

中文引用格式:陈述,王典学,杨应柳,等. 水电工程施工安全隐患语义匹配模型[J]. 中国安全科学学报,2024,34(12):40-47.

英文引用格式:CHEN Shu, WANG Dianxue, YANG Yingliu, et al. Semantic matching model of potential safety hazards in hydroelectric project construction [J]. China Safety Science Journal, 2024, 34(12): 40-47.

水电工程施工安全隐患语义匹配模型*

陈述^{1,2}教授, 王典学^{1,2}, 杨应柳^{**3}, 曹坤煜^{1,2}, 聂本武^{2,4}

(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 重庆大学 管理科学与房地产学院, 重庆 400044; 4 国家能源投资集团有限责任公司, 四川 成都 610095)

中图分类号: X948

文献标志码: A

DOI: 10.16265/j.cnki.issn1003-3033.2024.12.0795

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

【摘要】 为辅助制定水电工程施工安全隐患治理措施, 收集水电工程施工巡检积累的隐患文本, 借助 Python 工具对半结构化的隐患文本进行实体与关系抽取, 构建安全隐患知识图谱, 并导入到 neo4j 图数据库中进行存储; 搭建水电工程施工隐患语义匹配的基于双向编码器表征的句子嵌入 (Sentence-BERT) 模型, 学习目标隐患与历史隐患的深层语义特征, 推荐与目标隐患最相似的历史安全隐患; 利用 Cypher 查询语句, 检索该历史安全隐患对应的治理措施。结果表明: Sentence-BERT 模型对于施工隐患与历史相似隐患的识别准确率为 96.48%, 明显优于双向编码器表征 (BERT) 模型、基于词向量的深度语义匹配模型 (Word2vec-DSSM) 和基于 BERT 的 DSSM 模型 (BERT-DSSM)。在随机抽取的 150 条目标隐患数据中测试历史相似隐患推荐精确度达到 92%, 并通过隐患知识图谱展示隐患治理措施的检索效果, 验证了该方法的适用性和有效性。

【关键词】 水电工程施工; 安全隐患; 治理措施; 智能推荐; 知识图谱; 语义匹配

Semantic matching model of potential safety hazards in hydroelectric project construction

CHEN Shu^{1,2}, WANG Dianxue^{1,2}, YANG Yingliu³, CAO Kunyu^{1,2}, NIE Benwu^{2,4}

(1 Hubei Key Laboratory of Construction and Management in Hydropower Engineering, China Three Gorges University, Yichang Hubei 443002, China; 2 College of Hydraulic & Environmental Engineering, China Three Gorges University, Yichang Hubei 443002, China; 3 School of Management Science & Real Estate, Chongqing University, Chongqing 400044, China; 4 China Energy Investment Co., Ltd., Chengdu Sichuan 610095, China)

Abstract: In order to assist in the development of safety hazard management measures for hydropower project construction, the safety hazard texts accumulated during the construction inspection of hydropower projects were collected. Entities and relationships from the semi-structured safety hazard texts were extracted using Python. A knowledge graph of safety hazards was constructed and imported into the neo4j graph database for storage. A Sentence-Bidirectional Encoder Representations from Transformer (BERT)

* 文章编号: 1003-3033(2024)12-0040-08; 收稿日期: 2024-07-11; 修稿日期: 2024-09-15

** 通信作者: 杨应柳 (1996—), 女, 贵州遵义人, 博士研究生, 主要研究方向为水电工程施工安全。E-mail: YLyang112@126.com。

model based on bidirectional coding was built for the semantic matching of construction hazards in hydropower projects. The deep semantic features of target hazards and historical hazards were learned, and the historical safety hazards most similar to target hazards were recommended. Using the Cypher query statement, the governance measures corresponding to the historical security risk were searched. The results show that the Sentence-BERT model has an accuracy of 96.48% in identifying architecturally and historically similar safety hazards, which is significantly better than BERT, Word2vec-Deep Semantic Similarity Model (Word2vec-DSSM), and BERT-DSSM models. Among 150 randomly selected target safety hazard data, the accuracy rate of testing historical similar safety hazard suggestions reaches 92%, and the retrieval effect of hazard management measures is demonstrated through the hazard knowledge graph, which verifies the applicability and effectiveness of the method.

Keywords: hydropower project construction; safety hazard; semantic matching; management measures; intelligent recommendation; knowledge graph

0 引言

水电工程施工安全隐患多,人员伤亡事故屡有发生^[1]。目前,安全隐患治理措施主要由项目管理人员或邀请专家依据行业标准规范,结合项目实际制定而成,治理措施的有效性依赖于人员经验,制定效率相对较低,亟需通过人工智能技术,从历史隐患中智能推荐出治理措施,为制定隐患治理措施提供辅助支持。

近年来,有关水电工程施工安全隐患的研究多聚焦于隐患智能分类^[2-3]、特征提取^[4-5]、主题挖掘^[6-7]等。如王仁超等^[2]提出一种利用基于整个单词掩码的鲁棒优化双向编码器表征方法混合深度学习模型的隐患文本快速智能分类方法;陈述等^[4]采用短语提取技术,首次描述隐患特征、隐患部位、隐患类型之间的信息流动,挖掘隐患部位与隐患类型的关联规则;肖琪耀等^[6]提出一种基于双向长短期记忆网络、条件随机场和文档主题生成模型的煤矿隐患类型、地点和比重的主题挖掘模型,得到每个工作队全年出现安全隐患的类型、发生地点及比重。以上研究通过自然语言处理技术实现安全隐患的智能化管理,但局限于对安全隐患特征规律的分析,缺乏挖掘安全隐患的治理措施。

科学有效的治理措施是消除隐患的关键,直接关系到施工人员的生命财产安全^[8]。随着施工安全隐患历史文本数据的积累,从海量历史隐患中挖掘与目标隐患相似的隐患,可为当前隐患治理工作提供重要参考。然而,这一过程较为复杂,需要深度理解文本的语义信息并进行匹配^[9]。在自然语言处理技术的发展过程中,基于预训练的深度学习模型在语义匹配中表现出良好的效果,为安全隐患匹

配提供了技术支撑^[10]。当前,基于深度学习的语义匹配技术在医疗^[11]、农业^[12]、法律^[13]等领域应用广泛,较少应用于水电工程施工领域^[14]。此外,知识图谱通过图结构清晰表示各实体及其关系,有助于帮助理解和分析复杂的数据,为从历史隐患中提取与治理措施有关的信息提供了便利^[15]。

鉴于此,笔者拟构建安全隐患知识图谱,在此基础上利用深度学习的语义提取优势,搭建水电工程施工隐患语义匹配的基于双向编码器表征的句子嵌入(Sentence-Bidirectional Encoder Representations from Transformer, Sentence-BERT)网络模型,提取目标隐患与历史隐患的深层语义特征并进行历史相似隐患的智能推荐,运用 Cypher 语句进行隐患治理措施检索,以期为制定当前隐患治理措施提供参考。

1 安全隐患数据及语义匹配模型

1.1 安全隐患数据来源与预处理

从某水电工程施工巡检记录中获取安全隐患文本记录 32 221 条,涵盖触电、文明施工、高处坠落、物体打击等 16 种隐患类型,具有较强代表性。为确保安全隐患数据质量,清洗数据,剔除隐患描述不完整、重复记录以及整改要求缺失的数据,最终得到有效隐患数据 31 845 条。根据时间先后顺序,将发现时间较早的 21 230 条作为历史隐患数据集,发现较晚的 10 615 条作为目标隐患数据集。

为更好地训练模型,提高相似隐患文本对的语义匹配能力,组合历史隐患数据集与目标隐患数据集中相似的隐患,共得到 3 125 条相似隐患文本对作为正样本集。将目标隐患与其非相似的一条历史隐患组建文本对定义为负样本集,并且在非相似的历史隐患选取时,尽可能地选取具有部分相似特征

的历史隐患,据此构建负样本集。最终得到正负样本集共计 6 250 条数据,部分样例见表 1。并将该数据按照 7 : 1.5 : 1.5 划分为训练集、验证集和测试集。

表 1 目标隐患-历史隐患数据集样例

Table 1 Target safety hazard-historical safety hazard data set sample

编号	目标隐患记录	历史隐患记录	标签
1	顶拱挂网施工,汽车吊吊装范围未警戒防护	下岔洞,7号尾水下游钢筋台车吊装无警戒	1
2	5号尾调溜渣井下部未施工时未设置警示警戒	7号尾水连接管二层竖井下部警戒线缺失	1
⋮	⋮	⋮	⋮
3 126	安装间上游排架区,一配电箱有一闸多接现象	进水口 EL744 平台,二级配电箱箱门脱落	0
⋮	⋮	⋮	⋮
6 250	主厂南厂顶交通洞及 5-2 廊道南侧排架竹跳板破损	开挖道路路边警示带破损	0

1.2 安全隐患知识图谱构建

水电工程施工安全隐患文本通过表格形式记录隐患发现时间、地点、隐患描述、类型、整改要求、整改期限等主要内容,属于半结构化文本。为此,以隐患问题这一实体作为核心,通过定义隐患实体、实体

属性以及不同类型实体间的关系,建立隐患实体—关系,具体框架如图 1 所示。

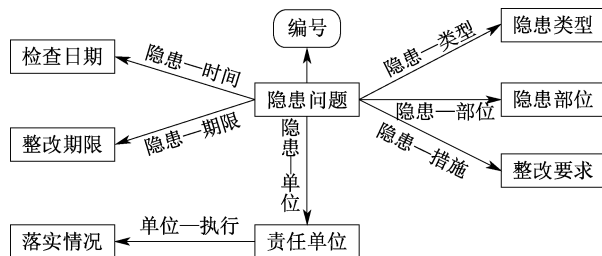


图 1 水电工程施工安全隐患主要实体和关系模式

Fig. 1 Main entities and relationship patterns of construction safety hazards in hydroelectric projects

首先,借助 Python 工具,采用基于规则结合人工修正的方式^[16],通过“将每列定义为一个实体”的规则抽取包括检查日期、隐患部位、隐患问题、隐患整改要求等 9 类实体,共计 191 070 个实体。其次,采用基于模板关系抽取方法^[17],依据定义的实体关系模板,抽取隐患-时间、隐患-类型、隐患-部位、隐患-措施、隐患-期限、隐患-单位、单位-执行 7 类关系,共计 148 610 个关系。最后,运行构建知识图谱的代码,并通过链接“http://localhost:7474/browser/”将知识图谱存储到 neo4j 数据库中。利用 Cypher 查询语句,以实体为节点,以实体关系为边,得到水电工程施工隐患知识图谱^[18],图 2 中展示了 2 000 个节点。

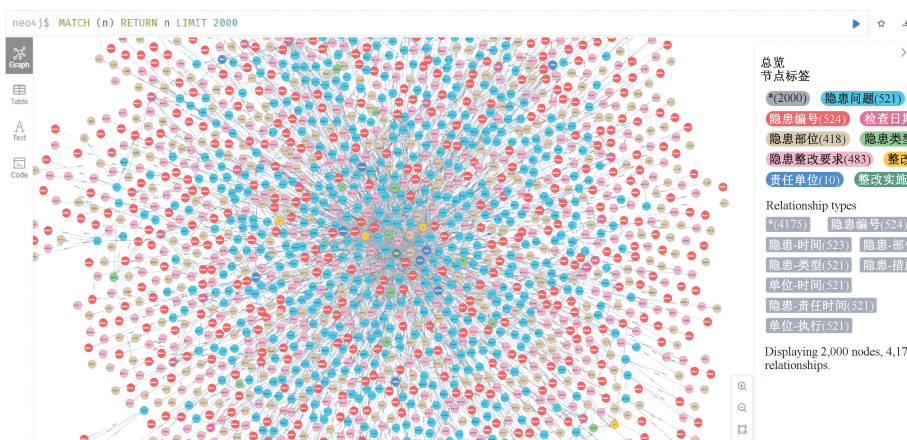


图 2 水电工程施工安全隐患知识图谱(部分)

Fig. 2 Knowledge graph of safety hazards in hydroelectric construction (partial)

1.3 隐患语义匹配模型构建

中文表达一词多义,增加了语义匹配难度,如风带被吹烂和风带破损表达意思相同但在句子向量化表示时差异较大,导致语义相似度较小。为解决这

一问题,采用隐患语义匹配的 Sentence-BERT 网络模型,将语义相近的 2 句话映射为相似的向量空间,提取文本深层语义信息为相似隐患间的语义匹配提供支持^[19]。

Sentence-BERT 模型包含输入层、BERT 模型层、池化层以及输出层,模型结构如图 3 所示。

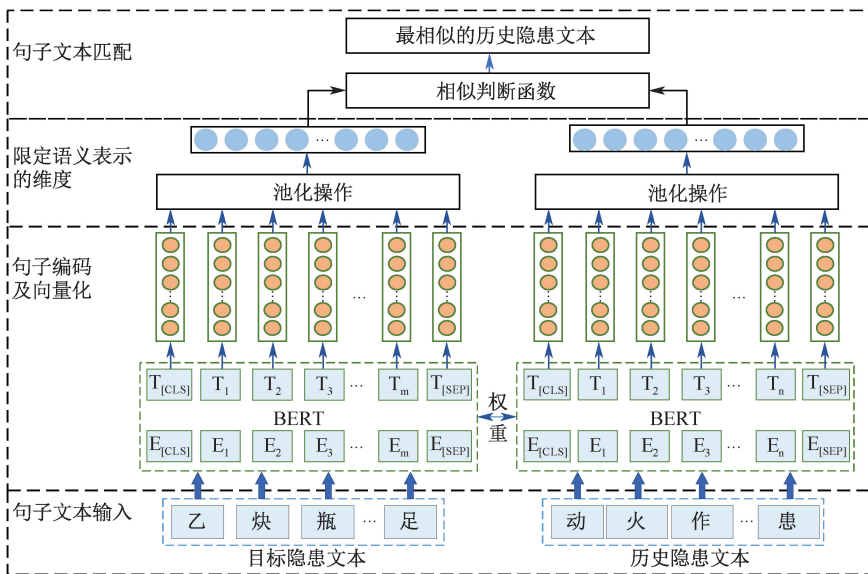


图 3 Sentence-BERT 模型

Fig. 3 Sentence-BERT model

在输入层,将安全隐患文本以及对应的历史相似隐患文本输入 BERT 模型层,将其转化为词嵌入向量。与此同时, BERT 模型层通过多个 Transformer 层,学习句子的上下文语义信息。Pooling 层对 BERT 输出进行平均池化操作,将整个句子的表示压缩为一个固定维度的向量。再经由输出层后计算目标隐患与历史相似隐患句子向量 x 和 y 的余弦相似度:

$$\text{Cosine_Similarity}(x_i, y_i) =$$

$$\frac{x_i \cdot y_i}{\|x_i\| \cdot \|y_i\|} = \frac{\sum_{j=1}^D x_{ij} \cdot y_{ij}}{\sqrt{\sum_{j=1}^D x_{ij}^2} \sqrt{\sum_{j=1}^D y_{ij}^2}} \quad (1)$$

式中: x_i 为第 i 条目标隐患对应的向量; x_{ij} 为目标隐患向量第 j 个位置上的值; y_i 为第 i 条历史隐患对应的向量; y_{ij} 为该历史隐患向量第 j 位置上的值; D 为向量维度。Cosine_Similarity(x, y) 值越大表示 2 个句子的语义相似度越高。

采用孪生网络的思想,训练 Sentence-BERT 模型,结构如图 4 所示。 X_1 和 X_2 作为孪生神经网络的 2 个输入,分别传入具有相同权重 W 的神经网络模型中进行编码映射,生成在新空间中的向量表示。由于通过设定阈值判断句子对的相似主观性较强,因此,采用余弦相似度计算损失函数 Loss,以优化模型在语义相似性任务上的性能,从而达到“最小化相似句子对的距离,最大化不相似句子对的距离”

的效果^[20]。

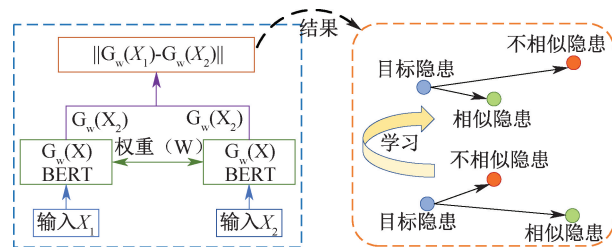


图 4 孪生网络结构

Fig. 4 Siamese network structure

$$\text{Loss} = \frac{\sum_{i=1}^m |l_i - \text{Cosine_Similarity}(x_i, y_i)|}{m} \quad (2)$$

式中: m 为训练集样本量; l_i 为第 i 个文本对标签对应值,正样本标签为 1,负样本标签为 0。

1.4 语义匹配模型性能评价指标

分别计算隐患文本对在曼哈顿距离、欧几里得距离、点积相似度以及余弦相似度 4 种语义相似度评估方法下的相似度得分。通过该得分与实际标签之间的 Pearson 系数(p)和 Spearman 系数(S)来衡量文本对语义相似度与标签之间的相关性,并将其作为模型性能的评估指标。

$$p = \frac{\sum_{i=1}^m (Y_i - \bar{Y})(l_i - \bar{l})}{\sqrt{\sum_{i=1}^m (Y_i - \bar{Y})^2} \sqrt{\sum_{i=1}^m (l_i - \bar{l})^2}} \quad (3)$$

$$S = 1 - \frac{6 \sum_{i=1}^m (\text{rank}(Y_i) - \text{rank}(L_i))^2}{m(m^2 - 1)} \quad (4)$$

式中: Y_i 为第 i 个文本对的相似度得分; \bar{Y} 为 m 个文本对相似度的均值; \bar{l} 为 m 个文本对标签的均值; $\text{rank}(a)$ 为向量 a 的秩。 p 与 S 取值范围均为 $[-1, 1]$, 当 p 和 S 值为 0 时表明 2 个所选样本无相关性, 值越趋于 1, 二者正相关性越强, 值越趋于 -1, 二者负相关性越强^[21]。

为进一步准确客观地检验水电工程施工安全隐患语义匹配模型的效果, 在测试阶段计算准确率 A 、精准率 P 、召回率 R 和 F_1 值(精确率和召回率的调和平均数), 评价模型对目标隐患和历史相似隐患的识别效果^[22]。

$$A = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (5)$$

$$P = \frac{TP}{TP + FP} \times 100\% \quad (6)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (7)$$

$$F1 = \frac{2 \times P \times R}{P + R} \times 100\% \quad (8)$$

式中: TP 为真阳性(正样本且识别正确); FN 为假阴性(正样本且识别错误); FP 为假阳性(负样本且识别错误); TN 为真阴性(负样本且识别正确)。

2 语义智能推荐结果与分析

在 Windows 操作系统下, 使用 PyTorch 2.1.0 深度学习框架, 采用 Python 3.9 编程语言和 PyCharm 编程平台搭建试验运行环境, 构建 Sentence-BERT 网络模型, 实现相似隐患之间的语义匹配, 利用微处理器进行计算, 显卡为 RTX3050, 8 G 显存, 内存 16 GB。为使模型达到全局最优, 参考设置学习率 0.000 5, 单次训练文本对数 32, 总迭代轮次 50^[14, 23]。基于以上超参数, 分析所提水电工程施工安全隐患语义匹配模型的效果。

2.1 模型训练结果分析

评价指标和模型迭代轮次之间的关系如图 5 所示。整体上, 4 种方法的 p 和 S 系数变化趋势基本保持一致, 在训练的前 5 轮增长幅度较为明显, 随着训练轮次的增加呈现出波动趋势, 在训练 5~35 轮时数值波动较小, 35 轮以后基本趋于稳定, 模型达到拟合状态。

从 p 和 S 的具体数值来看, 当达到拟合状态时, 曼哈顿距离、欧氏距离、点积相似度和余弦相似度的 p 值均稳定在 0.92 左右, S 值均稳定在 0.85 左右。表明经过训练的模型在不同度量方法下均取得较高语义识别能力, 语义相似度与实际标签之间具有强烈的线性关系, 模型性能良好。

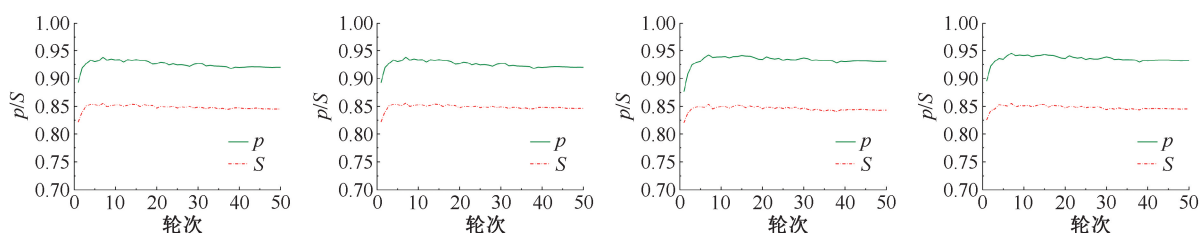


图 5 评价指标和迭代轮次关系

Fig. 5 Relationship between evaluation indicators and iteration rounds

2.2 语义匹配模型测试与评价

为评估 Sentence-BERT 模型对目标隐患和历史相似隐患的语义匹配效果, 在同一试验环境下, 采用常见的 BERT、基于词向量的深度语义匹配模型(Word2vec-Deep Semantic Similarity Model, Word2vec-DSSM)和基于 BERT 的深度语义匹配模型(BERT-DSSM)等文本语义匹配模型^[24]进行对比分析。

将数据集分别导入构建好的 BERT、Word2vec-DSSM 和 BERT-DSSM 模型中进行训练。处理批次

大小均为 100, 学习率均为 0.000 5, 迭代轮数均设置为 200 轮。为防止过拟合, 在每个线性层的输出后, 应用 Dropout 操作, 并将概率 P 设为 0.2, 对训练好的模型进行保存。采用 A 、 P 、 R 和 F_1 值评估测试集数据的模型性能, 结果见表 2。

由表 2 可知: Sentence-BERT 模型在进行水电工程目标隐患与历史相似隐患识别方面效果优于 BERT、Word2vec-DSSM 和 BERT-DSSM 模型, 其准确率等评价指标均在 95% 以上, 能够有效捕捉安全隐患语义信息。

表 2 不同模型测试性能对比

Table 2 Comparison of test performance of

different models				%
模型	A	P	R	F ₁ 值
Word2vec-DSSM	86.61	87.72	86.62	86.64
BERT-DSSM	84.55	83.34	84.49	84.12
BERT	90.43	89.55	91.57	91.24
Sentence-BERT	96.48	95.80	97.23	96.51

2.3 相似历史隐患治理信息检索

通过调用训练好的模型,推荐与目标隐患相似度最高的历史隐患。若推荐的历史隐患与目标隐患相对应,则通过 Cypher 语句进行隐患治理措施实体的查询;若推荐的相似度最高的历史隐患与目标隐患相似度较低,且二者相关性较小(即推荐的历史隐患相关治理措施并不能对目标隐患治理提供参考),则表明该目标隐患在以前并未发生,针对此情况,需要将目标隐患存储到历史隐患中以更新隐患图谱,从而提升图谱的泛化能力。

2.3.1 历史隐患匹配

在测试集样本中随机抽取 150 条目标隐患文本,将整理的 3 124 条历史隐患作为召回文本集,进行历史相似隐患推荐测试。最终发现,抽取的 150 条目标隐患文本中有 138 条都准确地推荐到了与其对应的相似历史隐患,推荐精度达到 92%,部分目标隐患的推荐结果见表 3。

表 3 历史隐患推荐结果展示

Table 3 Historical hazard recommendation result display

目标隐患记录	历史隐患记录	相似度值
左岸 4 号帷幕灌浆洞洞口 K0+10 位置,氧气瓶和乙炔瓶混放,未到达相关规定的安全距离,存在安全隐患	氧气瓶与乙炔气瓶未按安全距离摆放,存在爆炸危险	0.976 5
钢筋绑扎区域无上下通道,施工作业人员沿斜放钢筋上下	7 号机背坡备仓过程中,该段无通道,人员在已安装的钢筋上行走	0.963 2
施工现场防护栏底部挡脚板缺失,存在物体打击风险	654 马道防护挡脚板缺失,且文明施工较差	0.982 7
电焊机插头使用钢筋代替,存在触电风险	抗力体马道电焊机电焊线插头用钢筋代替,存在安全隐患	0.968 8
炸药临时存在点未警戒,无专人看守	0+230 处炸药临时存放无人看守,无警示灯,警示牌	0.944 6

续表 3

目标隐患记录	历史隐患记录	相似度值
砼泵管使用 K1800 塔机长短泵管混吊入仓易发生坠物伤害风险	下游围堰防渗墙浇筑溜槽稳定性不够,存在倾倒危险	0.653 1

2.3.2 历史隐患治理措施检索

为直观展示相似历史安全隐患治理措施的检索结果,运用 cypher 查询语句依次将表 3 中的相似历史安全隐患进行治理措施检索,具体结果如图 6 所示。

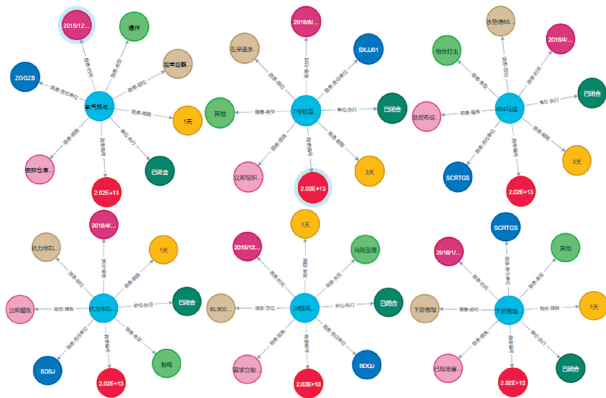


图 6 信息检索结果展示

Fig. 6 Display of information retrieval results

为清晰展示检索效果,以“左岸 4 号帷幕灌浆洞洞口 K0+10 位置,氧气瓶和乙炔瓶混放,未到达相关规定的安全距离,存在安全隐患”为例,由表 3 可知:通过语义推荐得到历史隐患“氧气瓶与乙炔气瓶未按安全距离摆放,存在爆炸危险”与之对应,二者语义相似度值为 0.976 5。利用 Cypher 语句,在 neo4j 数据库中输入“MATCH((n:‘隐患问题’{name:“氧气瓶与乙炔气瓶未按安全距离摆放,存在爆炸危险。”}))RETURN n LIMIT 25”,检索结果如图 7 所示。

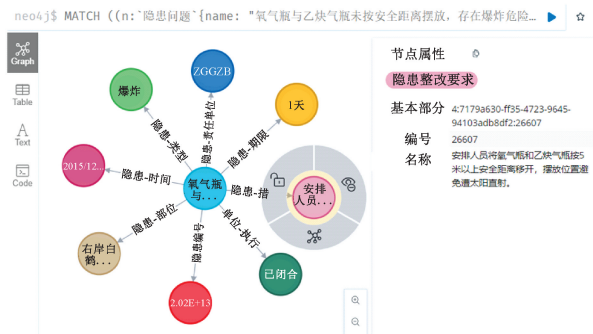


图 7 具体安全隐患治理措施检索结果展示

Fig. 7 Specific safety hazard rectification measures search results show

由图7可知:该隐患的解决需要安全人员将氧气瓶和乙炔瓶按5 m以上安全距离分开存放,并且需要注意氧气瓶和乙炔瓶的摆放位置避免阳光照射,此隐患治理须在一天内整改落实到位。由此可知:水电工程施工安全隐患智能推荐方法能够有效辅助目标隐患治理措施的制定。

3 结论

1) 针对水电工程施工安全隐患文本特征,定义安全隐患实体、实体属性以及实体间的相互关系,运用实体和关系抽取技术,构建施工安全隐患知识图谱,将安全隐患与其对应的治理措施进行有效关联,

清晰展示隐患各实体间的逻辑关系。

2) 搭建水电工程施工隐患语义匹配的 Sentence-BERT 网络模型,提取目标隐患与历史相似隐患的语义信息,计算2个文本之间的余弦相似度确定其语义匹配程度。通过测试模型,其准确率、精确率、召回率以及 F_1 值等评估指标均高于95%,明显优于 BERT、Word2vec-DSSM 以及 BERT-DSSM 模型,智能推荐效果优异。

3) 在历史相似隐患智能推荐的基础上,利用知识图谱的 Cypher 语句,成功实现历史相似隐患治理措施相关实体查询。这一实践验证文中所提方法在隐患治理领域的可行性,有效促进隐患的高效处理。

参考文献

- [1] 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.
- [2] 王仁超, 张毅伟, 毛三军. 水电工程施工安全隐患文本智能分类与知识挖掘[J]. 水力发电学报, 2022, 41(11): 96-106.
WANG Renchao, ZHANG Yiwei, MAO Sanjun. Intelligent text classification and knowledge mining of hidden safety hazards in hydropower engineering construction [J]. Journal of Hydroelectric Engineering, 2022, 41(11): 96-106.
- [3] 靳嵩, 朱艳, 吴可嘉, 等. 基于 BERT 的海上船舶安全隐患分类[J]. 船舶工程, 2023, 45(增1): 381-384.
JIN Song, ZHU Yan, WU Kejia, et al. Classification of ship safety hazards of marine vessel based on BERT [J]. Ship Engineering, 2023, 45(S1): 381-384.
- [4] 陈述, 刁俊博, 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.
- [5] 陈述, 刘文濯, 邵波, 等. 水电工程施工安全隐患时空分布特征研究[J]. 中国安全科学学报, 2022, 32(11): 90-96.
CHEN Shu, LIU Wenzhuo, SHAO Bo, et al. Research on spatial and temporal distribution characteristics of safety nazards in hydropower projects [J]. China Safety Science Journal, 2022, 32(11): 90-96.
- [6] 肖琪耀, 费宝山, 徐以诺, 等. 基于深度学习模型的煤矿安全隐患数据主题挖掘[J]. 中国安全生产科学技术, 2024, 20(4): 49-55.
XIAO Qiyao, JIA Baoshan, XU Yinuo, et al. Topics mining on potential safety hazard data of coal mine based on deep learning models [J]. Journal of Safety Science and Technology, 2024, 20(4): 49-55.
- [7] 宁桂峰, 高龙, 刘利平. 基于遗传算法的采煤工作面隐患数据关联规则挖掘[J]. 采矿与岩层控制工程学报, 2024, 6(2): 136-145.
NING Guifeng, GAO Long, LIU Liping. Mining hidden danger data association rules of coal mining face based on genetic algorithm [J]. Journal of Mining and Strata Control Engineering, 2024, 6(2): 136-145.
- [8] 陈述, 鲁世立, 王建平, 等. 缆机安全运行本体知识表示及复用方法[J]. 中国安全科学学报, 2024, 34(3): 93-100.
CHEN Shu, LU Shili, WANG Jianping, et al. Representation and reuse of ontology knowledge for safe operation of cable crane [J]. China Safety Science Journal, 2024, 34(3): 93-100.
- [9] 臧洁, 周万林, 王妍. 融合多头注意力机制和孪生网络的语义匹配方法[J]. 计算机科学, 2023, 50(12): 294-301.
ZANG Jie, ZHOU Wanlin, WANG Yan. Semantic Matching method integrating multi-head attention mechanism and siamese network [J]. Computer Science, 2023, 50(12): 294-301.
- [10] 阮光册, 钟静涵, 张祎笛. 基于深度学习的术语识别研究综述[J]. 数据分析与知识发现, 2024, 8(4): 64-75.
RUAN Guangce, ZHONG Jinghan, ZHANG Yidi. Review of term recognition studies based on deep learning [J]. Data Analysis and Knowledge Discovery, 2024, 8(4): 64-75.

- [11] 管立本, 李实. 融合多粒度语义信息和知识图谱的中文医疗问答匹配模型[J]. 计算机工程与应用, 2024, 60(14):152-161.
GUAN Liben, LI Shi. A Chinese medical Q&A matching model based on multi-granularity semantic information and knowledge graph [J]. Computer Engineering and Applications, 2024,60(14):152-161.
- [12] 许童羽, 赵冬雪, 周云成, 等. 基于 word2vec 和 Attention-Seq2Seq 的水稻病虫害智能问答方法研究[J]. 沈阳农业大学学报, 2019,50(3):378-384.
XU Tongyu, ZHAO Dongxue, ZHOU Yuncheng, et al. Research on method of intelligent Q&A for rice pests and diseases based on word2vec and Attention-Seq2Seq [J]. Journal of Shenyang Agricultural University, 2019,50(3):378-384.
- [13] 李琳, 段围, 周栋, 等. 基于深度语义匹配的法律条文推荐方法[J]. 软件学报, 2022,33(7):2 618-2 632.
LI Lin, DUAN Wei, ZHOU Dong, et al. Law article recommendation approach based on deep semantic matching [J]. Journal of Software, 2022,33(7):2 618-2 632.
- [14] 张晓健, 张栋梁, 李明超, 等. 面向质量检测的混凝土坝施工规范智能检索[J]. 水力发电学报, 2023,42(4):114-125.
ZHANG Xiaojian, ZHANG Dongliang, LI Mingchao, et al. Intelligent retrieval of concrete dam construction specification for quality inspection[J]. Journal of Hydroelectric Engineering, 2023,42(4):114-125.
- [15] 罗香玉, 杜浩, 华颖, 等. 一种煤矿顶板灾害防治知识图谱构建方法[J]. 工矿自动化, 2024,50(6):54-60.
LUO Xiangyu, DU Hao, HUA Ying, et al. A method for constructing a knowledge graph of coal mine roof disaster prevention and control[J]. Journal of Mine Automation, 2024,50(6):54-60.
- [16] 张吉祥, 张祥森, 武长旭, 等. 知识图谱构建技术综述[J]. 计算机工程, 2022,48(3):23-37.
ZHANG Jixiang, ZHANG Xiangsen, WU Changxu, et al. Survey of knowledge graph construction techniques [J]. Computer Engineering, 2022,48(3):23-37.
- [17] 王莉, 王建平, 许娜, 等. 基于知识图谱的地铁工程事故知识建模与分析[J]. 土木工程与管理学报, 2019, 36(5):109-114,122.
WANG Li, WANG Jianping, XU Na, et al. Knowledge graph-based metro engineering accidents knowledge modeling and analysis [J]. Journal of Civil Engineering and Management, 2019,36(5):109-114,122.
- [18] 郭天颖, 茆晓阳, 段齐骏, 等. 基于知识图谱的地震救援装备智能管理方法[J]. 中国安全科学学报, 2024, 34(7):239-245.
GUO Tianying, MAO Xiaoyang, DUAN Qijun, et al. Intelligent management and scheduling approach for earthquake rescue equipment based on knowledge graph[J]. China Safety Science Journal, 2024,34(7):239-245.
- [19] HU Caiping, SNU Xuekui, DAI Hua, et al. Research on log anomaly detection based on sentence-BERT [J]. Electronics, 2023,12(17):DOI: 10.3390/ELECTRONICS12173580.
- [20] 叶文豪, 王东波, 沈思, 等. 基于孪生网络的基金与受资助论文相关性判别模型构建研究[J]. 情报学报, 2020, 39(6):609-618.
YE Wenhao, WANG Dongbo, SHEN Si, et al. Research on constructing a model of correlation discrimination between funds and funded papers based on siamese network [J]. Journal of the China Society for Scientific and Technical Information, 2020,39(6):609-618.
- [21] 苏锦钿, 洪晓斌, 余珊珊. 基于多模型集成的语义文本相似性判断[J]. 华南理工大学学报:自然科学版, 2022, 50(4):1-9.
SU Jindian, HONG Xiaobin, YU Shanshan. Semantic textual similarity justification based on multi-model ensemble [J]. Journal of South China University of Technology:Natural Science Edition, 2022,50(4):1-9.
- [22] CAO Kunyu, CHEN Shu, ZHANG Xinkai, et al. Identification of causative factors for fatal accidents in the electric power industry using text categorization and catastrophe association analysis techniques [J]. Alexandria Engineering Journal, 2024,102:290-308.
- [23] 夏占杰, 张贝克, 高东. 基于数据增强的 HSE 检查纪要命名实体识别[J]. 中国安全科学学报, 2022,32(12):53-62.
XIA Zhanjie, ZHANG Beike, GAO Dong. Named entity recognition of HSE inspection minutes based on data enhancement[J]. China Safety Science Journal, 2022,32(12):53-62.
- [24] 钱杨舸, 秦小林, 张思齐, 等. 基于深度学习的文本语义匹配综述[J]. 软件导刊, 2022,21(12):252-261.
QIAN Yangge, QIN Xiaolin, ZHANG Siqi, et al. Survey of texts semantic matching based on deep learning [J]. Software Guide, 2022,21(12):252-261.
- [25] JADHAV S S, THEPADE S. Fake news identification and classification using dssm and improved recurrent neural network classifier [J]. Applied Artificial Intelligence, 2019,33(12):1 058-1 068.

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