

# 基于深度学习的施工安全隐患整改智能推荐系统

刘震<sup>1</sup>, 赵嵩<sup>2</sup>, 杨涛<sup>3</sup>, 蔡太伟<sup>4</sup>

1. 广东粤海珠三角供水有限公司, 广东 广州 511455;
2. 云南大学信息学院, 云南 昆明 650504;
3. 深圳市科荣软件股份有限公司, 广东 深圳 518063;
4. 华南师范大学华南先进光电子研究院, 广东 广州 510006

## 摘要

水利工程施工安全隐患治理正向信息化与智能化转型, 为了高效地从大量非结构化的施工安全隐患数据中挖掘出有价值的潜在信息, 提出了基于深度学习的施工安全隐患整改智能推荐系统。该算法基于词频-逆向文档频率算法, 提取施工安全隐患的特征词, 构建安全隐患关联桑基图, 展示施工标段、隐患特征、隐患类型之间的信息流动特征; 基于FP-Growth算法挖掘历史数据中的关联规则; 结合序列相似度匹配(sequence similarity matching, SSM)算法和Doc2Vec模型, 优化案例检索推荐的过程。该算法利用珠江三角洲水资源配置工程2019—2023年记录的80 953条施工安全隐患信息作为数据源。实例验证表明, 该算法能够为当前的施工安全隐患匹配出较为准确的整改措施, 可有效辅助施工安全管理者排查治理隐患问题。

## 关键词

施工安全; 关联分析; 深度学习; 智能推荐

中图分类号: TP183

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2023076

## *Intelligent recommendation system for rectification of construction safety hazards based on deep learning*

LIU Zhen<sup>1</sup>, ZHAO Song<sup>2</sup>, YANG Tao<sup>3</sup>, CAI Taiwei<sup>4</sup>

1. GD Holdings Pearl River Delta Water Supply Co., Ltd., Guangzhou 511455, China
2. School of Information Science and Engineering, Yunnan University, Kunming 650504, China
3. Shenzhen Koron Software Co., Ltd., Shenzhen 518063, China
4. South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou 510006, China

## Abstract

The management of safety hazards in water conservancy engineering construction is transitioning towards informatization and intelligence. In order to efficiently mine valuable potential information from a large amount of unstructured construction safety hazard data, an intelligent recommendation system for construction safety hazard

rectification based on deep learning is proposed. This paper is based on the TF-IDF algorithm to extract feature words of hidden danger, construct a safety hazard association Sankey diagram and display the information flow characteristics among construction sections, hazard features and hazard types. Then, this paper mines association rules in historical data based on the FP-Growth algorithm. In addition, the process of case retrieval recommendation is optimized by combining the sequence similarity matching algorithm and the Doc2Vec model. This paper uses 80 953 construction safety hazard information as the data source, which is recorded in the water resources allocation project of Pearl River Delta from 2019 to 2023. Example verification shows that the proposed method can match accurate rectification measures for current construction safety hazards, effectively assisting construction safety managers to identify and address hidden danger.

### Key words

construction safety, correlation analysis, deep learning, intelligent recommendation

## 0 引言

随着信息技术的不断发展,水利工程建设安全管理正向信息化和智能化转型<sup>[1]</sup>。施工安全隐患的排查治理是工程建设中安全管理的重要手段。随着信息化系统的普及与应用,人工排查、手动录入的施工安全隐患管理在排查过程中积累了大量非结构化的安全隐患文本数据,但这些数据尚未得到充分利用。因此,利用人工智能从海量的历史数据中挖掘出隐藏信息和潜在规律,从而促进水利工程建设由信息化模式向智能化模式发展<sup>[2]</sup>,对于提高施工安全隐患的排查治理效率具有重要的现实意义。

目前,不少学者对安全隐患文本数据的挖掘展开了研究。例如:刘梅等<sup>[3]</sup>利用相关性检验挖掘安全隐患特征之间的关联;谭章禄等<sup>[4]</sup>利用狄利克雷分配模型挖掘煤矿安全隐患,揭示了生产单位、责任主题与隐患致因之间的关系;陈述等<sup>[5]</sup>通过短语提取技术揭示了安全隐患时空分布特征;林旭杰等<sup>[6]</sup>采用Apriori关联算法挖掘煤矿安全隐患之间的关联规则;Le等<sup>[7]</sup>、Jatnika等<sup>[8]</sup>将深度学习的方法运用到提高建筑工程术语语义相似度计算

的准确性上,为隐患文本知识挖掘增加了可信度。为了从历史案例中挖掘出有用的信息和经验,1995年Kumar<sup>[9]</sup>首次将案例推理技术应用到工程设计领域,为案例推理技术在各种工程领域的应用提供了理论基础。例如:郑霞忠等<sup>[10]</sup>通过融合案例推理与深度学习的方法,结合历史安全隐患数据来辅助隐患治理方案的制订;原江涛等<sup>[11]</sup>基于案例推理技术,提出了一种煤矿安全隐患排查治理信息系统并应用于实际生产;夏登友等<sup>[12]</sup>利用情景元技术对案例进行描述和表示,提出了一种基于规则的推理方法,并实现了一个应急决策支持系统,对相关领域的应急决策提供了有效的帮助。

综上所述,以往的研究大多聚焦于隐患问题的智能分类和隐患问题关联规则的挖掘,忽略了历史隐患案例中潜藏的信息。为了充分挖掘安全隐患历史案例中的有用信息和经验,本文从历史案例视角对施工安全隐患进行分析,帮助安全管理者深入探析隐患事件发生的特征和规律,并根据相似的隐患案例制订有效的隐患预防措施,从而降低类似隐患的发生概率。类似的视角还包括事故因素分析和风险评估等。为此,本文结合文本特征提取、关联规则挖掘和文本相似度计算等方法,提出了施工安全隐患整改智能推荐系统。该方法

融合SSM算法和Doc2Vec模型来优化检索推荐过程,并在检索阶段分析相关历史案例信息。计算隐患描述之间的相似度时,考虑了上下文逻辑和短文本语义特征。最后,参考相似度最大的历史案例,将检索出的整改措施作为当前隐患问题的推荐整改方案。

## 1 数据来源与预处理

### 1.1 数据来源

以珠江三角洲水资源配置工程为研究对象,在该工程的建设过程中,安全检查单位每月对其负责的施工标段进行安全检查,检查过程中检查人员发现施工现场存在安全隐患问题,并指示施工单位在规定的期限内进行整改,之后将检查和整改记录上传到安全管理信息系统。本文的研究数据来源于从安全管理系统中获取到的2019—2023年期间80 953条安全隐患原始数据,其中,将2019—2022年期间的65 714条数据作为历史案例数据,2022—2023年期间的15 239条数据作为测试数据。每条安全隐患数据主要包含标段、隐患描述、隐患类型、整改措施和检查日期等字段,前4个字段均为非结构化的文本数据。其中,隐患类型分为环境隐患、人的不安全行为、管理隐患、设备及物料隐患4类。部分安全隐患记录见表1。

### 1.2 预处理

为了获得有效的施工安全隐患数据,本文结合工程施工安全隐患的判定标准等相关规范,手动对数据进行了处理。首先,人工记录的数据可能存在含有主观推断的信息、缺失值、异常值等数据,因此,手动剔除上述信息以获得有效的安全隐患数据。其次,针对水利工程施工安全领域的特点,构建了该领域的安全隐患字典用于辅助分词,包括手动添加专有名词到自定义词典,例如“高处坠落”“电气安全”“脚手架”等。这能够完善分词效果,有效避免术语被错误分开或合并的情况发生,从而提高数据处理和精确度。再次,采用哈工大停用词表,并将不规范的关键词、无意义的词添加到停用词表中,用于去除隐患问题描述文本中的停用词,例如空格、标点符号等影响文本处理与分析的无效信息。最后,采用了Jieba分词对隐患问题描述文本进行分词。

## 2 研究方法

### 2.1 基于TF-IDF算法的隐患特征提取

词频-逆向文档频率(term frequency-inverse document frequency, TF-IDF)是一种常用的文本特征提取算法。TF-IDF算法可以提取出文档中的关键词,评

表1 部分安全隐患记录

标段	隐患描述	隐患类型	整改措施	检查日期
A1	沉淀池无标识牌,无安全防护	环境隐患	悬挂标识牌,做好防护	2019-09-01
B1	现场人员存在吸烟现象	人的不安全行为	进行现场安全教育	2021-01-21
C1	现场多处电缆损坏且拖地	管理隐患	对电缆进行更换	2022-10-05

估提取出的关键词在文档集合中的重要程度。关键词的重要程度与该关键词在文档中的出现频率 (term frequency, TF) 成正比。TF的计算方式如式(1)。

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (1)$$

其中,  $n_{i,j}$  是安全隐患词语  $i$  在安全隐患文档  $j$  中出现的次数, 分母则表示安全隐患文档中所有词汇出现的次数总和。关键词的重要程度与该词在文档集中出现的频率 (inverse document frequency, IDF) 成反比。IDF值计算方式如式(2)。

$$IDF_i = \log \left( \frac{|D|}{|\{j:t_i \in d_j\}| + 1} \right) \quad (2)$$

其中,  $|D|$  表示语料库中的文档总数,  $d_j$  表示文档样本,  $|\{j:t_i \in d_j\}|$  表示包含词语  $t_i$  的文档数目。将关键词的TF值和IDF值进行乘积, 得到该词的TF-IDF值, 该值越大表示该关键词在文档中的重要程度越高<sup>[13]</sup>。文本特征选择还有互信息算法、信息增益算法、卡方检验算法等<sup>[14]</sup>。在本文的研究数据中, 每条数据通常只包含很少的关键词, 数据非常稀疏, 故采用TF-IDF算法来提取安全隐患中的关键词作为隐患特征。

## 2.2 基于深度学习的施工安全隐患整改智能推荐系统

在隐患排查治理中, 安全隐患具有高重复性, 因此, 可以借助历史安全隐患治理方案, 缩短查询隐患相关知识的时间, 及时制订隐患整改措施。除此之外, 安全隐患之间还具有相关性, 一个隐患的发生往往可能导致其他隐患的出现。在复用历史安全隐患治理经验的同时, 可挖掘出与当前隐患关联的一系列安全隐患问题并给出

整改措施, 从而提高隐患治理的效率, 实现无隐患早防控、有隐患早发现和早治理的目标。为此提出了基于深度学习的施工安全隐患整改智能推荐系统, 系统框架如图1所示。

在对施工安全隐患整改智能推荐系统的研究中, 面临整体数据规模大且存在较多稀疏数据的挑战。当入库一条安全隐患时, 首先, 采用TF-IDF算法提取出隐患的特征, 每个隐患特征都包括一个或多个安全隐患, 每个安全隐患至少归类到一个隐患特征中。其次, 采用FP-Growth算法从频繁项集列表中挖掘出与当前安全隐患特征相关联的频繁项 (安全隐患特征集), 再将这些安全隐患特征下的安全隐患案例作

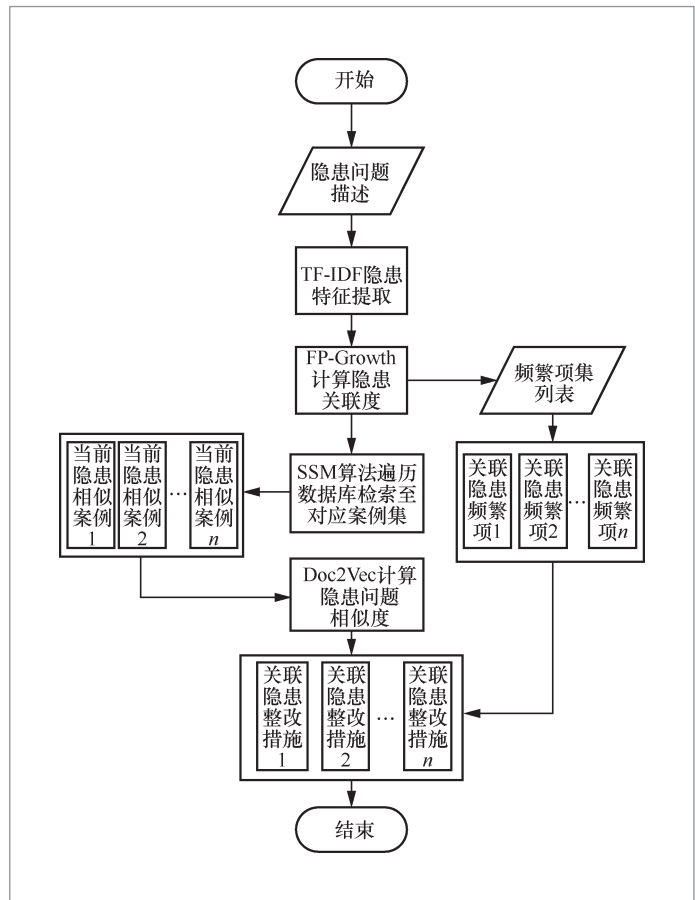


图1 基于深度学习的施工安全隐患整改智能推荐系统框架

为当前安全隐患潜在的预警信息。然后,利用SSM算法对当前入库的安全隐患与数据库中的历史安全隐患案例进行初步匹配,得到粗糙的相似案例集合,进一步采用Doc2Vec模型来计算当前安全隐患与相似案例集合中每一条安全隐患的相似度。最后,合并相似度最高的安全隐患案例与关联的安全隐患案例,将其推荐为当前安全隐患的整改方案。

### 2.2.1 基于FP-Growth算法的隐患关联规则挖掘

关联规则是一种描述不同项集之间关联关系的表达式,通常采用 $X \rightarrow Y$ 的形式表示,其中 $X$ 和 $Y$ 是不相交的项集。常用的关联规则挖掘算法包括Apriori算法、FP-Growth算法等。Apriori算法需要生成大量的候选集,在处理大规模数据时会非常耗时间和空间。而FP-Growth算法通过压缩数据、构建FP树去除了生成候选集的过程,大大减少了时间和空间的消耗。因此,本文采用FP-Growth算法挖掘安全隐患之间的关联规则,从而建立频繁项集列表。

FP-Growth算法主要分为两个步骤:构建FP树和基于FP树生成频繁项集<sup>[15]</sup>。FP树是一种基于频繁模式挖掘的数据结构,用于高效地存储和查找数据集中的频繁项集。FP树由一个根节点和多个项节点组成,每个项节点表示一个频繁项,每个节点包括一个计数器和指向相同项节点的链表指针。构建FP树的过程如下:遍历数据集统计每个项的支持度计数,根据支持度计数构建项头表;按照支持度降序遍历数据集,将事务中的每个项按顺序加入根节点;为每个节点创建一个初始值为1的计数器,如果该节点存在项节点,则计数器加1,以此递归地构建FP树。基

于FP树生成频繁项集的过程如下:从根节点开始,依次遍历每个频繁项的链表,生成以该项为结尾的频繁项集;采用递归方法,在每个以该项为结尾的前缀路径上构建条件模式基,从而继续生成更长的频繁项集。

### 2.2.2 SSM算法

SSM算法<sup>[16]</sup>的原理是通过计算两个序列之间的最长公共子序列(longest common subsequence, LCS)的长度来计算两个序列的相似度。假设两个序列分别为 $X$ 和 $Y$ ,LCS( $X, Y$ )的长度为 $\text{len}(\text{LCS}(X, Y))$ ,那么它们的相似度计算如式(3)。

$$\text{Similarity}(X, Y) = \frac{2 \times \text{len}(\text{LCS}(X, Y))}{\text{len}(X) + \text{len}(Y)} \quad (3)$$

与SSM算法类似的算法还有编辑距离算法、Jaccard相似度算法和余弦相似度算法等。相比于编辑距离算法<sup>[17]</sup>和Jaccard相似度算法<sup>[18]</sup>,SSM算法可以处理不同长度的序列。在施工安全隐患数据中,往往会存在序列长度不同的隐患数据。如果使用编辑距离算法和Jaccard相似度算法,则需要对数据进行维度对齐,这样会导致部分信息丢失。与余弦相似度算法相比,SSM算法考虑了序列中元素的顺序,能够发现相同子序列的位置和顺序。综上所述,SSM算法在文本相似度匹配、序列相似度匹配方面表现更优。因此,本研究利用SSM算法对当前入库的安全隐患与数据库中的历史安全隐患案例进行初步匹配,得到粗糙的相似案例集合。

### 2.2.3 基于Doc2Vec模型计算文本相似度

Doc2Vec<sup>[19]</sup>是一种深度学习模型,用于将一个文档表示为固定长度的向量,它是

Word2Vec模型<sup>[20-21]</sup>的扩展。Word2Vec模型可以将单个单词表示为向量,而Doc2Vec在训练模型时不仅考虑了每个单词的上下文信息,还考虑了整个文档的语境,为每个文档生成一个唯一的向量表示。Doc2Vec有两种算法,分别是分布记忆(distributed memory, DM)算法和分布词袋(distributed bag of words, DBOV)算法。DM算法将文档的向量作为额外的输入传递给模型,然后预测文档中的单词,其结构如图2所示。

DBOV算法结构如图3所示。在DBOV算法中,每个句子都被视为一个“袋子”,每个单词的顺序被忽略,每个单词都被独立地考虑。而该模型的目标是在不考虑上下文的情况下,根据整个句子预测中心词。相比DM算法,DBOV算法更简单和快速,通常适用于文本分类等任务,而DM模型则更适合语义相关性和相似性的建模任务。本文使用Doc2Vec模型中的DM算法来计算当前安全隐患与经过SSM算法匹配得到的粗糙相似案例集合中每一条安全隐患的相似度。在这个过程中,Doc2Vec模型首先会将相似案例集合中的每条安全隐患都转化为唯一的向量表示,通过计算它们之间的向量余弦相似度来衡量它们之间的相似度。

### 3 结果分析

各施工标段安全隐患的数量分布如图4所示。由图可知, B3、B4、C1是安全隐患高发的3个标段,因此,选用这3个标段的数据进行隐患特征挖掘与可视化。

#### 3.1 隐患特征分析

利用B3、B4和C1这3个标段的安全隐患数据来绘制桑基图<sup>[22]</sup>。首先,将每个标

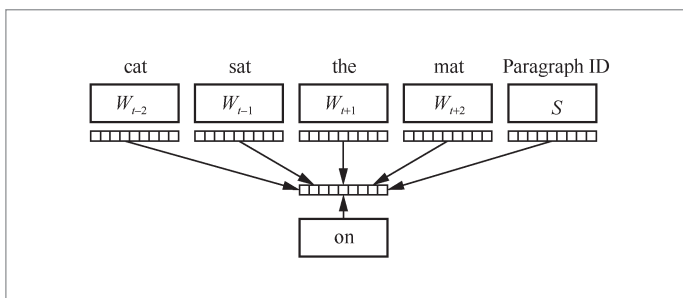


图2 DM 算法结构

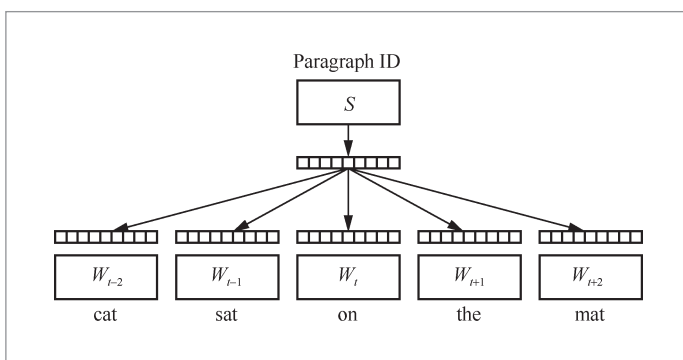


图3 DBOV 模型结构

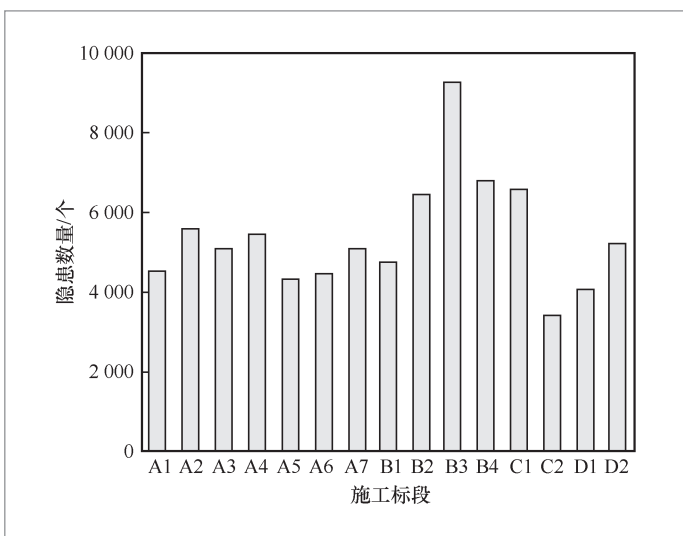


图4 各个施工标段隐患的数量分布

段的所有安全隐患数据作为一个文档,采用TF-IDF算法提取每个文档的关键词,选择TF-IDF值较大的前几个关键词作为对应施工标段的安全隐患特征;其次,利用RAWGraphs2.0软件绘制施工标段-隐

患类型桑基图,如图5所示。该桑基图从左至右依次表示施工标段、隐患特征和隐患类型,每个节点的宽度表示该隐患特征TF-IDF值大小,节点之间的分支代表信息的流动,分支的宽度则反映了信息流量的大小。从桑基图中可以直观地了解到不同的施工标段各自存在的主要安全隐患问题特征。竖向分析显示,配电箱、灭火器和钢筋加工棚这三个隐患特征对应的节点宽度最大,说明各标段发生与这三个特征有关的安全隐患最多。而从横向角度来看,这三个特征词包含的信息流宽度最大,这表明与它们有关的隐患问题发生的频率最高。从施工标段的角度来看,B3和B4标段易发生与配电箱和灭火器相关的隐患,

C1标段易发生与隧洞和钢筋加工棚相关的隐患。

## 3.2 安全隐患整改推荐结果分析

### 3.2.1 关联规则挖掘结果分析

在分析当前隐患问题时,首先采用TF-IDF算法对当前隐患数据进行特征提取,其次根据FP-Growth算法挖掘关联规则,置信度的阈值为0.5,支持度的阈值为0.002,再经过人工筛选,最终得到了7 688条关联规则。部分关联规则见表2。以第一条关联规则为例,它表示当配电箱出现时,通常会伴随着“不规范”“灭火器”“接线”这

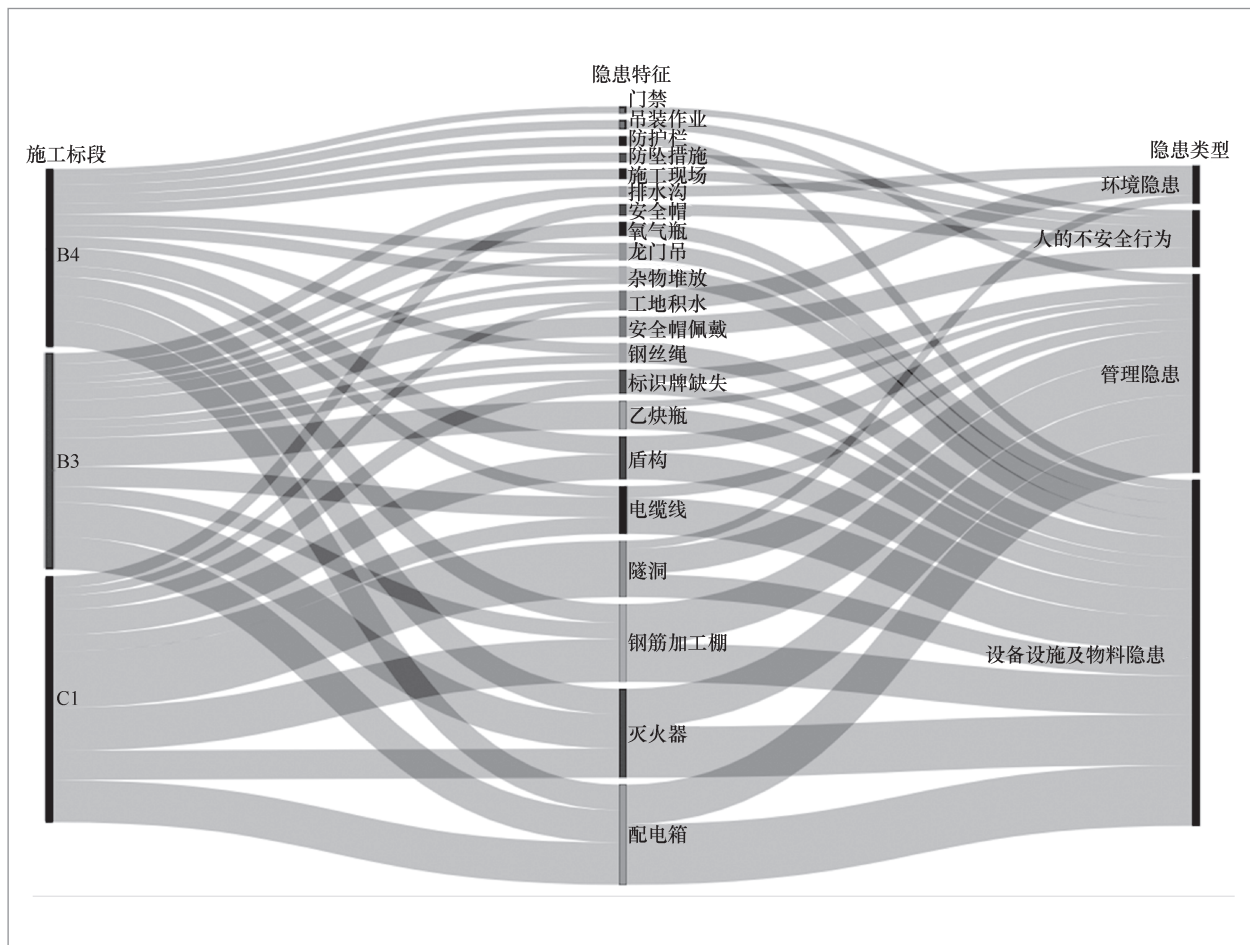


图5 施工标段 - 隐患类型桑基图

3个事务；支持度为0.0031，说明同时包含“配电箱”“不规范”“灭火器”和“接线”的事务数比较少；置信度为0.75，意味着当一条安全隐患记录中出现了配电箱时，有75%的概率出现“不规范”“灭火器”“接线”这3个事务。

如图6所示，当安全检查人员发现并记录隐患后，通过TF-IDF提取隐患的特征，得到“安全距离”“防护措施”等安全隐患特征。以“安全距离”这一特征为例，通过关联规则得到“警示”“设置”“标识牌”等频繁项集。再以“警示”这一频繁项为例，能够匹配到相似的安全隐患问题描述，并且检索出对应的整改方案。

### 3.2.2 相似度计算实验分析

从2022—2023年期间的测试数据中，分别从环境隐患、人的不安全行为、管理隐患、设备设施及物料隐患这4个类型中依次随机抽取300条作为隐患问题描述测试样本。为了使抽取的样本更具有代表性，

表2 部分关联规则

序号	条件	结果	置信度	支持度
1	配电箱	不规范、灭火器、接线	0.75	0.0031
2	区域、人员	警示、设置、安全帽	0.66	0.0023
3	电缆线	拖地、破损	0.67	0.0024
4	施工、警示	设置、临时	0.50	0.0024
5	杂物	堆放、灭火器	0.53	0.0027
6	标识牌	记录、拖地	1.00	0.0025
7	电缆、防护栏杆	防护、基坑	0.55	0.0030
8	高处作业	安全带、佩戴、安全帽	0.76	0.0032
9	气瓶	防倾倒、措施	0.82	0.0045
10	危险源	公示牌、警示	0.92	0.0026
...	...	...	...	...

抽取的样本涵盖了水利工程建设中的8种不同作业内容，如图7所示。

将从各个安全隐患类型中抽取出来的1200条测试样本分别通过SSM算法匹配到对应的案例集，再通过Doc2Vec模型计算案例集中的安全隐患与当前隐患的相似度<sup>[23]</sup>，最终综合准确率为0.869。部分安全隐患相似度匹配样例见表3。

为了进一步验证该推荐系统的有效性，将上述系统推荐的整改措施与安全

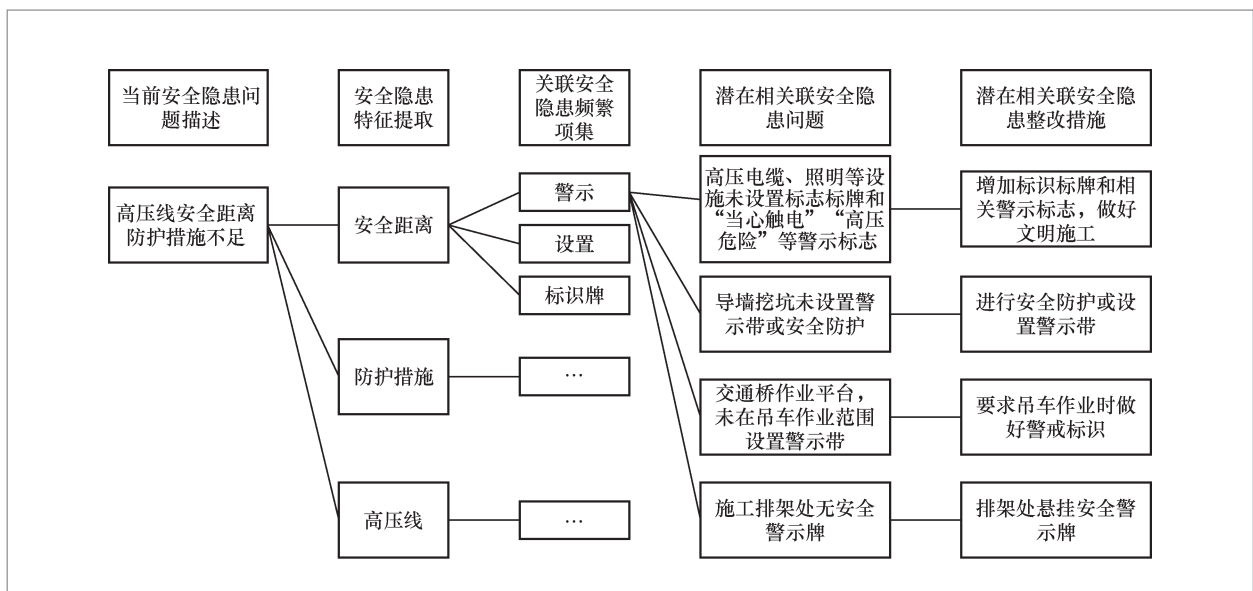


图6 关联规则分析

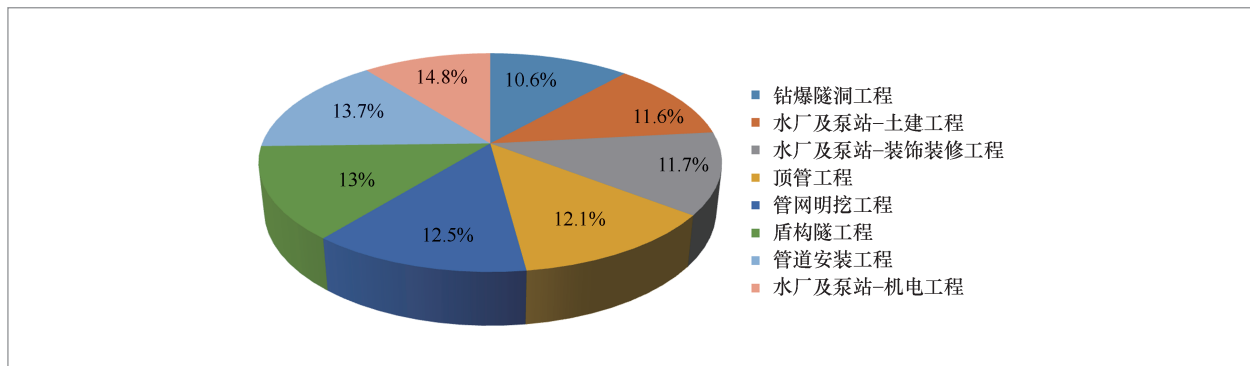


图7 不同作业内容测试样本占比

表3 安全隐患数据相似度表(部分)

编号	当前隐患描述	历史案例数据库	相似度
1	左线一电箱未及时贴标识牌	配电箱未张贴标识牌	0.903
2	高压线安全距离防护措施不足	220 kV高压线安全距离防护措施不足	0.891
3	仓库氧气瓶未专库存放	氧气瓶未使用时专库存放	0.869
4	现场安全警示牌排序未按标准进行挂设	工作井井底施工作业架安全警示牌排序未按标准进行挂设	0.782
5	入库输水箱涵处木工盘锯无防护罩	木工盘锯无防护罩	0.879
6	北三角营地宿舍私拉乱接, 违规使用插线板	西南角营地宿舍内私拉乱接, 违规使用插线板	0.891
7	灭火器无巡查记录	仓库灭火器无巡查记录	0.888
8	现场交叉道路无减速带	工区现场大门车道缺少减速带	0.785
9	工作井道路积土严重、扬尘较多	工作井场地积水、污泥严重	0.783
10	工作井泥浆池临边防护不到位	工作井临边防护不到位	0.880
...	...	...	...
综合准确率		0.869	

管理者制定的整改措施进行对比<sup>[24]</sup>, 推荐准确率采用Doc2Vec模型计算的文本相似度, 安全隐患整改推荐系统的综合准确率为0.914。见表4, 本文提出的施工安全隐患智能推荐系统得到的安全隐患整改措施与安全管理者制订的安全隐患整改措施一致性较强, 该推荐系统能较为准确地匹配出当前安全隐患的整改措施。

本文针对同一输入, 对比分析分别采用SSM、Doc2Vec、SSM+Doc2Vec这3种模型得到的相似度排名前5条的安全隐患, 各模型效果见表5。以输入“焊工棚二氧化碳气瓶无防护棚、无防倾倒措施”为例, SSM模型注意到了“无防倾倒措

施”“二氧化碳”“气瓶”等特征词, 模型表现一般; 在Doc2Vec模型中, “防护棚”这一特征词的权重较大, 模型匹配效果最差; SSM+Doc2Vec模型首先经过SSM筛选出相似度排名前1 000的安全隐患, 再利用Doc2Vec模型将这1 000条安全隐患转化为唯一的向量表示, 最后计算这些向量余弦相似度来衡量它们之间的相似度, 该模型同时注意到了“焊工棚”“无防倾倒措施”“二氧化碳”“气瓶”等特征词, 不仅降低了模型的计算量, 而且保留了关键的隐患特征词, 故其综合表现最好。

在从测试数据中随机采样得到的1 200个测试样本上, 采用Doc2Vec模型

表4 部分安全隐患整改措施推荐准确率

编号	相似案例推荐整改措施	安全管理者制定整改措施	相似度
1	加强现场人员教育,正确佩戴安全帽	对现场不正确佩戴安全帽人员进行教育	0.934
2	及时修复损坏警示标志牌	及时加设安全提示标语有安全标志牌	0.869
3	加设相应灭火器	及时加配灭火器	0.897
...	...	...	...
综合准确率		0.914	

表5 各模型效果对比表(部分)

输入	模型	输出
焊工棚二氧化碳气瓶无防护棚、无防倾倒措施	SSM	焊工棚二氧化碳气瓶无防倾倒措施
	Doc2Vec	二氧化碳气瓶无防倾倒措施 工作井内二氧化碳气瓶无防倾倒措施 工作井二氧化碳气瓶无防倾倒措施、无防震圈、无防撞帽 钢筋加工场所二氧化碳气瓶无防倾倒措施 颜屋工区料仓防护棚揽风绳花篮螺栓无保险扣 洞门墙上的2级配电箱需要加防护棚 检修交通洞工区拌和站氧气、乙炔瓶无防护棚和防倾倒措施 焊工棚二氧化碳气瓶无防倾倒措施 配电箱无防护棚
	SSM+Doc2Vec	焊工棚气瓶无防倾倒措施 焊工棚气瓶未做防倾倒 焊工棚氧气瓶无防倾倒 焊工棚氧气瓶无防倾倒 焊工棚二氧化碳气管开裂较多

计算文本的相似度,各模型的综合准确率见表6。SSM+Doc2Vec比SSM高0.032,比Doc2Vec高0.056。因此,本文提出的SSM+Doc2Vec模型在水利工程施工安全隐患文本上的表现优于单独采用SSM算法和Doc2Vec的方法。

### 3.2.3 算法性能对比与分析

为了进一步验证算法的优越性,采用相同的数据集并配置相同的实验环境,将本文提出的方法与文献[10]提出的基于Word2Vec计算目标案例与历史案例相似度的方法进行对比。结果表明,本文提出的方法在安全隐患整改推荐上的综合准确

表6 各模型综合准确率

模型	综合准确率
SSM	0.837
Doc2Vec	0.813
SSM+Doc2Vec	0.869

率达到0.869,优于文献[10]取得的0.802。当数据整体规模大且存在较多稀疏数据时,本文提出的方法采用SSM+Doc2Vec模型,可以更全面地挖掘案例描述的语义信息,在安全隐患整改推荐中能够提供更加准确的结果。

综上所述,基于深度学习的施工安全隐患整改智能推荐系统从多个方面提高了智能推荐方案的准确率和速度。首先,采

用TF-IDF算法和FP-Growth算法提取和挖掘安全隐患的关联规则,可以更加准确地找到与当前安全隐患相关联的案例,从而提高了整改方案的完整性。其次,利用SSM算法初步匹配历史案例和当前入库案例,减少了后续模型的计算量,提高了系统的运行效率。最后,采用Doc2Vec模型计算当前安全隐患与相似案例集合中每一条安全隐患的相似度,推荐最符合当前情况的整改方案,辅助安全管理人员在隐患管理工作中更好地进行决策。

## 4 结束语

本文构建了基于深度学习的施工安全隐患整改智能推荐系统。首先,基于TF-IDF算法进行安全隐患特征提取,并通过桑基图可视化安全隐患特征。其次,通过SSM算法对当前入库的安全隐患与数据库中的历史安全隐患案例进行初步匹配,得到粗糙的相似案例集合。然后,采用Doc2Vec模型计算当前安全隐患与相似案例集合中每一条安全隐患的相似度,合并相似度最高的安全隐患案例与关联的安全隐患案例,将其推荐为当前安全隐患的整改方案。本文在复用历史隐患治理经验的同时,挖掘出历史隐患数据中存在的关联规则,为安全隐患的排查治理提供了更全面的视角。验证结果表明,本文方法在安全隐患整改智能推荐任务上表现出色,能够快速、准确地为当前安全隐患问题推荐整改方案。在未来的工作中,将进一步完善安全隐患关联规则库和隐患整改推荐的方案。

## 参考文献:

- [1] 陈利. 对水利工程建设管理信息化、智能化的几点思考[J]. 农业科技与信息, 2022, 19(21): 126-128.  
CHEN L. Thoughts on informatization and intelligentization of water conservancy project construction management[J]. Information of Agricultural Science and Technology, 2022, 19(21): 126-128.
- [2] 钟登华, 时梦楠, 崔博, 等. 大坝智能建设研究进展[J]. 水利学报, 2019, 50(1): 38-52, 61.  
ZHONG D H, SHI M N, CUI B, et al. Research progress of the intelligent construction of dams[J]. Journal of Hydraulic Engineering, 2019, 50(1): 38-52, 61.
- [3] 刘梅, 许林宇, 廖彬超, 等. 基于数据驱动的施工安全隐患特征网络分析与预警策略[J]. 清华大学学报(自然科学版), 2023, 63(2): 191-200.  
LIU M, XU L Y, LIAO B C, et al. Data-driven network analysis of construction hazard characteristics and warning strategy[J]. Journal of Tsinghua University (Science and Technology), 2023, 63(2): 191-200.
- [4] 谭章禄, 王泽, 陈晓. 基于LDA的煤矿安全隐患主题发现研究[J]. 中国安全科学学报, 2016, 26(6): 123-128.  
TAN Z L, WANG Z, CHEN X. Research on topic extraction for coal mine hidden danger based on LDA[J]. China Safety Science Journal, 2016, 26(6): 123-128.
- [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 S, LIU W Z, SHAO B, et al. Research on spatial and temporal distribution characteristics of safety hazards in hydropower projects[J]. China Safety Science Journal, 2022, 32(11): 90-96.
- [6] 林旭杰. 基于数据挖掘算法的煤矿安全隐患预警研究及应用分析[J]. 现代信息科技, 2022, 6(4): 141-144.  
LIN X J. Research and application analysis of coal mine hidden danger warning based on data mining algorithm[J]. Modern Information Technology, 2022, 6(4): 141-144.
- [7] LE T, DAVID J H. NLP-based approach to semantic classification of heterogeneous

- transportation asset data terminology[J]. *Journal of Computing in Civil Engineering*, 2017, 31(6): 04017057.
- [8] JATNIKA D, BIJAKSANA M A, SURYANI A A. Word2Vec model analysis for semantic similarities in English words[J]. *Procedia Computer Science*, 2019, 157: 160–167.
- [9] KUMAR H S, KRISHNAMOORTHY C S. A framework for case-based reasoning in engineering design[J]. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 1995, 9(3): 161–182.
- [10] 郑霞忠, 汪珂, 陈云, 等. 水电工程施工安全隐患文本智能类推研究[J]. *安全与环境学报*, doi: 10.13637/j.issn.1009-6094.2022.2138.
- ZHENG X Z, WANG K, CHEN Y, et al. Study on Intelligent Analogy of Safety Hazards in Construction of Hydropower Projects[J]. *Journal of Safety and Environment*, doi: 10.13637/j.issn.1009-6094.2022.2138.
- [11] 原江涛, 张瑞新, 赵红泽, 等. 煤矿案例推理隐患排查治理信息系统开发及应用[J]. *中国安全科学学报*, 2018, 28(8): 135–141.
- YUAN J T, ZHANG R X, ZHAO H Z, et al. Development and application of case-based reasoning based information system for identifying and controlling hidden danger in coal mine[J]. *China Safety Science Journal*, 2018, 28(8): 135–141.
- [12] 夏登友, 李丞曜, 朱毅, 等. 基于情景元的案例推理及其在应急决策中的应用[J]. *安全与环境学报*, 2020, 20(3): 1028–1033.
- XIA D Y, LI C Y, ZHU Y, et al. Case-based reasoning method based on the scenario element and its application in the emergency decision-making[J]. *Journal of Safety and Environment*, 2020, 20(3): 1028–1033.
- [13] LIU B, ZHANG L. A survey of opinion mining and sentiment analysis[M]//Mining text data. boston: springer, 2012: 415–463.
- [14] MA G F, WU M. A big data and FMEA-based construction quality risk evaluation model considering project schedule for Shanghai apartment projects[J]. *International Journal of Quality & Reliability Management*, 2020, 37(1): 18–33.
- [15] HAN J W, PEI J, YIN Y W, et al. Mining frequent patterns without candidate generation[J]. *ACM Sigmod Record*, 2000, 29(2): 1–12.
- [16] PAL A, GILLAM L. Set-based similarity measurement and ranking model to identify cases of journalistic text reuse[C]//Proceedings of FIRE' 13 Forum for Information Retrieval Evaluation. [S.l.: s.n.], 2013.
- [17] TAKABATAKE Y, NAKASHIMA K, KUBOYAMA T, et al. siEDM: an efficient string index and search algorithm for edit distance with moves[J]. *Algorithms*, 2016, 9(2): 26.
- [18] WAHYUNINGSIH T, HENDERI H, WINARNO W. Text mining an automatic short answer grading (ASAG), comparison of three methods of cosine similarity, Jaccard similarity and Dice's coefficient[J]. *JADS*, 2021, 2(2): 45–54.
- [19] LE Q V, MIKOLOV T. Distributed representations of sentences and documents[EB]. arXiv preprint, 2014, arXiv: 1405.4053.
- [20] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[EB]. arXiv preprint, 2013, arXiv: 1301.3781.
- [21] VIEHWEGER A, KRAUTWURST S, KÖNIG B, et al. 1 Distributed representations of protein domains and genomes 2 and their compositionality[J]. *bioRxiv*, 2019: 524280.
- [22] ABDELALIM A, O' BRIEN W. Visualization of building performance using sankey diagrams to enhance the decision-making process[C]//Proceedings of the Symposium on Simulation for Architecture and Urban Design. San Diego: Society for Computer Simulation International, 2017: 1–8.
- [23] 李丹一, 王武魁, 井晖, 等. 基于案例推理的森林抚育智能决策算法研究[J]. *西北林学院学报*, 2022, 37(1): 211–216.

LI D Y, WANG W K, JING H, et al. Intelligent decisional gorithm of forest tending based on case reasoning[J]. Journal of Northwest Forestry University, 2022, 37(1): 211-216.

[24] 盛姝, 黄奇, 郭进京, 等. 基于知识库与案例库的在线健康社区诊疗解决方案自动推理模

型研究[J]. 情报科学, 2022, 40(5): 161-172.

SHENG S, HUANG Q, GUO J J, et al. Research on automatic reasoning model of online health community diagnosis and treatment solution based on knowledge base and case base[J]. Information Science, 2022, 40(5): 161-172.

#### 作者简介



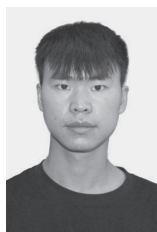
刘震(1996-),男,广东粤海珠三角供水有限公司工程师、信息管理副经理,主要研究方向为物联网、智慧工程、智慧水务、通信网络。



赵嵩(1998-),男,云南大学信息学院硕士生,主要研究方向为工业大数据建模、自然语言处理、目标检测与识别。



杨涛(1982-),男,深圳市科荣软件股份有限公司工程师、董事长,主要研究方向为水务信息化、人工智能。



蔡太伟(1999-),男,华南师范大学华南先进光电子研究院硕士生,研究方向为数据挖掘、深度学习、机器学习。

收稿日期: 2023-01-03

通信作者: 赵嵩, zsong0316@163.com

基金项目: 国家社会科学基金资助项目(No.21&ZD193)

Foundation Item: The National Social Science Foundation of China (No.21&ZD193)