



基于 LDA 模型的建筑安全事故原因分析

贾宜然¹, 庞玉成²

(1. 烟台大学 土木工程学院, 山东 烟台 264005; 2. 济宁医学院, 山东 济宁 272067)

摘要: 建筑行业推动着国民经济的发展, 是我国经济支柱行业。但建筑行业在高速发展的同时, 安全事故发生频次也较高, 给行业带来负面影响。文章应用隐含狄利克雷分配模型(latent dirichlet allocation, LDA), 通过事故案例文本资料的分析, 确定了 6 个主要主题, 包括施工器材安全管理、遵守法规、组织管理、特种作业管理、重型设备管理以及高处作业的安全保护措施。研究表明: 通过构建施工现场机械材料全周期安全管控体系, 健全安全生产法规与标准体系框架, 创新组织架构管理模式, 建立特种作业人员专项认证与动态监管机制, 完善重型设备拆装作业标准化流程及风险评估矩阵, 构建高处作业坠落防护系统与三维监测网络, 可系统提升建设工程本质安全水平, 为工程安全管理系统化建设提供参考。

关键词: 安全生产; LDA; 事故原因

中图分类号: TU714

文献标志码: A

文章编号: 1673-8993(2025)05-0072-07

doi: 10.13402/j.gcjs.2025.05.064

Cause analysis of building safety accidents based on Latent Dirichlet Allocation

JIA Yiran¹, PANG Yucheng²

(1. College of Civil Engineering, Yantai University, Yantai 264005, Shandong, China;

2. Jining Medical University, Jining 272067, Shandong, China)

Abstract: The construction industry drives the development of the national economy and is a pillar industry of China's economy. However, while the construction industry is developing rapidly, the frequency of safety accidents is also relatively high, bringing negative impacts to the industry. In this paper, the latent dirichlet allocation(LDA) is applied to analyze the text data of accident cases, and six main topics are identified, including construction equipment safety management, compliance with regulations, organizational management, special operation management, heavy equipment management, and safety protection measures for high-altitude operations. The research shows that by constructing a comprehensive safety control system for the entire life cycle of machinery and materials at construction sites, improving the framework of safety production regulations and standards, innovating organizational structure management models, establishing a special operation personnel certification and dynamic supervision mechanism, perfecting the standardized process and risk assessment matrix for heavy equipment installation and dismantling operations, and building a fall protection system and 3D monitoring network for high-altitude operations, the inherent safety level of construction projects can be systematically enhanced, providing a reference for the systematic construction of engineering safety management.

Key words: safety production; LDA; accident cause

收稿日期: 2024-8-15

作者简介: 贾宜然(1999—), 女, 硕士研究生, 从事建设项目安全方面的研究。

通信作者: 庞玉成(1977—), 男, 正高级工程师, 从事医院建设项目管理方面的研究。

建筑业是我国经济支柱行业, 其提供就业人数占全国就业总人数的 12.6%。建筑行业推动着国民经济的发展, 在国家经济中占很大的比重。通过查找《国家统计年鉴》相关数据, 在 2016 年至 2020 年间, 我国建筑业的国内生产总值逐年增加, 在 2020 年达到 72 995.7 亿元, 占国内生产总值的比例在 7% 左右。但也应注意到, 建筑行业在高速发展的同时, 建设过程中的工程安全事故也时有发生, 造成了严重的人员伤亡, 给行业带来较大的负面影响^[1]。因此, 研究建筑安全事故原因, 并提出有效的预防策略, 对促进工程建设行业高质量发展具有重要的现实意义。

安全生产是指通过一系列措施和活动, 旨在保护生产过程中的人员、设备和环境免受伤害、损害或破坏的状态。“十四五”规划提出, 要完善和落实安全生产责任制, 建立公共安全隐患排查和安全预防控制体系。每起建筑安全事故都会对一个家庭造成巨大打击, 对建筑项目造成重大损失, 建筑行业必须始终把安全生产放在首位, 切实落实各项安全措施, 为行业的持续健康发展提供坚实的保障。只有确保了安全生产, 我们才能构建更加安全、稳定、和谐的环境, 为国家的建设事业贡献力量。

鉴于当前安全事故频发的严峻形势, 国内外众多学者积极深入开展各类研究, 力求降低安全生产事故的发生率, 保障人民生命财产安全。丁明智等^[2]基于“超效率 SB-分层评-GML”模型构建建筑业宏观安全生产水平研究框架; 齐锡晶等^[3]基于 EPC 模式下设计、采购、施工之间的关联性, 运用系统动力学理论, 建立 EPC 模式的安全生产管理系统; 仝立杨等^[4]采用耦合动态聚类分析与双层前沿数据包络分析(DEA)的综合方法, 深入探索并科学合理地设定省域建筑业的安全生产目标; 郑霞忠等^[5]对于建筑业安全生产水平时空演变及影响因素进行分析; 胡铮^[6]基于模糊 DEMATEL 方法对建筑施工安全影响因子进行分析; 虞芷绮等^[7]对安全管理成熟度进行研究分析; OSWALD 等^[8]以英国某建筑项目为案例, 研究建筑工程项目生产压力对于安全的影响。LI 等^[9]提出 4D 交互式安全评估方法, 通过虚拟场景训练提

升建筑工人的安全知识和技能。现有研究多聚焦于静态风险分析或单一因素评估, 难以系统揭示事故成因的多维度动态耦合机制, 基于此, 本文引入 LDA 主题模型, 对事故报告进行文本挖掘, 通过量化主题共现规律揭示风险传导路径, 为建筑安全管理提供策略。

1 理论基础

潜在狄利克雷分配模型 (latent dirichlet allocation, LDA) 是一种用于发现文档集中潜在主题统计模型。其原理基于贝叶斯理论, 通过概率分布来推断文档中隐藏的主题结构。LDA 模型假设每个文档由多个主题混合而成, 而每个主题又是由一组特定的词汇以一定概率分布组成的。模型的核心在于学习这些潜在的主题以及它们与文档和词汇之间的关联。文档的主题分布及其主题内词语的分布通过吉布斯 (Gibbs) 采样等方法经多次迭代后收敛得出^[10]。LDA 主题分析的原理如图 1 所示。模型训练采用的是 Gibbs 采样的方法, 由两个参数 α 和 β 分别控制一个狄利克雷分布 (Dirichlet), α 随机生成文档对应主题的多项式分布 θ , θ 随机生成一个主题 Z , β 随机生成主题对应词语的多项式分布 φ , 综合主题 Z 和主题对应词语分布情况 φ 生成词语 W , 如此循环生成一个包含 M 个词语的文档, 最终生成 K 个主题下的 N 篇文档, 对比随机生成的 N 篇文档和原文档的相似性来判断模型的好坏, 最终在不同的 α 和 β 参数生成的模型中, 找到最优的模型, 即最佳的 α 和 β 参数的值。

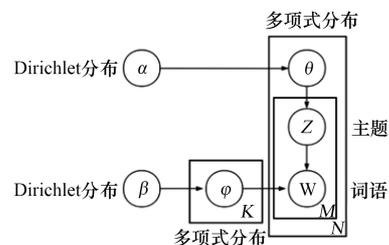


图 1 LDA 主题模型原理

国内外学者基于 LDA 主题分析模型进行了大量的研究, 关菁华等^[11]基于深度模型的命名实体识别方法与 LDA 模型的主题挖掘方法相结合, 进行专利技术主题演化分析; 宋军等^[12]基于 LDA 主题模型从上市公司风险文本中提取出 15 个主题风

险, 研究其披露程度对市场风险的影响; 杨益兴等^[13]基于 LDA 主题模型, 以京东为例, 研究多阶段的生鲜消费者需求; LI 等^[14]运用 LDA 主题模型, 探究建筑业数字化转型的关键影响因素。

2 研究分析

2.1 数据来源

事故报告作为记录和分析事故经过、原因、后果及预防措施的重要文献, 由具有丰富经验的专家团队编制, 内容详尽且贴近实际, 蕴含着丰富的安全生产信息和经验教训, 是研究安全风险因素的重要数据来源。LDA 主题分析模型能深入事故调查报告的文本数据, 自动挖掘潜在主题, 揭示事故发生的直接及间接原因, 高效利用专家团队的详尽报告, 全面且准确地识别建筑工程安全风险因素, 提炼出有效的安全管理策略, 以预防未来类似事故的发生。本研究的安全生产事故调查报告基于各省市及区级应急管理部门和住房和城乡建设部门官方网站上依法主动公开的信息, 建筑类型为房屋建筑, 本研究经过严格筛选, 剔除了信息不全或仅含事故简报的不完整数据, 最终筛选出格式一致、内容详尽的有效事故案例。经过数据清洗, 共整理出 132 份文本资料。

2.2 数据处理

由于 LDA 主题分析模型针对词语进行分析, 因此在进行 LDA 主题分析前, 需要将事故报告处理成分词, 即将文本分成一个个常用的词语, 以便后续进行分析处理。现在对于中文句子进行分词的技术已经非常成熟, 目前主要的分词软件有 THULAC4、LTP3、ICTCLAS2、jieba 分词等, 考虑到与 LDA 主题分析模型的适用性, 本研究采用 Python 的 jieba 分词对房屋建筑安全生产事故的文本进行分词。

2.3 数据分析

在进行模型分析前, 需要对参数 α 和 β 及主题数量 T 进行设置参数, 参数 α 和 β 的数值取常用数值, $\alpha = 0.1$, $\beta = 0.01$ 。除了 α 和 β 的参数值外, 研究的难点在于确定主题数量 T , 通过研究计算得出最优的主题数量 T 。在 LDA 主题分析模型中, 主题数量的确定常用主题困惑度

(Perplexity) 的方法来确定, 困惑度计算公式一般用公式(1)表示^[15]:

$$\text{Perplexity}(D_{\text{test}}) = \exp \left\{ \frac{-\sum_{d=1}^M \log(p(w_d))}{\sum_{d=1}^M N_d} \right\} \quad (1)$$

式中: D_{test} 为灾害语料库中的测试集, 篇; M 为收集的安全生产事故调查报告数目, 篇; $p(w_d)$ 为第 d 个文档中每个词项出现的概率; N_d 为第 d 篇文档中的词项个数, 个。Perplexity 的分子部分是对文档的似然估计, 即表示训练出的参数的生成能力, 分子值是一个正值且与文本生成能力正相关, 而分母是整个文档集的主题词数目, 模型生成能力越强, Perplexity 值越小, 对文本主题数量的测算越准确。

根据公式可以推出, 主题数量越多, 主题困惑度越小, 但是主题数量不是越多越好, 当主题数量过多时, 会出现过拟合的情况, 因此主题困惑度不能作为评估主题数量的唯一标准, 还需要结合主题一致性(Coherence)进行综合评估。主题一致性是衡量主题建模效果的一个重要指标, 它基于同一主题下词汇间语义关联性的紧密程度来判断。主题一致性越高, 主题内部词语之间的关联性越强, 即它们能够相互支撑、共同表达一个明确且集中的概念。主题一致性计算公式一般用公式(2)表示^[15]:

$$\text{Coherence}(k, V^k) = \sum_{M=2}^M \sum_{l=1}^{M-1} \log \frac{D(v_M^{(k)}, v_l^{(k)}) + 1}{D(v_l^{(k)})} \quad (2)$$

式中: $V^{(k)} = (v_1^{(k)}, v_2^{(k)}, \dots, v_M^{(k)})$ 表示在主题 k 中频率最高的 M 个词汇, 个; $D(v_l^{(k)})$ 表示测试文档集中出现词汇 $v_M^{(k)}$ 的个数, 个; $D(v_m^{(k)}, v_l^{(k)})$ 表示测试文档集中同时出现词汇 $v_M^{(k)}$ 和词汇 $v_l^{(k)}$ 的个数, 个; 1 是拉普拉斯平滑参数。

由主题困惑度确定最佳主题数量时, 采用的是手肘法(Elbow Method)。随着主题数的增加, 主题困惑度下降, 因为更多的主题可以更好地拟合数据分布; 但当主题数增加到一定程度时, 主题困惑度的下降速度会变缓, 形成类似手肘的拐点, 一般认为这个拐点是最佳主题数。由图 2、3 可以得出, 最佳主题数为 6。

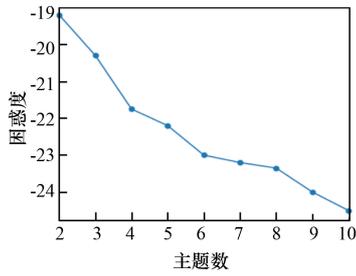


图 2 困惑度随主题变化的折线

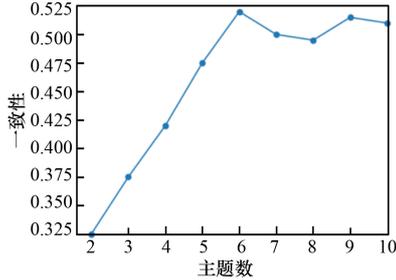


图 3 一致性随主题变化折线

借助 Python 软件运行 LDA 主题分析模型, 主题数为 6, 去除无用词, 选取每个主题中概率分布排名前十的主题词, 如表 1 所示, 主题词出现频率如表 2 所示。运用 Python 的数据可视化库绘制主题可视化图(图 4), 由图 4 可以看出, 主题之

间的重合度不高, 所选取的主题数量合理。各个主题分布频率如图 5 所示, 可以看出主题 6 的出现频率最高。

直观的图形能快速过滤掉无用信息, 高效捕捉文本中的核心要点。为了使主题词的展示更直观, 利用 Python 语言对 Wordcloud 程序包进行调用, 对主题词出现的词频进行可视化处理, 生成词云图, 如图 6 所示。在词云图中, 每个主题词出现的位置和大小反映了该主题词在整体语料库的重要性程度, 主题词的字体大小越大, 越靠近中心位置, 说明该主题词在语料库的重要性越高。以“生产”“施工”“作业”为例, 它们在词云图中的位置和大小相较其他主题词而言更加醒目, 突出他们的关键地位, 被认定为关键主题词条。

2.4 结果分析

根据 LDA 主题分析模型的结果, 可以分为 6 个主题如下。

2.4.1 施工器材安全管理

在主题 1 中, 如“塔机”“脚手架”“升降机”等均是房屋建筑项目所必需的施工器材, “进

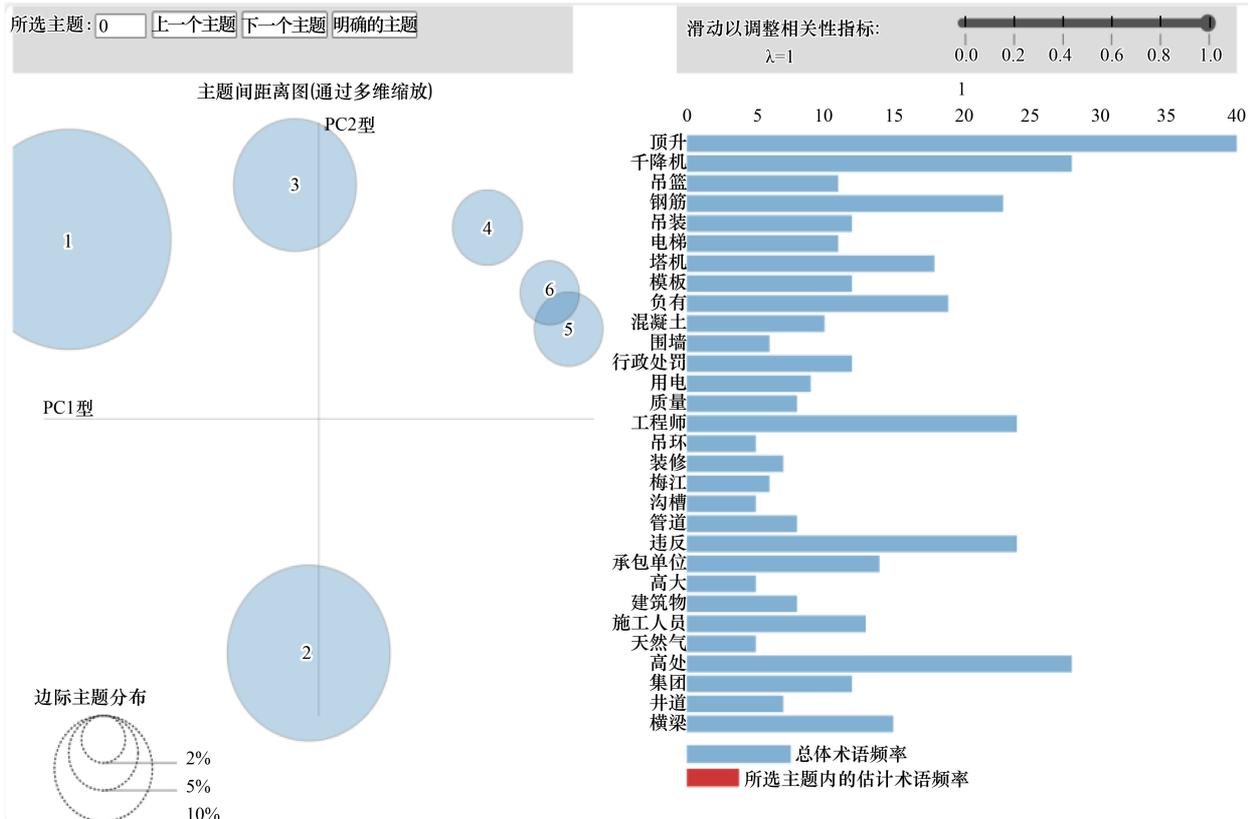


图 4 主题可视化结果

表 1 LDA 主题分析主题词汇

主题	主题词
1	装修、塔机、进料、提升机、金属、人防、脚手架、吊钩、外围、加挂
2	吊篮、吊环、沟槽、行政处罚、用电、电梯、混凝土、承包单位、井道、质量
3	钢筋、工程师、违反、电梯、责任、总监、监理、落实、管理、生产
4	围墙、吊装、天然气、处罚、电力工程、模板、管道、许可、网架、质量
5	顶升、升降机、安装、塔吊、塔式起重、拆卸、塔机、横梁、钢结构、单位
6	塔吊、拆除、构件、高处、有限公司、作业、塔式起重、防护、洞口、特种

表 2 关键词出现频率

序号	词	词频	序号	词	词频	序号	词	词频
1	生产	479	11	安装	228	21	工程	139
2	作业	420	12	有限公司	212	22	职责	138
3	施工	367	13	项目	199	23	情况	124
4	事故	356	14	责任	183	24	塔吊	119
5	监理	283	15	建设	178	25	顶升	111
6	管理	274	16	现场	169	26	导致	111
7	公司	261	17	原因	157	27	施工单位	107
8	单位	237	18	落实	155	28	工作	104
9	人员	234	19	发生	141	29	监督	104
10	到位	229	20	履行	140	30	教育	101

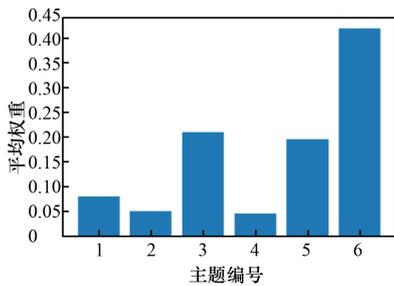


图 5 主题词分布频率



图 6 建设工程安全生产事故词云图

料”“金属”材料的储存与使用规范性，是预防安全事故的重要环节。在安全事故报告中频繁提及机械和材料相关的安全事故，表明在施工现场

进行机械和材料的安全管理是重要环节。在施工过程中需要加强机械安全和材料安全的管理，确保进场流程符合规定，进场的机械和材料都进行安全检查。

2.4.2 遵守法规

在主题 2 中，“行政处罚”是违规行为的结果，强调了法律约束力的重要性。“质量”不达标，会对安全和使用功能造成损害，面临行政处罚。“吊篮”“吊环”“电梯”等施工设备使用要严格遵守操作规程，确保合法合规、安全使用，避免不当操作带来的风险，“混凝土”等建筑材料及“用电”等能源的使用，同样需遵循安全规范，降低事故发生风险。在项目中做到有法可依、有章可循，注重项目质量，能有效地防止生产安全事故的发生。

2.4.3 组织管理

在主题 3 中，“工程师”“违反”“责任”“总监”“监理”“落实”“管理”等主题词表现了组织管理的重要性。在安全生产中，组织管理工作意义重大，组织管理保证了生产活动的整体性、效率性和可持续性。人员职责不明，没有尽到责

任义务,工作职责完成不到位都是导致安全生产事故发生的重要因素。在生产中,要保证各项工作按安全标准开展,需要建立完善的安全生产管理制度。

2.4.4 特种作业管理

在主题4中,“吊装”“天然气”“电力工程”等主题词均是特种作业,相较于其他生产过程会有更大的安全生产风险,这些特种作业工程的事故发生率较高。在生产过程中,需要建立健全的安全管理和安全检查工作,保证特种作业工作人员必须持有有效的特种作业证书,并做好教育培训工作。

2.4.5 重型设备管理

在主题5中,“升降机”“塔吊”“塔式起重”“塔机”等主题词均是重型设备,“安装”“拆卸”反映了重型设备安装与拆卸的过程,这些重型设备在安装与拆卸过程中易于发生安全生产事故,需要加强对重型设备的管理,提升管理水平,细化作业流程,强化人员培训,减少安全生产事故的发生。

2.4.6 高处作业的安全保护措施

在主题6中,“塔吊”“构件”“高处”等主题词关联高处作业,属于高风险作业,需要采取安全措施。“防护”“洞口”“脚手架”等主题词表明具体的安全保护措施。“特种”则强调高处作业所需的专业性和技术性。在生产过程中,需要规范作业流程,定期对高处作业设备进行维护,对于洞口、临边等危险区域设置警示标志,有效预防高处作业中的安全事故,保障作业人员的生命安全和身体健康。

3 建议

根据对前文文本数据的分析及主题词的提取,提出以下建议:

(1) 构建施工现场机械材料全周期安全管控体系。建立健全的机械与材料进出场检验制度,确保进场流程符合规定,进场的机械和材料能达到设计要求,符合国家相关规定标准。定期对机械设备保养维护,使其处于良好运行状态。加强对材料储存和使用过程的管理,防止因材料储存

不当或者使用不当造成的安全生产事故。

(2) 健全安全生产法规与标准体系框架。组织全体员工学习并遵守国家及地方的安全生产法律法规,确保施工活动合法合规。加强与外部机构的合作与交流,与政府部门、行业协会、科研机构等建立紧密的合作关系,及时了解最新的法律法规、安全标准和行业动态。建立完善的质量管理体系,从设计、采购、施工到验收,全程把控质量,减少因质量问题导致的事故。

(3) 创新组织架构管理模式。制定详细的组织管理制度,明确各级管理人员与作业人员的具体职责及严格的操作规程。推动智能化管理,通过智能化监控设备、数据分析平台等手段,实现对作业现场的实时监控和数据分析,及时发现并处理潜在的安全隐患。对施工现场人员进行虚拟现实(VR)培训,提升他们应对突发事件的能力。

(4) 建立特种作业人员专项认证与动态监管机制。对特种作业人员实行持证上岗制度,确保其具备相应的专业技能和安全知识。在进行特种作业前,制定详细的作业方案和安全措施,并进行技术交底和安全教育。加强现场监管,确保特种作业按照既定方案和安全措施进行,防止违章操作。

(5) 完善重型设备装拆作业标准化流程及风险评估矩阵。对于塔吊、升降机等重型设备需要定期对起重设备的钢丝绳、吊钩和螺栓等部件进行检查和更换,在安装和拆卸时严格按照规定操作流程进行,减少发生起重伤害事故的风险。

(6) 构建高处作业坠落防护系统与三维监测网络。在需要进行高空作业时,施工人员需佩戴符合要求的安全帽、安全带,在洞口、临边等危险区域布置好安全网并设置警示标志,防止人员坠落。实施有效的监控与通讯手段,在高空作业现场安装监控摄像头,实时监控作业过程,确保高空作业人员与地面指挥人员之间的通讯畅通无阻,可以实时沟通施工作业情况。

4 结论

本文基于房屋建筑工程安全事故报告文本数据,采用jieba工具实施语料分词处理,构建LDA

主题模型,通过困惑度指标与主题一致性检验确定最优主题数量为6,结合可视化词云呈现高频特征词分布,大致分为施工器材安全管理、遵守法规、组织管理、特种作业管理、重型设备管理及高处作业的安全保护措施6个主题。基于研究结果,建议从构建施工现场机械材料全周期安全管控体系,健全安全生产法规与标准体系框架,创新组织架构管理模式,建立特种作业人员专项认证与动态监管机制,完善重型设备拆装作业标准化流程及风险评估矩阵,构建高处作业坠落防护系统与三维监测网络6个方面优化现有安全管理体系,为提升建设工程安全生产管理提供了借鉴和参考。

参考文献:

- [1] 李政道,曾佳,吴恒钦.融合GWO和SVR的建筑安全事故预测模型[J].安全与环境学报,2024,24(3):1079-1086.
- [2] 丁明智,白永兰.中国建筑业安全生产水平测度与评价[J].安全与环境学报,2024,24(5):1740-1748.
- [3] 齐锡晶,康伟鑫,赵磊.基于EPC模式的安全生产管理系统分析与仿真[J].东北大学学报(自然科学版),2021,42(7):1041-1048.
- [4] 全立杨,郑霞忠,陈国梁.建筑业省域安全生产目标设置和政策优化路径研究[J].中国安全生产科学技术,2020,16(8):155-160.
- [5] 郑霞忠,全立杨,陈国梁.建筑业安全生产水平时空演变及影响因素分析[J].中国安全科学学报,2020,30(1):27-34.
- [6] 胡铮.基于模糊DEMATEL方法的建筑施工安全影响因子分析[J].工程建设,2022,54(1):63-68.
- [7] 虞芷绮,祝连波,王婷婷,等.安全管理成熟度研究现状及发展趋势分析[J].工程建设,2024,56(4):1-6.
- [8] OSWALD D,SHERRATT F,SMITH S. Managing production pressures through dangerous informality:a case study [J]. Engineering Construction and Architectural Management, 2019,26(11):2581-2596.
- [9] LI H,CHAN G,SKITMORE M. Visualizing safety assessment by integrating the use of game technology [J]. Automation in Construction,2012,22:498-505.
- [10] 谭春辉,陈晓琪,梁远亮,等.隐私泄露事件中社交媒体围观者情感分析[J].情报科学,2023,41(3):8-18.
- [11] 关菁华,毛子凡,王婷,等.基于LDA和深度模型的专利主题演化[J].数学的实践与认识,2024,54(4):68-81.
- [12] 宋军,肖超.上市公司年报风险信息披露与市场风险:基于LDA主题模型的文本研究[J].复旦学报(社会科学版),2024,66(2):165-176.
- [13] 杨益兴,吴刚,陈兰芳,等.基于LDA主题模型的多阶段生鲜消费者需求研究:以京东为例[J].管理案例研究与评论,2024,17(1):105-122.
- [14] LI H,HAN Z,ZHANG J, et al. Systematic identification of the influencing factors for the digital transformation of the construction industry based on LDA-Dematel-ANP [J]. BUILDINGS,2022,12(9):1409.
- [15] 姚成铭.富水地质深基坑施工安全风险系统动力学仿真研究[D].宜昌:三峡大学,2023.