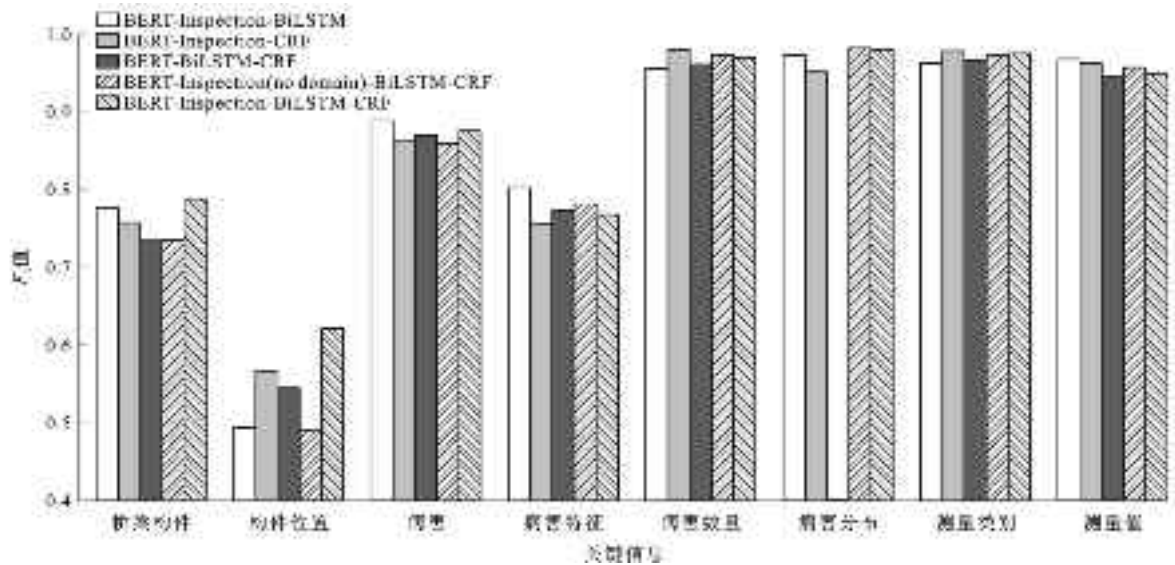


图 7 消融试验中各模型对关键桥检信息的识别性能

Fig. 7 Recognition performances of various models on key bridge inspection information in ablation experiment

表 8 提出模型在消融试验中对不同关键信息的 F₁ 值

Table 8 F₁ scores for different key information achieved by proposed model in ablation experiment

关键信息	BERT-Inspection-BiLSTM	BERT-Inspection-CRF	BERT-BiLSTM-CRF	BERT-Inspection (no domain)-BiLSTM-CRF	BERT-Inspection-BiLSTM-CRF
桥梁构件	0.776 7	0.757 9	0.735 0	0.735 6	0.788 2
构件位置	0.494 8	0.567 6	0.545 3	0.492 1	0.621 8
病害	0.889 1	0.864 3	0.870 1	0.860 2	0.875 8
病害特征	0.803 5	0.757 3	0.773 2	0.782 0	0.768 7
病害数量	0.957 3	0.981 0	0.962 4	0.973 7	0.970 2
病害分布	0.974 5	0.952 2	0.193 7	0.982 5	0.981 0
测量类别	0.962 7	0.979 0	0.968 0	0.973 5	0.975 7
测量值	0.968 3	0.962 8	0.947 3	0.958 0	0.949 0

BiLSTM 的增加显著提高了模型对于构件位置这类需要结合长语义关系才能准确提取的信息的准确率,其对于桥梁构件、病害和病害特征,这些领域信息也具备较好的提取效果。这表明 BiLSTM 通过充分挖掘长文本的依赖关系从而实现了对于桥检报告这类具有复杂长语句文本蕴含信息的精准提取,

可以在少样本情况下提高模型的性能。此外,当样本量减少时,没有添加 BiLSTM 的模型的性能指标变化较大,F₁ 值减小了 11.1%,表明 BiLSTM 对于模型整体稳定性有着积极的作用。

对 CRF 层进行消融试验评估,根据图 7 和表 8 所示,添加 CRF 层对模型 BERT-Inspection-BiLSTM-

CRF 在提取桥梁构件、构件位置、病害数量、病害分布和测量类别这 5 类具有明显语义特性的关键信息时,相较于 BERT-Inspection-BiLSTM 均有显著提升,这说明使用 CRF 能够在一定程度上提高模型预测标签类别的正确率,减少了信息边界的错误识别,使得模型在复杂信息的提取上面具有较好的性能。

上述结果证明了提出模型各组件的有效性。双阶段迁移策略将大量领域和任务相关的信息特征嵌入在预训练模型内的参数信息,实现了对桥梁相关领域信息和任务显著特征的快速提取和挖掘;BiLSTM 通过挖掘双向依赖关系,实现了对单句包含较多信息的桥检文本的高效和稳定提取;CRF 通过减少信息边界的错误识别,提高了模型在少样本情况下信息提取性能。

4.5 实例分析

对于信息抽取任务,识别信息的边界至关重要,不仅需要信息中间部分提取正确,还需要 B 和最后一个 I 同时正确,才能认为该信息提取成功。表 9 为 5 个真实桥检报告中的句子的提取结果,其中:be 代表桥梁构件,bl 代表构件位置,dd 代表病害特征,d 代表病害,q 代表病害数量,s 代表病害分布,mc 代表测量类别,m 代表测量值。根据表 5,发现所提出的方法在包含多样化的桥型及构件的文本上表现出色,能够精确地提取出桥梁领域的专业术语,如主拱圈、湿接缝、支座和主塔等,以及检测领域的特定信息,包括裂缝、剥落、掉角和串动。此外,对于病害数量、分布和测量等关键数值信息,该方法同样展现出了良好的提取能力。

表 9 提出模型在 5 个实例的信息提取结果

Table 9 Information extraction results of proposed model in five instances

实例 1	文本输入	主拱圈存在横向裂缝 1 条,裂缝长度为 1.1 m,裂缝宽度为 0.12 mm
	模型输出	[B-be, I-be, I-be, O, O, B-dd, I-dd, B-d, I-d, B-q, I-q, O, B-d, I-d, B-mc, I-mc, O, B-m, I-m, I-m, I-m, O, B-d, I-d, B-mc, I-mc, O, B-m, I-m, I-m, I-m, I-m, O]
	提取结果	桥梁构件为主拱圈;病害特征为横向;病害为裂缝;病害数量为 1 条;测量类别为长度,测量值为 1.1 m;测量类别为宽度,测量值为 0.12 mm
实例 2	文本输入	湿接缝存在剥落、掉角 6 处,分布于 6 根构件上,总面积为 0.85 m ²
	模型输出	[B-be, I-be, I-be, O, O, B-d, I-d, O, B-d, I-d, B-q, I-q, O, O, O, O, B-s, I-s, O, O, O, O, O, B-mc, I-mc, O, B-m, I-m, I-m, I-m, I-m, I-m, O]
	提取结果	桥梁构件为湿接缝;病害 1 为剥落;病害 2 为掉角;病害数量为 6 处;病害分布为 6 根;测量类别为面积,测量值为 0.85 m ²
实例 3	文本输入	支座存在串动和脱空 58 个,其中脱空 58 个,脱空率介于 15.0%~50.0%
	模型输出	[B-be, I-be, O, O, B-d, I-d, O, B-d, I-d, B-q, I-q, O, O, O, B-d, I-d, B-q, I-q, O, B-d, I-d, I-d, O, O, B-m, I-m, I-m, I-m, I-m, I-m, I-m, I-m, I-m, O]
	提取结果	桥梁构件为支座;病害为串动;病害为脱空;病害数量为 58 个;测量类别为脱空率,测量值为 15.0%~50.0%
实例 4	文本输入	排水系统存在排水孔堵塞、排水不畅现象,共 3 处
	模型输出	[B-be, I-be, I-be, I-be, O, O, B-d, I-d, I-d, I-d, I-d, O, B-d, I-d, I-d, I-d, O, O, O, O, B-q, I-q, O]
	提取结果	桥梁构件为排水系统;病害①为排水孔堵塞;病害②为排水不畅;病害数量为 3 处
实例 5	文本输入	主塔纵向阻尼器存在油漆脱落现象,共 1 处,面积为 0.02 m ²
	模型输出	[B-be, I-be, B-dd, I-dd, B-d, I-d, I-be, O, O, B-dd, I-dd, B-d, I-d, O, O, O, O, B-q, I-q, O, B-mc, I-mc, O, B-m, I-m, I-m, I-m, I-m, I-m, O]
	提取结果	桥梁构件为主塔;构件位置为纵向阻尼器;病害特征为油漆;病害为脱落;病害数量为 1 处;测量类别为面积,测量值为 0.02 m ²

在训练集中仅有 5 个样本与拱桥相关,且“主拱圈”一词在训练集中未曾出现。尽管如此,提出方法对第 1 个实例仍实现了精准识别,这体现了领域迁移在少样本下信息提取任务的有效性,其显著提升了模型对专业术语的挖掘和识别性能。在第 2 个实例中,该方法能够准确提取出同时出现的不同病害信息,即“剥落”和“掉角”。第 3 个实例表明,通过在任务迁移阶段对具有明显结构特征的复杂数字信息

进行深入学习,模型在处理如“15.0%~50.0%”此类复杂和少见的数字信息时,展现出优异的性能。在第 4 个实例中,借助 BiLSTM 捕获的长距离依赖信息,方法能准确区分“排水系统”作为桥梁构件与“排水孔堵塞、排水不畅”作为病害信息,凸显了模型对于复杂桥检文本的出色识别能力。尽管第 5 个实例中的斜拉桥在训练数据中未曾出现,方法依然能识别“主塔”这一核心构件。在识别“阻尼器”时,出

现了错误识别,方法将“阻尼”误识别为病害,将“器”误识别为结构构件。这可能是由于预训练数据中包含“阻碍”等类似病害信息,使方法错误将“阻尼”识别为病害。这类少见的信息,后续将加强对该类信息的迁移学习,以进一步提升信息提取效果。

上述分析表明,虽然桥检报告的文本在自然语言形式上具有高度异质性,但是提出方法在复杂句子的误差有限,充分证明了所提方法的实用性和可行性。

5 结 语

(1)提出模型在仅使用100和50个标注样本进行训练的情况下,可以在包含1491个样本的测试集上实现高精度的桥梁检测信息识别, F_1 值分别高达0.8607和0.8202,均优于既有方法。

(2)提出的模型通过领域预训练和任务微调,实现了可靠的领域和任务迁移,在训练样本量发生变化时模型性能变化较小,对比全部基线模型,提出的模型具有较好的鲁棒性。

(3)采用的领域迁移和任务迁移的预训练方法,可以提高模型在少样本上快速捕捉领域信息和任务特征的能力,实现了快速提取和挖掘领域和任务相关特征,提高了模型在少样本数据下的信息提取性能。

(4)未来将针对数据稀疏、丢失、错误等缺陷问题持续优化,并考虑引入文本生成模型、桥检领域本体和对抗训练等多种技术综合提高模型的鲁棒性,进一步提升其在实际工程中的应用范围和使用性能。

参考文献:

References:

- [1] XIA Ye, LEI Xiao-ming, WANG Peng, et al. A data-driven approach for regional bridge condition assessment using inspection reports[J]. Structural Control and Health Monitoring, 2022, 29(4): e2915.
- [2] YANG Jian-xi, YANG Xiao-xia, LI Ren, et al. BERT and hierarchical cross attention-based question answering over bridge inspection knowledge graph[J]. Expert Systems with Applications, 2023, 233: 120896.
- [3] JIANG Ya-li, YANG Gang, LI Hai-jiang, et al. Knowledge driven approach for smart bridge maintenance using big data mining[J]. Automation in Construction, 2023, 146: 104673.
- [4] YANG Jian-xi, XIANG Fang-yue, LI Ren, et al. Intelligent bridge management via big data knowledge engineering[J]. Automation in Construction, 2022, 135: 104118.
- [5] KALE A, RICKS B, GANDHI R. New measure to understand and compare bridge conditions based on inspections time-

series data[J]. Journal of Infrastructure Systems, 2021, 27(4): 04021037.

- [6] WANG Y C, CAI C S, HAN B, et al. A deep learning-based approach for assessment of bridge condition through fusion of multi-type inspection data[J]. Engineering Applications of Artificial Intelligence, 2024, 128: 107468.
- [7] YANG Xiao-xia, YANG Jian-xi, LI Ren, et al. Complex knowledge base question answering for intelligent bridge management based on multi-task learning and cross-task constraints[J]. ENTROPY, 2022, 24(12): 1805.
- [8] LIU Heng, ZHANG Yun-feng. Bridge condition rating data modeling using deep learning algorithm[J]. Structure and Infrastructure Engineering, 2020, 16(10): 1447-1460.
- [9] LIU Kai-jian, EL-GOHARY N. Ontology-based semi-supervised conditional random fields for automated information extraction from bridge inspection reports[J]. Automation in Construction, 2017, 81: 313-327.
- [10] LI T S, ALIPOUR M, HARRIS D K. Context-aware sequence labeling for condition information extraction from historical bridge inspection reports[J]. Advanced Engineering Informatics, 2021, 49: 101333.
- [11] LI Ren, MO Tian-jin, YANG Jian-xi, et al. Bridge inspection named entity recognition via bert and lexicon augmented machine reading comprehension neural model[J]. Advanced Engineering Informatics, 2021, 50: 101416.
- [12] MOON S, CHUNG S, CHI S. Bridge damage recognition from inspection reports using NER based on recurrent neural network with active learning[J]. Journal of Performance of Constructed Facilities, 2020, 34(6): 04020119.
- [13] WANG X Y, EL-GOHARY N. Deep learning-based named entity recognition and resolution of referential ambiguities for enhanced information extraction from construction safety regulations[J]. Journal of Computing in Civil Engineering, 2023, 37(5): 04023023.
- [14] ZHANG R C, EL-GOHARY N. A deep neural network-based method for deep information extraction using transfer learning strategies to support automated compliance checking[J]. Automation in Construction, 2021, 132: 103834.
- [15] 贺拴海,王安华,朱 钊,等.公路桥梁智能检测技术研究进展[J].中国公路学报,2021,34(12):12-24.
- [16] HE Shuan-hai, WANG An-hua, ZHU Zhao, et al. Research progress on intelligent detection technologies of highway bridges[J]. China Journal of Highway and Transport, 2021, 34(12): 12-24.
- [17] MOON S, LEE G, CHI S. Automated system for construction specification review using natural language processing[J]. Advanced Engineering Informatics, 2022, 51: 101495.
- [18] LI Chuan-jiang, LI Shao-bo, WANG Huan, et al. Attention-based deep meta-transfer learning for few-shot fine-grained fault diagnosis[J]. Knowledge-Based Systems, 2023, 264: 110345.
- [19] ZHANG Hu, GUO Jia-yu, WANG Yu-jie, et al. Judicial

- nested named entity recognition method with mrc framework[J]. International Journal of Cognitive Computing in Engineering, 2023, 4: 118-126.
- [19] MOIRANGTHEM D S, LEE M. Hierarchical and lateral multiple timescales gated recurrent units with pre-trained encoder for long text classification[J]. Expert Systems with Applications, 2021, 165: 113898.
- [20] HU Zhong-jian, YANG Peng, LI Bing, et al. Biomedical extractive question answering based on dynamic routing and answer voting[J]. Information Processing and Management, 2023, 60(4): 103367.
- [21] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[J]. arXiv, 2017, DOI: 10.48550/arXiv.1810.04805.
- [22] LIU K J, EL-GOHARY N. Semantic neural network ensemble for automated dependency relation extraction from bridge inspection reports[J]. Journal of Computing in Civil Engineering, 2021, 35(4): 04021007.
- [23] LIU K J, EL-GOHARY N. Improved similarity assessment and spectral clustering for unsupervised linking of data extracted from bridge inspection reports[J]. Advanced Engineering Informatics, 2022, 51: 101496.
- [24] 劳武略, 崔 闯, 张登科, 等. 基于计算机视觉的钢桥面板裂纹识别方法[J]. 中国公路学报, 2023, 36(3): 188-201.
- LAO Wu-lue, CUI Chuang, ZHANG Deng-ke, et al. Crack identification method of orthotropic steel deck based on computer vision[J]. China Journal of Highway and Transport, 2023, 36(3): 188-201.
- [25] 景 强, 郑顺潮, 梁 鹏, 等. 港珠澳大桥智能化运维技术与工程实践[J]. 中国公路学报, 2023, 36(6): 143-156.
- JING Qiang, ZHENG Shun-chao, LIANG Peng, et al. Technologies and engineering practices of intelligent operation and maintenance of Hong Kong-Zhuhai-Macao Bridge[J]. China Journal of Highway and Transport, 2023, 36(6): 143-156.
- [26] VO S N, VO T T, LE B. Interpretable extractive text summarization with meta-learning and bi-lstm: a study of meta learning and explainability techniques [J]. Expert Systems with Applications, 2024, 245: 123045.
- [27] LI Dai-yi, YAN Li, YANG Jian-zhong, et al. Dependency syntax guided BERT-BiLSTM-GAM-CRF for chinese NER[J]. Expert Systems with Applications, 2022, 196: 116682.
- [28] ZHOU Shang-lian, SONG Wei. Deep learning-based roadway crack classification using laser-scanned range images: a comparative study on hyperparameter selection[J]. Automation in Construction, 2020, 114: 103171.
- [29] LAMBRECHTS G, DE GEETER F, VECOVEN N, et al. Warming up recurrent neural networks to maximise reachable multistability greatly improves learning [J]. Neural Networks, 2023, 166: 645-669.
- [30] LIU Ya-fei, WEI Si-qi, HUANG Hai-jun, et al. Naming entity recognition of citrus pests and diseases based on the BERT-BiLSTM-CRF model[J]. Expert Systems with Applications, 2023, 234: 121103.