

杨阳蕊, 朱亚萍, 刘雪梅, 等. 知识驱动的南水北调工程巡检信息推荐方法[J]. 南水北调与水利科技(中英文), 2024, 22(2): 368-377.
YANG Y R, ZHU Y P, LIU X M, et al. Knowledge-driven recommended method for inspection information of South-to-North Water Transfers Project[J]. South-to-North Water Transfers and Water Science & Technology, 2024, 22(2): 368-377. (in Chinese)

知识驱动的南水北调工程巡检信息推荐方法

杨阳蕊¹, 朱亚萍¹, 刘雪梅^{1,2}, 陈思思¹, 李慧敏³

(1. 华北水利水电大学信息工程学院, 郑州 450000; 2. 黄河流域水资源高效利用省部共建协同创新中心, 郑州 450000;
3. 华北水利水电大学水利学院, 郑州 450000)

摘要: 基于南水北调工程巡检专报结合专家经验构建巡检知识图谱概念模型, 在此基础上利用实体关系联合抽取框架进行巡检知识抽取, 并以 Neo4j 图数据库为载体进行巡检知识图谱可视化。基于巡检知识图谱进行南水北调工程巡检信息推荐, 利用 BERT(Bi-directional encoder representation from transformers)预训练模型设计 BERT 孪生网络, 通过知识检索及字符串相似度计算等技术, 实现关联工程风险信息推荐, 以辅助巡检人员进行工程风险等级诊断。通过实验评估了知识抽取及字符串相似度计算模型的准确性, 知识抽取 F_1 值达到 88.42%, 字符串相似度计算 F_1 值为 86.00%。该方法可提高南水北调工程风险管理能力、推动工程运维的数字化发展。

关键词: 知识图谱; 南水北调工程; 工程巡检信息推荐; 工程风险等级诊断

中图分类号: TP391.1 **文献标志码:** A **DOI:** 10.13476/j.cnki.nsbdkq.2024.0038

工程巡检作为工程运维中的关键性工作, 可以有效保障工程的安全稳定运行。传统工程巡检方法主要依赖人工经验, 存在着两方面的局限: 一是巡检人员专业水平参差不齐, 难以形成统一问题记录标准; 二是巡检专报信息冗余, 有效信息的可访问性差。随着河流演变带来的工程规模的不断扩大及工程复杂性不断增加, 传统的工程巡检方法已经无法满足高效智能的管理需求^[1]。因此, 利用知识图谱及深度学习技术进行南水北调工程巡检信息智能推荐, 以推动南水北调工程运维数字化、智能化进展亟待研究。

知识图谱作为新一代信息技术, 是知识组织和管理的有力工具^[2-4]。目前已有一些研究^[5-6]提出了基于知识图谱和深度学习的水利工程智能运维方法。例如: Tang 等^[7]利用深度学习和贝叶斯算法进行设备故障的智能诊断; 刘雪梅等^[8]通过构建水利工程应急方案知识图谱, 实现了风险事件应急方案的智能生成与推送。这些研究^[9-11]主要集中在水务和水利工程信息的组织和推理方法, 而对于工程巡检方法的数字化及智能化研究仍有待进一步探索。

为了缓解传统工程巡检方法的局限性, 提升工程巡检效率, 将知识图谱与深度学习技术相结合, 提出知识驱动的南水北调工程巡检信息推荐方法。具体而言, 基于工程巡检文本构建巡检知识图谱, 并设计基于巡检知识图谱的工程风险信息推荐方法^[12]。在巡检知识图谱构建过程中, 结合专家经验定义巡检知识图谱概念模型^[13], 并在此基础上利用实体关系联合抽取框架从非结构化的工程巡检文本中抽取出结构化三元组知识, 以 Neo4j 图数据库^[14]为载体进行知识可视化。巡检知识图谱以实体-关系-实体的三元组形式清晰呈现出工程站点、部位、风险事件及处置措施等巡检信息, 同时支持知识可视化与知识检索^[15], 这缓解了巡检专报中有效信息可访问性差的局限。在工程风险信息推荐方法中, 利用 BERT (Bi-directional encoder representation from transformers)预训练模型^[16]和孪生网络框架^[17]设计了 BERT 孪生网络, 通过计算部位实体的字符串相似度, 将巡检知识图谱中与当前部位匹配的实体及实体关联的工程风险信息推荐给巡检人员^[18-19], 辅助巡检人员进行工程风险等级

收稿日期: 2023-08-10 修回日期: 2024-03-11 网络出版时间: 2024-03-28

网络出版地址: <https://link.cnki.net/urlid/13.1430.TV.20240326.1603.002>

基金项目: 国家自然科学基金项目(72271091); 河南省科学院科技开放合作项目(220901008); 华北水利水电大学硕士创新能力提升工程资助项目(NCWUYC-2023091)

作者简介: 杨阳蕊(1982—), 女(回族), 河南南阳人, 副教授, 博士, 主要从事自然语言处理与智慧水利研究。E-mail: yangyangrui@ncwu.edu.cn

诊断。将知识图谱和深度学习技术引入到工程巡检智能化应用中,实现了巡检知识的深度关联和有效利用,对提高南水北调工程运维效率、强化工程风险管理能力具有重要作用。

1 研究方法

以知识图谱赋能南水北调工程智能运维为出发点,提出利用知识图谱进行巡检文本的组织与管理。

基于南水北调工程运行管理质量问题巡检专报定义明确的关系关联各巡检实体,并在此基础上构建巡检知识图谱以结构化形式清晰呈现工程巡检信息。同时,利用BERT模型及孪生网络框架设计了BERT孪生网络框架,提出基于巡检知识图谱的工程风险信息推荐方法,以辅助巡检人员进行工程风险等级诊断。方法总体框架见图1。

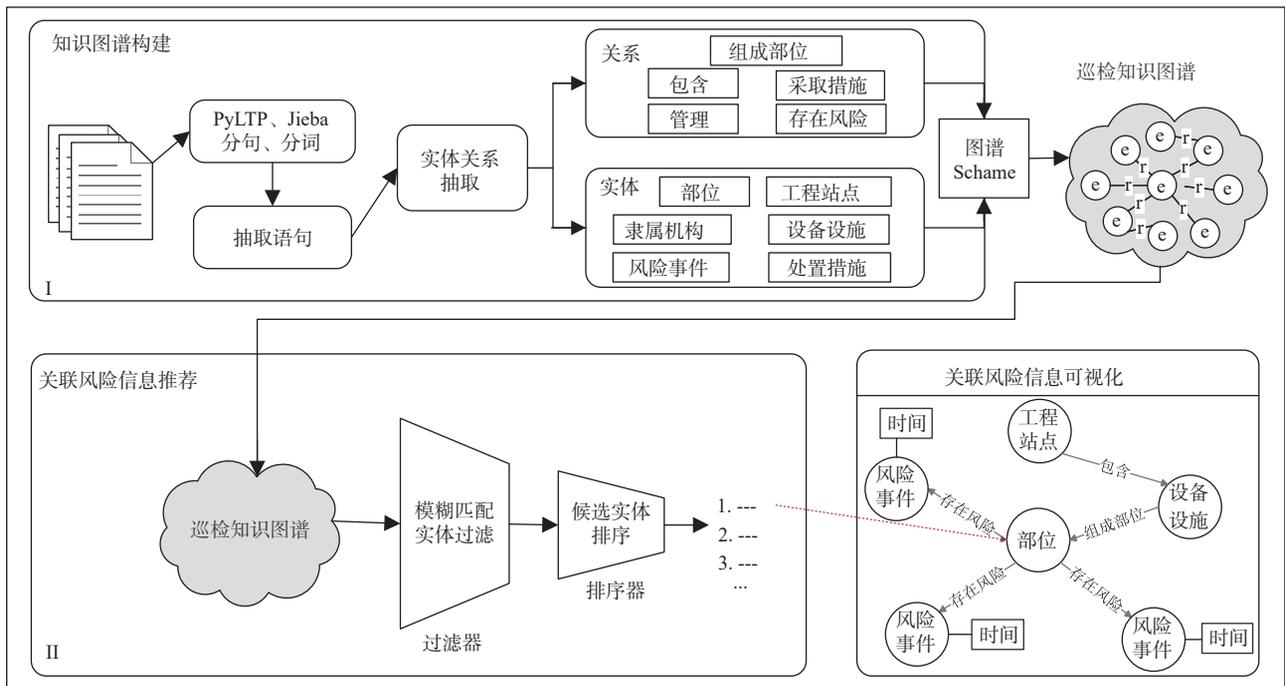


图1 方法总体框架

Fig. 1 Overall framework

1.1 巡检知识图谱概念模型构建

工程巡检知识图谱为专业知识图谱,采用自顶向下的构建模式^[20]。概念模型是工程巡检知识图谱的基础和核心^[21-22]。南水北调工程运行管理质量问题巡检专报是语义描述清晰的高质量文本数据,其通常按要求顺序记录问题要点,包括管理机构、工程站点、设备设施或建筑物等,本文将此称为巡检实体。同时巡检实体间关系也隐含在文本中,如隶属机构与工程站点之间的“管理”关系,部位与风险事件之间的“存在风险”关系,将巡检实体与其间关系共称为工程巡检知识。根据工程巡检任务要求,分析南水北调工程运行管理质量问题巡检专报文本及隐患排查记录数据类型,定义出“隶属机构、工程站点、设备设施/建筑物、部位、风险事件,处置措施”共6种基本实体类型,以及“发生时间”等属性信息。在这里,设备设施和建筑物属于同等级实体,因为问题有可能存在于设备设施,也有可能存在于

建筑物。同时定义出各实体类型之间的关系,关系类型为“管理、包含、组成部位、存在风险、采取”,实体类型和关系类型共同构成了巡检知识图谱的基础概念模型。概念模型及对应实体关系实例见图2。

1.2 巡检知识图谱数据层构建

1.2.1 初始数据源筛选

考虑两种类型的文本数据,分别是南水北调工程运行管理质量问题巡检专报(以下称专报)与南水北调工程隐患排查记录(以下称记录)。“专报”文本以自然语言结合视觉图像的形式提供了工程问题的非结构化声明,包含工程养护缺陷与工程运行管理违规行为等问题类型,并解释了问题表现与问题详细等级。“记录”以自然语言形式提供了隐患排查的半结构化知识,包含隐患排查时间与隐患详情,同时,详细描述了不同隐患的应对措施。通过人工对以上文本数据进行初筛,过滤掉无关和噪声数据获得规范目标文本。

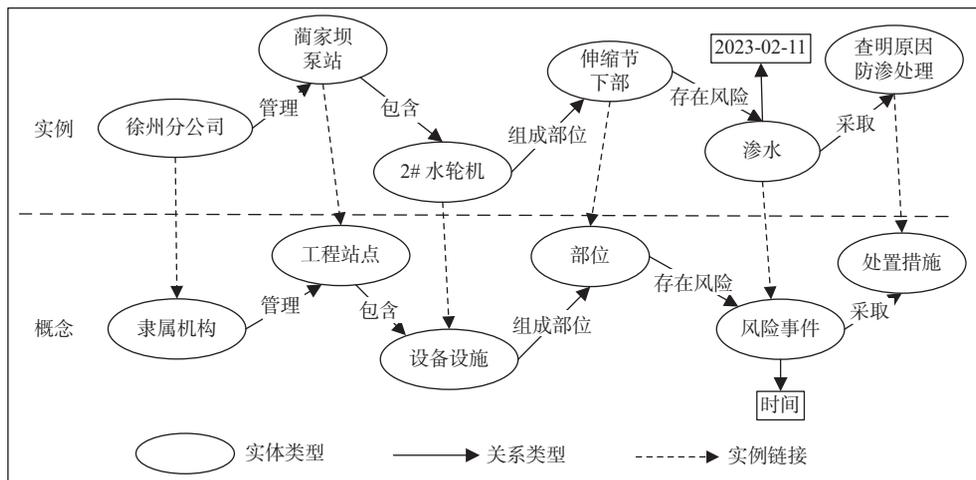


图 2 图谱概念模型及对应实例
Fig. 2 Conceptual model of knowledge graph and example

1.2.2 知识抽取方法

为进一步完成知识抽取,初始自然语言文本必须经分句、分词等处理。利用 PyLTP 对文本完成分句,最终处理得到 2 092 条描述南水北调工程问题的文本数据。分词过程中,通用 Jieba 分词器可能会将一个完整实体错分,例如“午河渡槽进口节制闸”被分为“午河”、“渡槽”、“进口”、“节制闸”,这会破坏实体完整性,进而对后续知识抽取产生负面影响。因此,加载自定义词典扩展 Jieba 词典库,使扩展后的 Jieba 分词器在分词过程中保留意义完整的实体。

水利领域现有知识抽取工作大多基于传统深度学习学习方法以任务流形式进行,使用传统深度学习进行领域知识抽取的一个关键挑战是标注样本稀缺问题,手工标注大规模训练样本造成极大的人力开销^[23]。任务流形式的知识抽取方式,即先进行实体识别再完成关系分类,实体识别结果的正误直接决定着关系分类结果的正误^[24],这一固有特性会造成实体识别和关系分类两个任务之间的误差累积,给最终三元组知识的质量控制带来负面影响。针对以上两点,本文基于迁移学习方法,使用大型预训练语言模型(T5)^[25]对南水北调工程巡检文本数据中的实体和关系进行联合提取。与传统深度学习模型^[26]不同,T5学习了海量语言知识,可以作为现实世界实体和关系的神经知识库^[27-28]。同时,T5可提供更优的模型初始化以及学习能力^[29]。使用一小组任务相关的训练数据对 T5 进行微调,通用语义信息结合南水北调工程巡检领域语义信息,也即是预训练(通用信息)+微调(领域信息)的最新范式,

可以使 T5 具有识别出当前语句中存在的巡检实体和关系的能力。受近期通用信息抽取以及提示调整^[30]等工作的启发,采取基于 T5 的实体关系联合抽取框架-WRERJE^[31],整体框架见图 3。WRERJE 首先将输入语句 X 处理为固定提示序列 P ,该过程可表示为

$$[p_1, p_2, \dots, p_n] = PG([x_1, x_2, \dots, x_n]) \quad (1)$$

式中:PG 为基于文本分类模型实现的提示生成器; $X=[x_1, x_2, \dots, x_n]$ 为输入语句; $[p_1, p_2, \dots, p_n]$ 为所生成的提示序列 P 。由 T5 从该提示序列中同时提取出巡检实体和关系,并处理为结构化三元组知识(structured triplet knowledge, S_k)序列^[32],该生成过程可表示为

$$[y_1, y_2, \dots, y_y] = JE([p_1, p_2, \dots, p_n]) \quad (2)$$

式中:JE 为基于 Transformer 的大型预训练语言模型 T5, $[y_1, y_2, \dots, y_y]$ 为包含巡检实体和关系的 S_k 序列。 S_k 序列具体形式见图 3(III)。

1.2.3 知识可视化实现

基于 1.1 节中构建的图谱概念模型,以南水北调工程运行管理质量问题巡检专报与南水北调工程隐患排查记录等数据为基础构建工程巡检知识图谱。同时,利用水利领域期刊文献、水利百科等公开数据对图谱内实体及关系数据进行补充。抽取的实体类型包括隶属机构、工程站点、设备设施、部位、风险事件及处置措施等,构建的关系有管理、包含、组成部位、存在风险及采取等。知识抽取结果以三元组形式表达,例如(下部油管路,存在风险,2处漏油)、(两侧土体,存在风险,局部沉降)等,在知识图谱中以节点表达实体,以有向边代表关系,

通过有向边连接两个节点从而表达出三元组陈述,最终以图数据库 Neo4j^[5]为载体完成三元组中实体

和关系到图谱中节点和边元素的映射,实现知识可视化。

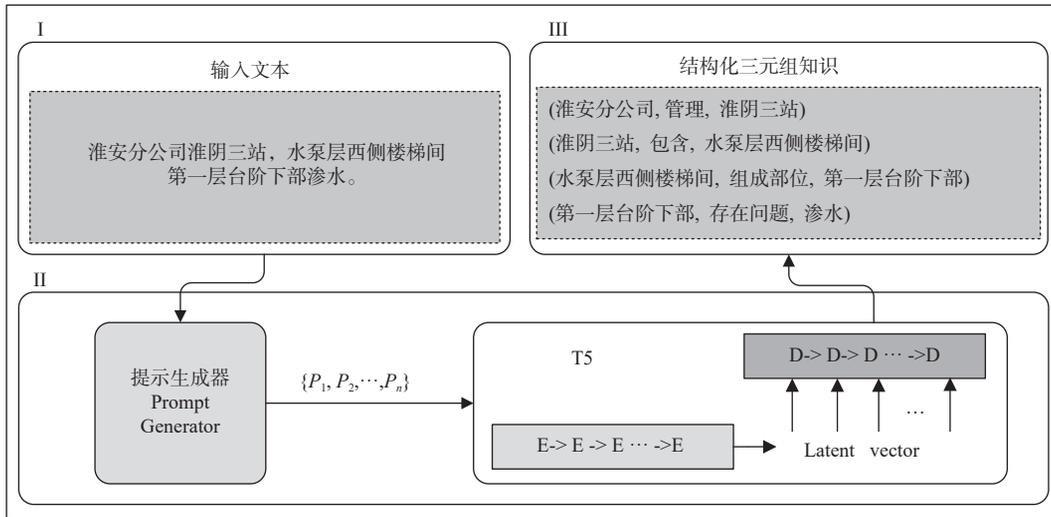


图3 知识抽取框架

Fig. 3 Knowledge extraction framework

1.3 基于知识图谱的工程风险信息推荐

历史风险信息对辅助当前风险事件的诊断具有重要意义,例如巡检发现某一部位出现裂缝;若是首次出现的较小裂缝,则可能风险等级较低;若当前部位历史已发生过较严重程度的裂缝,则当前裂缝风险等级可能较高,需引起工作人员高度重视。本文将与当前部位有关的工程风险信息推荐给巡检人员,进而辅助巡检人员进行工程风险等级诊断。传统基于协同过滤的推荐方法存在冷启动等关键问题^[33]影响推荐效果,针对此问题,构建了巡检知识图谱,设计了基于知识图谱的“两阶段”工程风险信息推荐方法^[34],将当前巡检部位关联的风险信息可视化^[35],辅助巡检人员定位当前风险严重程度。

推荐系统被设计为一个“漏斗”,其中包含两部分核心模块,分别是过滤器和排序器,见图1(II)。过滤器的任务为生成候选实体列表,排序器旨在对候选实体列表排序,从而获得最佳推荐结果。首先获取当前部位实体作为输入,通过模糊匹配^[36]方式将当前部位实体映射到底层巡检知识图谱中的目标实体,并生成候选实体列表。在排序模型中,设计了一个以BERT预训练模型作为骨干模型的孪生网络框架,用来计算当前输入实体与候选列表中实体的相似度,并根据相似度值进行排序,将相似度最高的候选实体以及与该实体关联的所有风险信息可视化呈现,从而辅助巡检人员进行风险等级诊断。

1.3.1 候选实体生成

过滤器旨在生成与当前部位相匹配的候选实体列表,其基本思想为字符串相似度映射。具体而言,根据巡检任务中涉及的主要实体类型设计了一个关键字匹配模板^[37],其中关键字包括工程站点、设备设施、部位、风险事件。由于“部位”实体对工程风险信息推荐任务的重要性,在生成候选实体列表时将“部位”作为匹配的主要实体类型,基于“部位”类型构建查询语句,在巡检知识图谱中查询出“部位”类型的所有实体。当巡检人员输入当前出现问题的部位时,将当前部位视为字符串 S ,过滤器任务是从巡检知识图谱查询到的所有部位实体中,生成与字符串 S 匹配的一个候选实体集合 $E=(e_1, e_2, \dots, e_n)$ 。由于巡检人员对水利领域专业知识了解程度不同,会出现对相同实体的不同描述,导致巡检记录文本存在描述不统一。因此,字符串 S 可能并非 e_n 的严格子集,而是与 e_n 相似的字符串。例如当前字符串 S 为“伸缩节下部”,而巡检知识图谱中存储的实体为“伸缩节”、“伸缩节下沿”等。此时,字符串 S 不能严格匹配巡检知识图谱中实体。为了缓解这个局限,使用模糊匹配的方法,将字符串 S 与查询到的实体进行模糊匹配,得到一组候选实体,即为候选实体列表 E 。

1.3.2 候选实体排序

完成候选实体生成任务之后,从巡检知识图谱中得到与当前部位相匹配的一个候选实体集合 E ,这个集合通常会包含多个实体。排序器的任务是

对候选实体进行排序,以确定候选实体中最终可以与当前 S 正确匹配的实体,从而呈现最佳推荐。具

体而言,利用基于 BERT 预训练模型的双生网络用于计算 S 与候选实体 e 的相似度,总体框架见图 4。

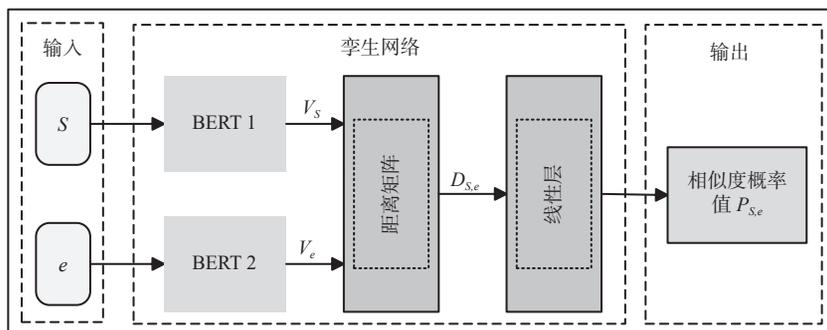


图 4 Bert 孪生网络框架

Fig. 4 Bert twin networking framework

输入当前部位实体 S 和集合 E 中的一个候选词 e , 使用骨干模型 BERT 提取 S 和 e 的特征, 将 BERT 中由 CLS 标记的隐藏向量作为 S 和 e 的向量表示 V_s 与 V_e , CLS 标记的隐藏向量通常包含了更丰富的语义信息, 能够更好地表示字符串特征。接着使用余弦相似度作为相似度比较方法, 由距离矩阵计算 V_s 与 V_e 之间的距离, 公式为

$$D_{S,e} = \cos_similarity(V_s, V_e) = (V_s * V_e) / (||V_s|| * ||V_e||) \quad (3)$$

式中: $D_{S,e}$ 为字符串 S 与 e 之间的距离; V_s 与 V_e 为 S 和 e 的特征向量。将计算所得距离 $D_{S,e}$ 通过线性层映射得到一个概率值 $P_{S,e}$, $P_{S,e} = (D_{S,e} + 1) / 2$, 概率越大表示两个字符串越相似。对集合 E 中所有候选实体的相似度概率进行排序, 相似概率最大的候选实体 e 即为与当前 S 最匹配的实体, 将与该实体关联的所有风险信息可视化返回给巡检人员, 从而辅助其进行风险等级诊断。

2 实验结果分析

2.1 知识抽取结果分析

如 1.2.2 节所述, 知识抽取框架 WRERJE 主要从初始语料中抽取实体关系三元组。通过人工对初始数据进行标注, 按照语料分割常用方法将初始文本语料按照 8 : 1 : 1 的比例划分为训练集、验证集与测试集, 对每条语句中各类关系三元组进行标注测试, 采用 P 、 R 和 F_1 值作为知识抽取框架的性能评估指标, 各指标计算公式为

$$P = \frac{T_p}{T_p + F_p} \times 100\% \quad (4)$$

$$R = \frac{T_p}{T_p + F_N} \times 100\% \quad (5)$$

$$F_1 = \frac{2PR}{P+R} \times 100\% \quad (6)$$

式(4)~(6)中: T_p 为预测为正例的正样本; F_p 为预测为正例的负样本; F_N 为预测为负例的正样本; P 为精确率(预测为正例的正样本与预测为正例的正负样本总数之比); R 为召回率(预测为正例的正样本与总的正样本之比); F_1 为 P 和 R 的调和平均值。实验结果显示该知识抽取框架在南水北调工程巡检实体关系三元组抽取任务中, P 、 R 和 F_1 值分别达到 87.80%、89.00% 和 88.42%, 见表 1。知识抽取结果具有较高准确性, 认为语料一致性可信赖^[38]。

表 1 各类型三元组抽取结果

Tab. 1 Extraction results of each type of triplet %

三元组类型	P	R	F_1
隶属机构-工程站点	84.51	84.63	84.57
工程站点-设备设施	87.06	91.16	89.05
设备设施-部位	91.84	92.89	92.36
部位-风险事件	83.50	87.54	85.47
风险事件-时间	91.17	91.95	91.56
风险事件-处置措施	89.12	85.86	87.51
平均值	87.80	89.00	88.42

2.2 图谱抽取结果分析

以南水北调工程运行管理质量问题巡检专报、南水北调工程隐患排查记录等文本为抽取语料库, 基于 1.2.2 节中所述方法对该语料库进行实体关系三元组的抽取。其中包括隶属机构、工程站点、设备设施、部位及风险事件等实体类型以及发生时间等属性, 构建的关系类型包括隶属机构与工程站点之间的“管理”关系, 工程站点与设备设施之间的“包含”关系, 设备设施与部位之间的“组成部位”

关系, 部位与风险事件之间的“存在风险”关系, 风险事件与处置措施之间的“采取”关系。累计抽取实体关系三元组数量见表 2。此外, “发生时间”这一属性三元组抽取数量为 14 022 条。

表 2 图谱三元组数量统计
Tab. 2 Triplet count statistics

实体类型1	关系类型	实体类型2	关系数量/条
隶属机构	管理	工程站点	3 680
工程站点	包含	设备设施	5 779
设备设施	组成部位	部位	8 318
部位	存在风险	风险事件	14 160
风险事件	采取	处置措施	13 985

2.3 候选实体排序模型结果分析

为了全面评估本文提出的基于 BERT 预训练模型的孪生网络计算实体相似度的性能, 将其与单一 BERT 模型、Word2Vec 模型及传统 Jaccard 相似度算法进行对比, 其中 BERT、模型参数与孪生网络的骨干网络参数保持一致, 以检验孪生网络框架的有效性。模型的主要实验参数设置: epoch(训练轮数)为 50, Max_len(每次输入最大长度)为 128, Batch_size(每次训练实体数)为 16, Learning_rate(初始学习率)为 5×10^{-5} , Drop_rate(随机丢失率)为 0.1, Optimizer(优化器)为 Adam。手动构建了包含 2 536 对部位实体的相似度算法测试集, 对比了 3 种相似度算法在该测试集上的效果, 结果见表 3。

表 3 相似度算法比较结果
Tab. 3 The similarity algorithm compares the results %

算法	P	R	F_1
Jaccard	87.03	74.45	80.25
Word2Vec	85.24	71.86	77.98
Bert	82.00	80.90	81.43
Bert孪生网络	85.34	86.67	86.00

表 3 结果表明传统 Jaccard 算法和 Word2Vec 模型的 F_1 为 80.25% 和 77.98%, 与之相比 BERT 模型 F_1 值达到了 81.43%, 这验证了 BERT 中由 CLS 标记的隐藏向量能更好地表示字符串语义特征。与单一 BERT 相比, 本文设计的基于 BERT 预训练模型的孪生网络框架 F_1 提升了将近 5 个百分点, 这验证了 BERT 孪生网络框架的有效性。总体来看, BERT 孪生网络在字符串相似度计算任务上表现出

了较好的性能, 候选实体排序结果基本可靠。

2.4 工程风险信息推荐结果

工程相关风险信息对巡检人员定位当前风险等级意义重大, 基于 1.3 节设计的工程风险信息推荐算法, 结合巡检知识图谱的知识检索及可视化技术, 将巡检知识图谱中与当前部位实体匹配到的实体及实体关联信息可视化返回, 实现关联风险信息推荐。

以“排冰闸闸门前”为例, 使用模糊匹配从巡检知识图谱中生成包括“闸门、排冰闸闸门、平板闸门、检修闸门、进口闸门”等候选实体列表。接着使用 BERT 孪生网络框架计算“排冰闸闸门前”的字符串向量表示与候选实体列表中的候选实体逐一计算相似度, 最终相似度排名第一的候选实体为“排冰闸闸门”。将该实体及与该实体关联的所有工程风险信息可视化返回, 结果见图 5。该部位在不同时间里存在风险事件有“油缸压力异常、电融冰设备故障、电动葫芦接线故障、扰冰设备不合理”, 这些风险事件及处置措施等信息可有效提升巡检人员工作效率。

3 结论

本文对南水北调工程运行管理质量问题巡检专报文本及隐患排查记录的内容组成进行了系统分析和梳理, 结合工程巡检任务要求, 构建了巡检知识图谱概念模型, 并在此基础上利用实体关系联合抽取框架抽取结构化三元组知识, 以此构建巡检知识图谱, 提高了巡检信息的可访问性。

利用 BERT 预训练模型和孪生网络框架设计了 BERT 孪生网络, 将知识图谱与深度学习技术相结合用于工程风险信息的推荐, 辅助风险等级决策。实验验证 BERT 孪生网络在实体字符串相似度计算任务上表现出了较好的性能, 候选实体排序结果可靠, 为巡检人员进行风险诊断时提供可参考信息和风险处置方法, 有效提升巡检人员工作效率。

本研究成果具有强迁移性和推广价值, 利用知识图谱的知识组织和关联信息推荐方法可为其他相关领域中数字化工程运维方法的研究提供借鉴和参考。

本研究初步实现了巡检知识图谱的构建和工程风险信息的推荐, 但仍存在部分问题, 例如知识图谱的维护和知识更新问题, 以及巡检数据有时存在噪声或缺失现象可能会对深度算法的性能产生影响。在下一步工作中, 需进一步探索知识维护及更新算法, 同时扩大语料收集范围并优化巡检数据集,

以提升知识抽取框架的性能,实现更可靠、更高效、智能化发展。
 的南水北调工程智能巡检方法,推动工程运维数字

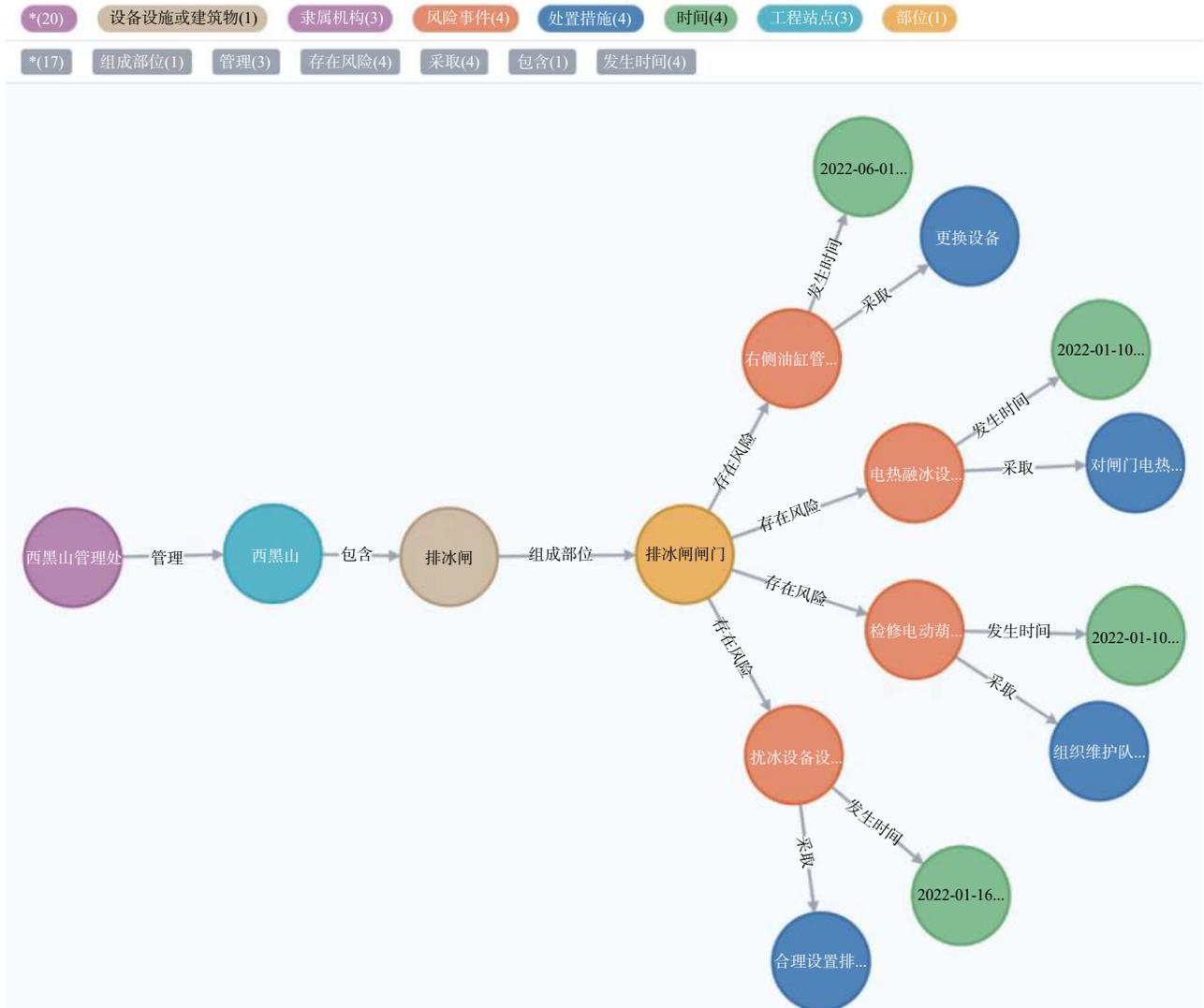


图 5 工程风险信息推荐结果

Fig. 5 Project risk information recommendation results

参考文献:

[1] 赵薛强, 凌峻. 无人机自动巡检智慧监控系统研究与应用[J]. 人民长江, 2022, 53(6): 235-241. DOI: 10.16232/j.cnki.1001-4179.2022.06.036.

[2] WU T, QI G, LI C, et al. A survey of techniques for constructing Chinese knowledge graphs and their applications[J]. Sustainability, 2018, 10(9): 3245. DOI: 10.3390/su10093245.

[3] 徐有为, 张宏军, 程恺, 等. 知识图谱嵌入研究综述[J]. 计算机工程与应用, 2022, 58(9): 30-50. DOI: 10.3778/j.issn.1002-8331.2111-0248.

[4] CHEN Z, WANG Y, ZHAO B, et al. Knowledge graph completion: A review[J]. IEEE Access, 2020, 8: 192435-192456. DOI:10.1109/ACCESS.2020.3030076.

[5] 欧阳文字, 叶磊, 顾学志, 等. 深度学习水文预报研究进展综述 II: 研究进展及展望[J]. 南水北调与水利科技(中英文), 2022, 20(5): 862-875. DOI: 10.13476/j.cnki.nsbdkq.2022.0087.

[6] 褚钰, 付景保, 朱九龙. 基于图模型的南水北调中线水源地生态环境多主体治理冲突分析[J]. 南水北调与水利科技(中英文), 2022, 20(6): 1179-1187. DOI: 10.13476/j.cnki.nsbdkq.2022.0116.

[7] TANG S, ZHU Y, YUAN S. Intelligent fault diagnosis of hydraulic piston pump based on deep learning and Bayesian optimization[J]. ISA transactions, 2022, 129: 555-563. DOI: 10.1016/j.isatra.2022.01.013.

[8] 刘雪梅, 卢汉康, 李海瑞, 等. 知识驱动的水利工程应急预案智能生成方法: 以南水北调中线工程为例

- [J]. *水利学报*, 2023, 54(6): 666-676. DOI: [10.13243/j.cnki.slxb.20220994](https://doi.org/10.13243/j.cnki.slxb.20220994).
- [9] XIA L, LIANG Y, LENG J, et al. Maintenance planning recommendation of complex industrial equipment based on knowledge graph and graph neural network[J]. *Reliability Engineering & System Safety*, 2023, 232: 109068. DOI: [10.1016/j.res.2022.109068](https://doi.org/10.1016/j.res.2022.109068).
- [10] RASMUSSEN M H, LEFRANÇOIS M, PAUWELS P, et al. Managing interrelated project information in AEC Knowledge Graphs[J]. *Automation in Construction*, 2019, 108: 102956. DOI: [10.1016/j.autcon.2019.102956](https://doi.org/10.1016/j.autcon.2019.102956).
- [11] SUN A Y, SCANLON B R. How can Big Data and machine learning benefit environment and water management: A survey of methods, applications, and future directions[J]. *Environmental Research Letters*, 2019, 14(7): 073001. DOI: [10.1088/1748-9326/ab1b7d](https://doi.org/10.1088/1748-9326/ab1b7d).
- [12] YAN J, LV T, YU Y. Construction and recommendation of a water affair knowledge graph[J]. *Sustainability*, 2018, 10(10): 3429. DOI: [10.3390/su10103429](https://doi.org/10.3390/su10103429).
- [13] HAO X, JI Z, LI X, et al. Construction and application of a knowledge graph[J]. *Remote Sensing*, 2021, 13(13): 2511. DOI: [10.3390/rs13132511](https://doi.org/10.3390/rs13132511).
- [14] 蒋玮, 王铭华, 陈锦铭, 等. 基于Neo4j图数据库的配电网供电可靠性计算[J]. *电力系统自动化*, 2022, 46(15): 104-111. DOI: [10.7500/AEPS20211031001](https://doi.org/10.7500/AEPS20211031001).
- [15] 岳丽欣, 刘自强, 许海云. 基于交互式可视化的领域知识图谱构建研究[J]. *情报科学*, 2020, 38(6): 145-150. DOI: [10.13833/j.issn.1007-7634.2020.06.021](https://doi.org/10.13833/j.issn.1007-7634.2020.06.021).
- [16] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [EB/OL]. <http://arxiv.org/abs/1810.04805>, 2019-05-24.
- [17] 赵志宏, 吴冬冬, 窦广鉴, 等. 一种基于CNN-Bi-GRU孪生网络的轴承故障诊断方法[J]. *振动与冲击*, 2023, 42(6): 166-171,211. DOI: [10.13465/j.cnki.jvs.2023.06.020](https://doi.org/10.13465/j.cnki.jvs.2023.06.020).
- [18] DÖRPINGHAUS J, STEFAN A, SCHULTZ B, et al. Context mining and graph queries on giant biomedical knowledge graphs[J]. *Knowledge and Information Systems*, 2022, 64(5): 1239-1262. DOI: [10.1007/s10115-022-01668-7](https://doi.org/10.1007/s10115-022-01668-7).
- [19] YANG Y, ZHU Y, JIAN P. Application of knowledge graph in water conservancy education resource organization under the background of Big Data[J]. *Electronics*, 2022, 11(23): 3913. DOI: [10.3390/electronics11233913](https://doi.org/10.3390/electronics11233913).
- [20] 段浩, 韩昆, 赵红莉, 等. 水利综合知识图谱构建研究[J]. *水利学报*, 2021, 52(8): 948-958. DOI: [10.13243/j.cnki.slxb.20200924](https://doi.org/10.13243/j.cnki.slxb.20200924).
- [21] MA X. Knowledge graph construction and application in geosciences: A review[J]. *Computers & Geosciences*, 2022, 161: 105082. DOI: [10.1016/j.cageo.2022.105082](https://doi.org/10.1016/j.cageo.2022.105082).
- [22] WU H, ZHONG B, LI H, et al. Combining computer vision with semantic reasoning for on-site safety management in construction[J]. *Journal of Building Engineering*, 2021, 42: 103036. DOI: [10.1016/j.jobe.2021.103036](https://doi.org/10.1016/j.jobe.2021.103036).
- [23] WANG X, EL-GOHARY N. Deep learning-based relation extraction and knowledge graph-based representation of construction safety requirements[J]. *Automation in Construction*, 2023, 147: 104696. DOI: [10.1016/j.autcon.2022.104696](https://doi.org/10.1016/j.autcon.2022.104696).
- [24] LIN J, ZHAO Y, HUANG W, et al. Domain knowledge graph-based research progress of knowledge representation[J]. *Neural Computing and Applications*, 2021, 33: 681-690. DOI: [10.1007/s00521-020-05057-5](https://doi.org/10.1007/s00521-020-05057-5).
- [25] RAFFEL C, SHAZEER N, ROBERTS A, et al. Exploring the limits of transfer learning with a unified text-to-text transfer transformer[J]. *The Journal of Machine Learning Research*, 2020, 21(140): 1-67.
- [26] 王艺霖, 靳燕国, 陈晓楠, 等. LSTM神经网络和量纲分析法在弧形闸门过流计算中的对比[J]. *南水北调与水利科技(中英文)*, 2022, 20(3): 590-599. DOI: [10.13476/j.cnki.nsbdkq.2022.0060](https://doi.org/10.13476/j.cnki.nsbdkq.2022.0060).
- [27] PETRONI F, ROCKTÄSCHEL T, LEWIS P, et al. Language models as knowledge bases?[EB/OL]. <https://doi.org/10.48550/arXiv.1909.01066>, 2019-09-04.
- [28] 杨阳蕊, 朱亚萍, 陈思思等. 融合群体智能策略的AI链在大坝防汛抢险知识推理中的应用[J]. *水利学报*, 2023, 54(9): 1122-1132. DOI: [10.13243/j.cnki.slxb.20230188](https://doi.org/10.13243/j.cnki.slxb.20230188).
- [29] QIU X, SUN T, XU Y, et al. Pre-trained models for natural language processing: A survey[J]. *Science*

- China Technological Sciences*, 2020, 63(10): 1872-1897. DOI: [10.1007/s11431-020-1647-3](https://doi.org/10.1007/s11431-020-1647-3).
- [30] YANG Y, ZHU Y, CHEN S, et al. API comparison knowledge extraction via prompt-tuned language model[J]. *Journal of Computer Languages*, 2023, 75: 101200. DOI: [10.1016/j.coala.2023.101200](https://doi.org/10.1016/j.coala.2023.101200).
- [31] 杨阳蕊, 朱亚萍, 刘雪梅, 等. 水利工程文本中抢险实体和关系的智能分析与提取[J]. *水利学报*, 2023, 54(7): 818-828. DOI: [10.13243/j.cnki.slxb.20220992](https://doi.org/10.13243/j.cnki.slxb.20220992).
- [32] LU Y, LIU Q, DAI D, et al. Unified structure generation for universal information extraction[EB/OL]. <https://doi.org/10.48550/arXiv.2203.12277>, 2022-03-23.
- [33] SARWAR B M. Sparsity, scalability, and distribution in recommender systems[M]. Minnesota: University of Minnesota, 2001.
- [34] HE L, YE W, WANG Y X, et al. Using knowledge graph and RippleNet algorithms to fulfill smart recommendation of water use policies during shale resources development[J]. *Journal of Hydrology*, 2023, 617: 128970. DOI: [10.1016/j.jhydrol.2022.128970](https://doi.org/10.1016/j.jhydrol.2022.128970).
- [35] AMADOR E, SERRANO E, MANRIQUE D, et al. Prediction and decision-making in intelligent environments supported by knowledge graphs, a systematic review[J]. *Sensors*, 2019, 19(8): 1774. DOI: [10.3390/s19081774](https://doi.org/10.3390/s19081774).
- [36] 张茂宇, 李海明. 融合双重语义相似度的混合协同过滤推荐算法[J]. *计算机工程与设计*, 2023, 44(9): 2671-2678. DOI: [10.16208/j.issn1000-7024.2023.09.015](https://doi.org/10.16208/j.issn1000-7024.2023.09.015).
- [37] LIU X, LU H, LI H. Intelligent generation method of emergency plan for hydraulic engineering based on knowledge graph: take the South-to-North Water Transfers Project as an example[J]. *LHB: Hydroscience Journal*, 2022, 108(1): 2153629. DOI: [10.1080/27678490.2022.2153629](https://doi.org/10.1080/27678490.2022.2153629).
- [38] ARTSTEIN R, POESIO M. Inter-coder agreement for computational linguistics[J]. *Computational Linguistics*, 2008, 34(4): 555-596. DOI: [10.1162/coli.07-034-R2](https://doi.org/10.1162/coli.07-034-R2).

Knowledge-driven recommended method for inspection information of South-to-North Water Transfers Project

YANG Yangrui¹, ZHU Yaping¹, LIU Xuemei^{1,2}, CHEN Sisi¹, LI Huimin³

(1. School of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450000, China; 2. Collaborative Innovation Center for Efficient Utilization of Water Resources, Zhengzhou 450000 China; 3. School of Water Conservancy, North China University of Water Resources and Electric Power, Zhengzhou 450000, China)

Abstract: There are many risks in the operation of the South-to-North Water Transfers Project, and the inspection work of the South-to-North Water Transfers Project is of great significance to ensure the safety and stability of the project. Traditional project inspection methods mainly rely on manual experience and have a low degree of digitalization. Due to the uneven professional level of inspection personnel, it is difficult to form a unified record standard, which in turn leads to redundant inspection special report information, among which the accessibility of effective information is poor, making the traditional project inspection method inefficient.

As a new generation of information technology, knowledge graph is a powerful tool for knowledge organization and management. In order to alleviate the limitations of traditional project inspection methods and improve the efficiency of project inspection, knowledge graph with deep learning technology was combined, and using knowledge graph to empower intelligent inspection of the South-to-North Water Transfers Project was proposed. Specifically, the inspection knowledge graph was constructed based on the project inspection text, and an project risk information recommendation method was designed based on the inspection knowledge graph. In the process of building the inspection knowledge graph, the conceptual model of the inspection knowledge graph is defined based on expert experience, and on this basis, the entity relationship joint extraction framework is used to extract the structured triplet knowledge from the unstructured project inspection text, and the knowledge visualization is carried out with the Neo4j graph database as the carrier. The inspection knowledge graph clearly presents inspection information such as engineering sites, parts, risk events, and disposal measures in the form of entity-relationship-

entity triples, and supports knowledge visualization and knowledge retrieval, which alleviates the limitation of poor accessibility of effective information in inspection reports. In the project risk information recommendation method, the Bert pre-training model and twin network framework were used to design the Bert twin network, and recommends the project risk information of the entities and entities associated with the current part in the inspection knowledge graph to the inspectors by calculating the string similarity of the part entities, and assists the inspectors in the project risk level diagnosis.

The quality of the knowledge graph and the effectiveness of the method are evaluated experimentally. The experimental results show that the average F_1 value of various relational triples extracted is 88.42%, and the knowledge extraction results have high accuracy, and the quality of the knowledge graph is considered to be reliable. The F_1 value of the candidate entity ranking model designed reaches 86%, which is higher than that of the traditional Jaccard algorithm and Word2Vec model.

In general, the Bert twin network designed shows good performance in the string similarity calculation task, and the project risk information recommendation results based on patrol inspection knowledge graph are basically reliable. Knowledge graph and deep learning technology were introduced into the intelligent application of project inspection, which realizes the deep correlation and effective use of inspection knowledge, which can provide reference significance for improving the operation and maintenance efficiency of the South-to-North Water Transfers Project and strengthening the risk management capability of the project.

Key words: knowledge graph; South-to-North Water Transfers Project; recommended engineering inspection information; project risk level diagnosis

水利部办公厅日前发出通知,确定我国纪念2024年“世界水日”“中国水周”活动主题为“精打细算用好水资源,从严从细管好水资源”,并明确了44条宣传口号。2024年3月22日是第三十二届“世界水日”,3月22日—28日是第三十七届“中国水周”。联合国确定2024年“世界水日”主题为“Water for Peace”(以水促和平)。

来源:2024年“世界水日”“中国水周”活动主题发布[N].中国水利报,2024年3月12日(第1版)。

摄影|曲帅超