

中文引用格式:牛茂辉,李威君,刘音,等. 基于24Model的动火作业事故致因文本挖掘[J]. 中国安全科学学报,2025,35(3):151-158.

英文引用格式:NIU Maohui, LI Weijun, LIU Yin, et al. Text mining of causes of hot working accidents based on 24Model[J]. China Safety Science Journal, 2025, 35(3): 151-158.

## 基于24Model的动火作业事故致因文本挖掘\*

牛茂辉<sup>1</sup>, 李威君<sup>\*\*1</sup>副教授, 刘音<sup>1</sup>教授, 王璐<sup>2</sup>

(1 山东科技大学 安全与环境工程学院, 山东 青岛 266590;

2 山东省港口集团有限公司, 山东 青岛 266000)

中图分类号: X928.02

文献标志码: A

DOI: 10.16265/j.cnki.issn1003-3033.2025.03.0757

基金项目: 国家自然科学基金青年基金资助(51904169); 山东省自然科学基金资助(ZR2023ME093)。

**【摘要】** 为探究工业动火作业事故的根源,提出一种基于“2-4”模型(24Model)的文本挖掘方法。首先,收集整理220篇动火作业事故报告,并作为数据集,构建基于来自变换器的双向编码器表征量(BERT)的24Model分类器,使用预训练模型训练和评估事故报告数据集,构建分类模型;然后,通过基于BERT的关键词提取算法(KeyBERT)和词频-逆文档频率(TF-IDF)算法的组合权重,结合24Model框架,建立动火作业事故文本关键词指标体系;最后,通过文本挖掘关键词之间的网络共现关系,分析得到事故致因之间的相互关联。结果显示,基于BERT的24Model分类器模型能够系统地判定动火作业事故致因类别,通过组合权重筛选得到4个层级关键词指标体系,其中安全管理体重的权重最大,结合共现网络分析得到动火作业事故的7项关键致因。

**【关键词】** “2-4”模型(24Model); 动火作业; 事故致因; 文本挖掘; 指标体系

### Text mining of causes of hot working accidents based on 24Model

NIU Maohui<sup>1</sup>, LI Weijun<sup>1</sup>, LIU Yin<sup>1</sup>, WANG Lu<sup>2</sup>

(1 College of Safety and Environmental Engineering, Shandong University of Science and Technology, Qingdao Shandong 266590, China; Shandong Port Group Co. Ltd., Qingdao Shandong 266000, China)

**Abstract:** In order to systematically explore the root causes of industrial hot work accidents through a large amount of text data, a text mining method based on 24Model was proposed. Firstly, 220 hot work accident reports were collected and sorted as datasets, and a 24Model classifier based on Bidirectional Encoder Representations from Transformers (BERT) was constructed. The pre-trained model was used to train and evaluate the accident report dataset to construct a classification model. Then, through the combination weight of the Keyword extraction algorithm based on BERT (KeyBERT) and Term Frequency-Inverse Document Frequency (TF-IDF) algorithms, combined with the 24Model framework, a keyword index system for hot work accident text was established. Finally, the interrelationships between accident causes were obtained through the analysis of the network co-occurrence relationship between text-mining keywords. The results show that the BERT-based 24Model classifier model can systematically and

\* 文章编号:1003-3033(2025)03-0151-08; 收稿日期:2024-10-19; 修稿日期:2024-12-21

\*\* 通信作者:李威君(1988—),女,山东烟台人,博士,副教授,主要从事油气生产过程的风险评估与预警、事故预防与风险控制理论、应急管理过程评价等方面的研究。E-mail: weijunli2013@163.com。

accurately determine the causative categories of hot work accidents. The weight of the safety management system was the largest among the 4-level keyword index systems obtained through the combination of weights. Furthermore, 7 key causative factors of hot work accidents were obtained by combining them with the co-occurrence network analysis. This shows that 24Model can strengthen the interpretability of text mining results, which provides an important reference for the prevention and management of hot work accidents.

**Keywords:** "2-4" model (24Model); hot work; accident causes; text mining; index system

## 0 引言

动火作业是工业领域极为常见的特殊作业,包括焊接、切割、电钻、喷灯等产生明火或火花的作业活动<sup>[1]</sup>。动火作业的危险性较高,尤其对于存在易燃、易爆物质的环境,极易引发火灾爆炸事故。近年来多起重特大事故都与动火作业有关,如2021年山东栖霞笏山金矿“1·10”重大爆炸事故等<sup>[2]</sup>、2024年江西新余“1·24”特大火灾事故<sup>[3]</sup>。在船舶修造火灾事故中,59.19%的点火源是动火作业引起的<sup>[4]</sup>。因此,研究动火作业事故致因,对于维护工业稳定、安全发展至关重要。

目前,关于动火作业事故防控的研究方法主要分为系统建模和数据挖掘2个方面。在系统建模方面,董秉聿等<sup>[5]</sup>运用事故分析图,从国家法律法规、政府政策与监管机构、技术与企业管理、事故进程与人员活动、设备与环境6个方面,分析了化工企业动火作业火灾事故案例的原因;LI Weijun等<sup>[6]</sup>利用改进蝶形图和Petri网开展动火作业风险演化分析和定量评估。此类研究虽然得到系统分析结果,但在分析过程中依赖领域专家经验,易产生主观和认知误差,且耗时费力、资源消耗大,而数据挖掘技术为解决上述问题提供了可能。其中,文本挖掘可通过特征提取、主题聚类、分类模型等技术分析事故报告,发现隐藏在事故文本中的信息<sup>[7]</sup>。例如:XU Hui等<sup>[8]</sup>利用文本挖掘技术分析处理了我国化工领域的267个动火作业事故,得到导致事故的关键原因有:未进行气体检测和持续监控、可燃物未清理、防护措施不当、工人违反规定。然而,现有的动火作业数据挖掘主要局限于技术运用,未与事故致因分析等领域知识有效融合。

鉴于此,笔者拟选用实用性较强的“2-4”模型(24Model)对事故原因初步分类,基于24Model预训练文本模型开展事故致因分类训练,分析事故致因的潜在联系和相互作用,识别关键致因,以期为工业动火作业事故的预防和控制提供理论依据。

## 1 基于24Model的文本挖掘方法

文本挖掘是收集、处理、分析、可视化文本数据的过程。传统的文本挖掘主要以词频作为分类依据,缺乏系统的理论指导,导致结果的可解释性较差。针对该问题,文中提出利用事故致因模型引导事故文本挖掘的方法。目前,国内外代表性的事故致因模型有近30种<sup>[9]</sup>;考虑到事故文本挖掘过程的特点,需要选取一种对事故原因分类和界定较为明确的模型。24Model<sup>[10]</sup>能够将事故致因划分为2个层面和4个关键过程,系统、多层次地分析事故致因。故按照24Model将事故原因分为安全文化、安全管理体系、安全知识、安全意识、安全习惯、安全心理、安全生理、不安全动作和不安全状态,驱动构建文本分类模型,建立指标体系,并开展关联网络分析,具体流程如图1所示。

### 1.1 数据收集和预处理

动火作业事故报告记录了事故发生的过程、原因和预防措施等细节,蕴含大量的事故信息,故选用动火作业事故报告作为文本挖掘的数据集。

### 1.2 构建基于24Model的文本分类器

基于24Model的文本挖掘方法以模型中的9项要素作为分类标签,解析事故致因,表1为每个分类的解释词。

为减少噪声文本对模型训练的影响,加入噪声文本标签,得到用于训练分类模型的标签化数据集。将数据集划分为训练集(80%)和测试集(20%),训练集用于模型学习特征,测试集用于验证模型的性能。

基于来自变换器的双向编码器表征量(Bidirectional Encoder Representations from Transformers, BERT)的预训练模型通过无监督学习方式,能够全面地理解语义和语法信息,在11项自然语言处理任务中表现出更卓越的性能<sup>[11]</sup>。相较于从零开始构建模型,基于BERT构建文本分类

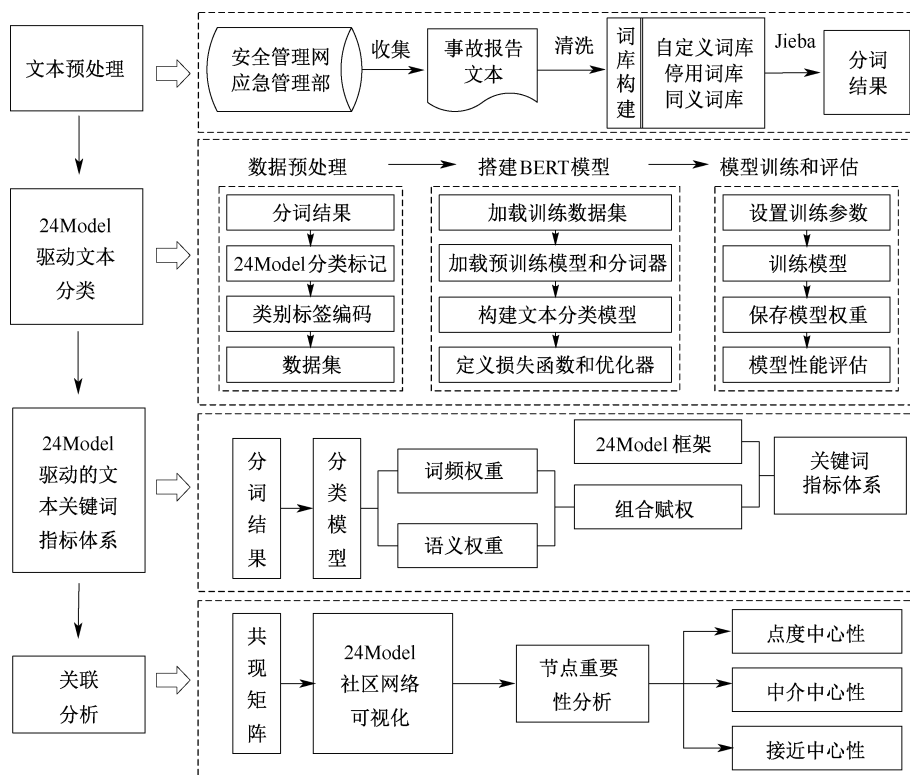


图 1 基于 24Model 的文本挖掘流程

Fig. 1 24Model-based text mining process

表 1 24Model 解释词

Table 1 Explanation of terms in the 24 Model

层级要素	解释词
安全文化	文化建设、思想、责任、价值观、投入、态度、理念、方针、指导思想
安全管理体系	许可证、培训、监管、规程、风险管控、隐患排查、管理体系、标准、资质、规章、制度、岗位责任
安全知识	教育、技能、知识、能力、资格、经验
安全意识	不重视、疏忽、逃避、动机
安全习惯	多次违章、多次违反
安全心理	冒险、压力、紧张、情绪
安全生理	疲劳、迟钝、年龄、性别、体重
不安全动作	未正确使用、违法采用、违规操作、违反操作规程、整改不力、违章作业、违章指挥
不安全状态	条件、环境、防护用品、工具、设备、附件、安全防护措施、标志

器<sup>[12]</sup>,所消耗的训练时间和标注数据更少<sup>[13]</sup>。因此,24Model 分类器的搭建采用 BERT 预训练中文语言模型,通过 BERT 分词器将输入的数据集划分成标记序列,生成“谷歌张量流”数据集。

准确率  $A$ 、精确率  $P$ 、召回率  $R$  和  $F_1$  分数是评估模型性能的主要指标<sup>[14]</sup>:

$$A = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$P = TP / (TP + FP) \quad (2)$$

$$R = TP / (TP + FN) \quad (3)$$

$$F_1 = 2PR / (P + R) \quad (4)$$

式中:TP (True Positive) 为预测和实际结果均为正例的样本数;TN (True Negative) 为预测和实际均为反例的样本数;FP (False Positive) 为预测是正例、实际是反例的样本数;FN (False Negative) 为预测是反例、实际是正例的样本数。

### 1.3 24Model 指标体系及词频-语义组合权重

不同于传统文本挖掘中由单一权重确定致因指标的方式,基于 24Model 的文本挖掘通过模型确定事故致因框架,并根据权重以指标填充的方式建立事故致因指标体系。为综合考量关键词的词频和语义相关性,采用基于 BERT 的关键字提取算法 (KeyBERT) 和词频-逆文档频率 (Term Frequency-Inverse Document Frequency, TF-IDF) 算法的组合权重。通过比较词汇间的相似性分析出最具有代表性和语义相关性的关键词,相似度  $W_k$  的计算过程为:

$$W_k = (S \times K) / (|S| \times |K|) =$$

$$\sum_{i=1}^n w_{si} \times w_{ki} / (\sqrt{\sum_{i=1}^n (w_{si})^2} \times \sqrt{\sum_{i=1}^n (w_{ki})^2}) \quad (5)$$

式中: $S$ 和 $K$ 为句子 $S$ 和关键词 $K$ 各自 $n$ 个分量的向量; $w_{si}$ 和 $w_{ki}$ 分别 $S$ 为和 $K$ 的第 $i$ 个分量。

TF-IDF 算法通过分析一个文档中词语出现的频率确定该词语在文档中的重要程度<sup>[14]</sup>, 权重  $W_T$  表示为:

$$W_T = w_T \times w_I = (C_w/N) \times \ln(d/w_D) \quad (6)$$

式中: $w_T$ 为特征词 $w$ 在文本中出现的次数; $w_I$ 为 $w$ 的逆文档频率; $C_w$ 为 $w$ 在文档中出现次数; $N$ 为文档中总词语个数; $d$ 为文档总数; $w_D$ 为特征词所在文档数。

考虑词频和语义相关性,组合权重  $W$  表示为:

$$W = (W_K + W_T)/2 \quad (7)$$

### 1.4 事故致因关联分析

共现网络分析有助于揭示潜在致因因素、阐明各因素之间的联系<sup>[15]</sup>,通过计算致因因素的共现频率,得出致因因素的共现矩阵;将数据导入 Gephi 可视化工具<sup>[16]</sup>,构建事故致因网络图。在网络图中,每个事故致因被视为一个节点,节点大小反映共现次数;节点之间的边表示其共现关系,边的粗细可反映共现关系的强度。在网络分析中,节点的重要性由3个评价指标来衡量:点度中心性、中介中心性和接近中心性<sup>[17]</sup>。

度中心性是指与某节点相连的边的数量。度中心性越大表明该节点在网络中的影响力更大。节点的点度中心性  $C_D(v)$  计算过程为:

$$C_D(v) = \sum_i^n a_{vi}/(n-1) \quad (8)$$

式中: $a_{vi}$ 为连接到节点 $v$ 和节点 $i$ 关系,相连记为1,不相连记为0; $n$ 为网络中的总节点数。

接近中心性衡量节点与其他所有节点的平均最短路径长度,即节点间的距离。数值大表明因素间联系紧密,是事故发生的直接原因或重要环节。节点 $v$ 的接近中心性  $C_C(v)$  的计算过程为:

$$C_C(v) = (n-1)/\sum_{i=1}^n b_{vi} \quad (9)$$

式中 $b_{vi}$ 为节点 $v$ 到节点 $i$ 之间的最短路径长度。

中介中心性用于衡量节点在网络中的控制能力和信息传播能力。数值高表明该节点是连接其他节点的关键节点,对事故发生和发展有重要影响。节点的中介中心性  $C_B(v)$  表示为:

$$C_B(v) = \sum_{s \neq v \neq t} (\sigma_{st}(v)/\sigma_{st}) \quad (10)$$

式中: $\sigma_{st}$ 为从节点 $s$ 到节点 $t$ 的所有最短路径数量; $\sigma_{st}(v)$ 为从节点 $s$ 到节点 $t$ 的所有最短路径数量。

## 2 基于 24Model 的动火作业事故致因文本挖掘

### 2.1 数据收集和处理

事故数据源自截至2024年1月的安全管理网(<https://www.safehoo.com/Case/>)和中华人民共和国应急管理部(<https://www.mem.gov.cn/>)公布的476篇动火作业事故报告,筛选去除相同案例,整理得到220篇事故报告文本,作为文本挖掘数据集。

为防止错误切分表示事故致因的短语,在使用 Python 的 Jieba 工具分词时,先汇总安全生产和化工领域的术语,再添加动火作业致因词,建立自定义词库。同时整合“哈工大停用词表”和“百度停用词表”,并加入“企业”、“车间”等与事故致因分析无关的词语,构建用于动火作业事故报告文本分析的停用词库。另外,整合事故报告中的同义表达,建立同义词库。

### 2.2 分类模型的训练和评估

设定关键参数:迭代次数、批次大小和最大序列长度。记录每次迭代的损失函数值和  $A$ , 绘制训练曲线,如图2所示。观察到在经过10次迭代后, $A$ 曲线和损失值函数值曲线趋于稳定;准确率曲线未出现下降趋势,损失值曲线也未显示上升趋势,表明未出现过拟合或欠拟合现象。故将最佳迭代次数设置为10。由于数据集相对较小,选择批次大小为16。考虑到分类对象为词或词组,将最大序列长度设定为16。

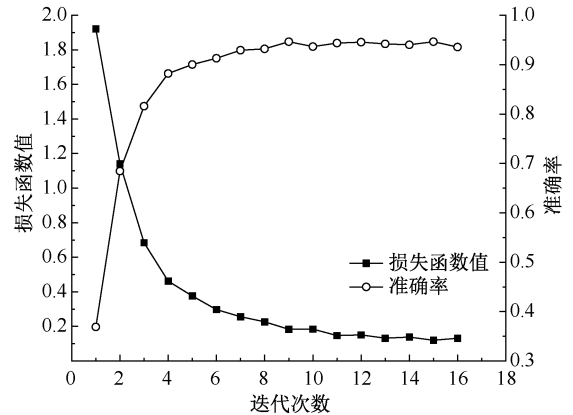


图2 分类模型训练曲线

Fig.2 Training curve of classification model

将测试数据投入训练好的分类模型,得到分类预测结果。由式(1)一式(4)计算得到24Model要

素分类的性能参数  $A$ 、 $P$ 、 $R$  和  $F_1$  分数,见表2。可以看出,基于24Model的文本分类在各类别上展现出相对均衡的  $P$ 、 $R$  和  $F_1$  分数,  $A$  为 82%,说明该方法能够较为准确地辨识事故致因的类别。

表2 模型评估参数

Table 2 Model evaluation parameters

层级要素	$P$	$R$	$F_1$
安全文化	0.750 0	0.750 0	0.750 0
安全管理体系	0.918 4	0.882 4	0.900 0
安全知识	1.000 0	0.800 0	0.888 9
安全意识	0.750 0	0.857 1	0.800 0
安全习惯	1.000 0	0.500 0	0.667 0
安全心理	0.833 0	1.000 0	0.909 1
安全生理	1.000 0	1.000 0	1.000 0
不安全动作	0.666 7	0.909 1	0.769 2
不安全状态	0.727 3	0.761 9	0.744 2
噪声文本	0.761 9	0.695 7	0.727 3
$A$	0.820 1		

### 2.3 指标体系构建及权重计算

由式(7)计算关键词的  $W$  部分,结果见表3。使用经过训练的事故致因分类模型,对关键词提取

结果进行分类和二次过滤,根据权重值筛选出每个分类中的致因关键词,得到基于24Model的文本挖掘关键词指标体系,如图3所示。

表3 关键词权重赋值结果(部分)

Table 3 Partial results of keyword weight assignment

关键词	$W_k$	$W_T$	$W$
安全防护措施	0.074 9	0.028 1	0.051 5
安全管理	0.05 94	0.026 5	0.042 9
安全制度不完善	0.054 6	0.022 5	0.038 6
教育和培训	0.039 9	0.028 3	0.034 1
安全意识淡薄	0.051 0	0.016 3	0.033 6
动火作业审批	0.041 7	0.013 3	0.027 5
动火作业许可证	0.038 2	0.016 1	0.027 2
安全生产主体责任	0.036 0	0.012 1	0.024 1
现场监护不力	0.027 3	0.020 0	0.023 7
违规动火作业	0.032 2	0.011 4	0.021 8

对词频-语义权重值归一化处理得到动火作业事故致因各层级权重,见表4。可以看出,安全管理体系的权重最大,是最重要的层级要素;不安全状态和不安全动作在历年动火作业事故中的致因占比也相对较大。



图3 基于24Model的文本挖掘关键词指标体系

Fig.3 24Model-based text mining keyword indicator system

### 2.4 工业动火作业事故致因关联分析及可视化

表5给出了体现动火作业文本挖掘关键词间共现关系的共现矩阵。将数据导入 Gephi 可视化工具,按照24Model的框架将动火作业事故的致因要素划分为组织文化、管理体系、个体能力和个体行为4个层级指标作为网络社区单元,得到共现网络,如图4所示。

表4 动火作业事故致因权重

Table 4 Weight of causes of hot work accidents

层级要素	分层权重	关键词	归一化权重	层级要素	分层权重	关键词	归一化权重
A	0.075 3	A <sub>1</sub>	0.070 0	E	0.0142	D <sub>5</sub>	0.002 4
		A <sub>2</sub>	0.002 0			E <sub>1</sub>	0.008 4
		A <sub>3</sub>	0.002 3			E <sub>2</sub>	0.002 8
		A <sub>4</sub>	0.001 0			E <sub>3</sub>	0.003 0

续表 4

层级要素	分层权重	关键词	归一化权重	层级要素	分层权重	关键词	归一化权重		
B	0.480 0	B <sub>1</sub>	0.088 2	F	0.001 8	F <sub>1</sub>	0.001 8		
		B <sub>2</sub>	0.048 6	G	0.005 8	G <sub>1</sub>	0.005 8		
		B <sub>3</sub>	0.055 9	H	0.128 2	H <sub>1</sub>	0.044 8		
		B <sub>4</sub>	0.056 6			H <sub>2</sub>	0.033 1		
		B <sub>5</sub>	0.049 5			H <sub>3</sub>	0.010 6		
		B <sub>6</sub>	0.040 5			H <sub>4</sub>	0.018 6		
		B <sub>7</sub>	0.079 3			H <sub>5</sub>	0.008 9		
		B <sub>8</sub>	0.039 3			H <sub>6</sub>	0.012 2		
		B <sub>9</sub>	0.022 1			I <sub>1</sub>	0.105 9		
C	0.017 6	C <sub>1</sub>	0.010 2			I	0.186 0	I <sub>2</sub>	0.024 0
		C <sub>2</sub>	0.003 2					I <sub>3</sub>	0.018 5
		C <sub>3</sub>	0.002 1	I <sub>4</sub>	0.014 4				
		C <sub>4</sub>	0.002 1	I <sub>5</sub>	0.010 2				
		D <sub>1</sub>	0.069 2	I <sub>6</sub>	0.007 9				
D	0.091 1	D <sub>2</sub>	0.007 3	I <sub>7</sub>	0.003 6				
		D <sub>3</sub>	0.006 2	I <sub>8</sub>	0.001 5				
		D <sub>4</sub>	0.006 0						

表 5 共现矩阵(部分)

Table 5 Co-occurrence matrix (partial)

关键词	B <sub>8</sub>	A <sub>1</sub>	I <sub>1</sub>	B <sub>1</sub>	B <sub>7</sub>	B <sub>9</sub>	B <sub>6</sub>	B <sub>2</sub>
B <sub>8</sub>	0	67	69	72	72	62	45	45
A <sub>1</sub>	7	0	0	4	2	9	46	3
I <sub>1</sub>	12	65	0	16	74	56	43	21
B <sub>1</sub>	7	66	58	0	75	51	47	4
B <sub>7</sub>	7	75	0	4	0	50	47	3
B <sub>9</sub>	0	52	0	2	0	0	39	0
B <sub>6</sub>	2	5	1	4	4	2	0	2
B <sub>2</sub>	1	32	39	41	39	23	23	0

施)、B<sub>1</sub>(安全管理)、B<sub>7</sub>(安全制度不完善)、B<sub>9</sub>(隐患排查)及 D<sub>1</sub>(安全意识淡薄)7 个事故致因具有较高的接近中心性和中介中心性,说明这些因素在事故发生的过程中起着重要作用,是导致事故的关键因素。

### 3 与传统文本挖掘的比较分析

传统文本挖掘方法是在没有事故致因模型驱动的情况下依赖于自然语言处理技术和统计学方法挖掘文本中的有用信息<sup>[18]</sup>。而基于 24Model 的文本

由式(8)一式(10)计算节点的 C<sub>D</sub>、C<sub>C</sub> 和 C<sub>B</sub>, 表 6 为事故致因节点重要性定量分析结果。可以看出, B<sub>8</sub>(监督检查)、A<sub>1</sub>(教育培训)、I<sub>1</sub>(安全防护措

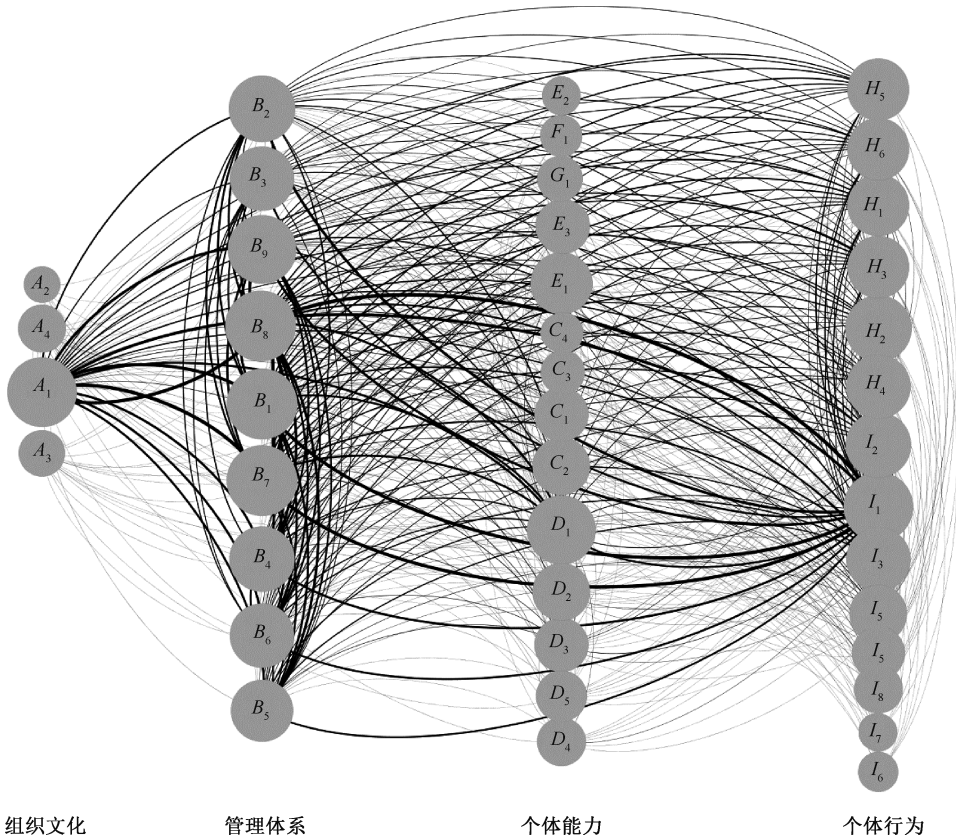


图 4 24Model 驱动文本挖掘关键词共现网络

Fig. 4 24Model-driven text mining keyword co-occurrence network

表6 事故致因网络中心性

Table 6 Centrality in accident causation network

关键词	$C_D$	$C_C$	$C_B$	关键词	$C_D$	$C_C$	$C_B$
$B_8$	40	1.00	22.25	$H_2$	36	0.91	9.91
$A_1$	39	0.98	18.20	$H_3$	32	0.83	3.99
$I_1$	39	0.98	18.73	$H_6$	32	0.83	5.00
$B_1$	40	1.00	22.25	$H_5$	32	0.83	5.11
$B_7$	40	1.00	22.25	$H_4$	34	0.87	6.54
$B_9$	38	0.95	18.03	$I_5$	28	0.77	2.44
$B_6$	34	0.87	8.15	$E_1$	32	0.83	4.62
$B_2$	36	0.91	12.54	$D_3$	25	0.73	0.61
$B_4$	35	0.89	9.06	$I_4$	24	0.71	1.21
$B_3$	34	0.87	8.01	$C_2$	28	0.77	1.84
$B_5$	32	0.83	3.62	$D_4$	20	0.67	1.31
$I_2$	35	0.89	11.46	$C_1$	25	0.73	0.74
$D_1$	38	0.95	16.99	$D_5$	22	0.69	0.28
$I_3$	34	0.87	9.87	$E_3$	25	0.73	0.96
$H_1$	32	0.83	6.96	$D_2$	28	0.77	2.18

挖掘方法与传统文本挖掘在目标、技术应用、结果呈现方面均存在明显差异。

1) 目标差异:传统文本挖掘侧重于从文本自身中提取信息、分类文档或完成聚类任务。模型驱动的文本挖掘旨在从文本中更加系统化地识别事故原因,分层次地分析事故原因。

2) 技术应用:传统文本挖掘方法通常使用词频统计、TF-IDF 等技术。模型驱动的文本挖掘结合自然语言处理技术、机器学习算法和专门针对事故分析设计的模型来进行文本分析。选用 24Model 设计分类搭建预训练模型,增强了结果的可解释性,降低分类模型训练所需数据量。为文本挖掘结果维度高,降维算法可能造成关键信息缺失的问题提供解决思路。

3) 结果呈现:传统文本挖掘直接给出关键词,且对关键词分类时,事故原因之间存在重叠的情况,导致结果缺乏层次分明的解释,见表 7。基于 24Model 的文本挖掘方法得到的结果更加注重对事故原因的解释和可视化,帮助决策者理解事故发生的机制。

表7 传统文本挖掘方法结果

Table 7 Results of traditional text mining methods

影响因素分类	事故初始因素	影响因素分类	事故初始因素	影响因素分类	事故初始因素
安全教育与培训	未组织安全教育培训	安全管理制度	管理制度不健全	安全监管	未进行技术交底
	未掌握安全技能		动火审批制度未落实		特种作业资格证过期
	安全知识储备不足		安全生产主体责任制度不完善		安全监管不到位
违章违规行为	员工不重视安全教育培训		未办理动火票		特种作业人员无证上岗
	违章作业	通风设备故障	监督检查不严格		
	冒险作业	电源线老化损坏	安全意识	存在侥幸心理	
	违反操作规程	未设置安全警示标志		监护人疏忽大意	
未安排现场指挥人员	消防设施不完善	缺乏自我保护意识			
安全防护	违章指挥	隐患排查	事故隐患排查不彻底	应急处置	应急救援不及时
	安全防护措施不完善		未组织危险源识别		未制定应急处置预案
	未正确使用安全防护措施		未进行风险评估		未组织应急演练
	未正确佩戴劳动防护用品		存在易燃易爆物	人员配备	技术人员缺失
非法储存危险化学品			人员配备不合理		

## 4 结论

1) 通过将 24Model 与文本挖掘相结合,构建基于 BERT 的 24Model 分类器,将安全文化、安全管理体系、安全知识、安全意识、安全习惯、安全心理、安全生理、不安全动作、不安全状态 9 项要素作为分类标签,以准确判定事故致因类别,增强文本挖掘过程

的系统性。

2) 采用 KeyBERT 和 TF-IDF 组合赋权构建以 24Model 为框架的动火作业事故文本关键词指标体系,为确定关键事故致因层级及因素提供框架,结合基于 24Model 的文本挖掘关键词共现网络得到 7 项关键致因,为动火作业事故预防和管理提供更确切的定量信息。

## 参考文献

[1] WANG Xin. Analysis and improvement of hot work management in China[J]. Process Safety Progress, 2023, 42(1): 72-78.

- [2] 山东省人民政府. 山东五彩龙投资有限公司栖霞市笏山金矿“1·10”重大爆炸事故调查报告[EB/OL]. (2021-02-23). [http://www.shandong.gov.cn/art/2021/2/23/art\\_305296\\_10334792.html](http://www.shandong.gov.cn/art/2021/2/23/art_305296_10334792.html).
- [3] 中华人民共和国应急管理部. 佳乐苑小区特别重大火灾事故调查报告[EB/OL]. (2024-09-21). [https://www.mem.gov.cn/xw/bndt/202409/t20240921\\_501982.shtml](https://www.mem.gov.cn/xw/bndt/202409/t20240921_501982.shtml).
- [4] 李显, 焦宇, 陈文涛, 等. 船舶修造企业火灾事故特征挖掘与致因分析[J]. 中国安全科学学报, 2024, 34(3): 29-38.  
LI Xian, JIAO Yu, CHEN Wentao, et al. Feature mining and causative analysis of fire accidents in ship repair and building enterprises[J]. China Safety Science Journal, 2024, 34(3): 29-38.
- [5] 董秉聿, 王宇昊, 孙鹏依, 等. 基于 AcciMap 模型的动火作业火灾事故案例分析[J]. 安全, 2023, 44(6): 94-99.  
DONG Bingyu, WANG Yuhao, SUN Pengyi, et al. Case analysis on fire accident caused by fire operation based on AcciMap model[J]. Safety and Security, 2023, 44(6): 94-99.
- [6] LI Weijun, SUN Qiqi, ZHANG Jiawang, et al. Quantitative risk assessment of industrial hot work using adaptive bow tie and Petri nets[J]. Reliability Engineering and System Safety, 2024, 242; DOI: 10.1016/j.res.2023.109784.
- [7] 田水承, 黄权. 基于文本挖掘的煤矿瓦斯事故致因分析[J]. 煤矿安全, 2022, 53(5): 241-245.  
TIAN Shuicheng, HUANG Quan. Analysis of causes for coal mine gas accidents based on text mining method[J]. Safety in Coal Mines, 2022, 53(5): 241-245.
- [8] XU Hui, LIU Yi, SHU Chimin, et al. Cause analysis of hot work accidents based on text mining and deep learning[J]. Journal of Loss Prevention in the Process Industries, 2022, 76; DOI: 10.1016/j.jlp.2022.104747.
- [9] FU Gui, XIE Xuecai, JIA Qingsong, et al. The development history of accident causation models in the past 100 years: 24Model, a more modern accident causation model[J]. Process Safety and Environmental Protection, 2020, 134: 47-82.
- [10] 傅贵, 陈奕燃, 许素睿, 等. 事故致因“2-4”模型的内涵解析及第6版的研究[J]. 中国安全科学学报, 2022, 32(1): 12-19.  
FU Gui, CHEN Yiran, XU Surui, et al. Detailed explanations of 24Model and development of its 6<sup>th</sup> version[J]. China Safety Science Journal, 2022, 32(1): 12-19.
- [11] DEVLIN J, CHANG Mingwei, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019: 4171-4186.
- [12] GARRIDO-MERCHAN E C, GOZALO-BRIZUELA R, GONZALEZ-CARVAJAL S. Comparing BERT against traditional machine learning models in text classification[J]. Journal of Computational and Cognitive Engineering, 2023, 2(4): 352-356.
- [13] 李思锐. 基于 BERT 的子词级中文文本分类方法[J]. 计算机科学与应用, 2020, 10(6): 1075-1086.  
LI Sirui. Subword-level Chinese text classification method based on BERT[J]. Computer Science and Application, 2020, 10(6): 1075-1086.
- [14] 兰晓芳, 刘卓, 许志豪, 等. 基于 TF-IDF 和 TextRank 结合的中文文本关键词提取方法: 以体育新闻为例[J]. 软件工程, 2023, 26(8): 6-10.  
LAN Xiaofang, LIU Zhuo, XU Zhihao, et al. A Chinese text keyword extraction method based on the combination of TF-IDF and TextRank; a case study of sports news[J]. Software Engineering, 2023, 26(8): 6-10.
- [15] 郑彬彬, 冯婷婷, 王佳贺, 等. 基于文本挖掘的城镇燃气事故致因及关联分析[J]. 中国安全科学学报, 2023, 33(7): 190-195.  
ZHENG Binbin, FENG Tingting, WANG Jiahe, et al. Causes and correlation analysis of urban gas accidents based on text mining[J]. China Safety Science Journal, 2023, 33(7): 190-195.
- [16] BASTIAN M, HEYMANN S, JACOMY M. Gephi: an open source software for exploring and manipulating networks[C]. International AAAI Conference on Weblogs and Social Media, 2009: 361-362.
- [17] LIU Muyang, LUO Xiaowei, WANG Guangbin, et al. Intelligent information extraction from government on-site inspection reports of construction projects: a graph-based text mining approach[J]. Advanced Engineering Informatics, 2023, 58; DOI: 10.1016/j.aei.2023.102163.
- [18] 王亮, 赖佳燕, 张自欣, 等. 基于文本挖掘和改进 DEMATEL 法的化工事故关键因素识别[J]. 中国安全科学学报, 2024, 34(3): 20-28.  
WANG Liang, LAI Jiayan, ZHANG Zixin, et al. Identification of critical factors in chemical accidents based on text mining and improved DEMATEL method[J]. China Safety Science Journal, 2024, 34(3): 20-28.

**作者简介:** 牛茂辉 (1999—), 男, 河南濮阳人, 硕士研究生, 研究方向为油气生产过程的风险评估与预警、事故预防与风险控制理论。E-mail: nmaohui@163.com。

