**文献综述**

2020年我国制造业所产GDP的比重和利润较2019年分别增长3.4%和7.6%[1]。在制造业中，船舶制造业是一个很具有代表性的行业，已经成为我国具备国际竞争力的行业之一，中国造船业份额在全球市场上所占的比重正在明显上升[2],据中国船舶工业行业协会数据，我国已经连续多年占据国际造船市场份额第一的位置[3]。

中国船舶工业协会网站的数据显示，中国船舶企业的新接订单量保持领先，2020年1月到10月新接订单量1980万载重吨，占全球总新接订单量56.6%[4]。同时也应该注意到，船舶制造行业是一个劳动密集、技术密集、资金密集的高风险综合性行业，涉及船舶和海洋工程装备建造、修理、改装以及配套服务，作业过程取材纷繁，工种和工序多样且复杂，涉及技术面较广，属于事故高发的行业[5]。如：韩国现代重工企业自1974年成立到2020年的46年间，因重大生产安全事故导致死亡的工人累计达470人，平均每年有超过10人因工伤事故死亡[6]。我国船舶制造行业的安全状况也不容乐观，数量庞大的险兆事件、大量的轻伤等职业伤害事件、时有发生的重伤和死伤事件是当前我国船舶修造企业的现状。2001年7月17日，上海市沪东中华造船有限公司在门式起重机在吊装作业过程中，发生下坠特大事故，造成36人死亡，3人受伤。2006年11月武昌造船厂发生易燃气体爆燃事故造成2人死亡，6人受伤。 2008年12月在青岛北海船舶重工有限公司厂区内，正改装的外籍货轮货仓内的钢管支架突然发生大面积倒塌，造成2人死亡、6人受伤[7]。

事故的发生不仅影响企业的生产经营，同时也考验国家的治理能力。为了预防事故的发生，多数发达国家都建立了安全生产事故数据库，并基于事故数据进行风险分析和事故预测[2]。安全生产事故数据库可以提供可靠而全面的信息，一方面用于事故案例学习和研究，另一方面分享经验教训，从而达到预防事故发生的目的。有些国家（相关机构）所构建的事故数据库不仅收集本国发生的事故信息，而且还收集国际上其他国家发生的重大事故信息。如：美国的全球性综合灾害数据库、澳大利亚的EMA灾害数据库等。对中国来说，一方面政府在不断地强调安全生产的重要性，且对于事故的查处与问责程度越来越严厉，另一方面企业为了经营需求和安全管理考虑，一般都会创建企业事故数据库。事故数据库是由一系列事故的字段信息组成，例如事故的时间、事故的地点、事故的原因、事故的类型、事故详细报告等字段。

事故数据库中的事故信息记录了发生事故的特征，包括结构化数据和非结构化数据。结构化数据是能够用数据或统一的结构表示的信息，如时间、数字、符号。非结构化数据不符合任何预定义的模型，较难格式化，但是它却承载了绝大多数有价值的数据。结构化数据与非结构化数据或半结构化数据相比，在数据结构上更加“整齐”，更易于分析利用。非结构化数据详细描述事故发生始末缘由，对企业的安全管理有着重要的作用，但是事故的内容描述没有一种统一格式，会因事故类别和登记人而有所不同，这就为有效内容的提取带来困难，需要借助一定的技术手段。文本挖掘技术是指使用计算机技术来分析大量文本数据并从分析中揭示有价值的信息，已被证明可以运用于事故预防研究。

目前大多数研究都是针对非结构化数据进行挖掘研究，如Botao Zhong等[8]人基于文本挖掘和深度学习的技术，对一个大型交通基础设施项目中收集的危险记录进行危险分析，提出并验证了具有自动分析危险记录的框架。Joon-Soo Kim等[9]人通过整理收集近十年发生的火灾事故数据并对其进行文本挖掘方法和主成分分析方法研究，讨论了事故类型、事故因素等特征，结果显示常见影响事故的因素为火灾中直接气体和火焰，季节影响因素为冬季零度以下温度和干燥天气。Zhaochen Wang等[10]人根据中国海事局收集的内河水上运输事故报告，基于FP-Growth算法的关联性分析方法分析了事故报告中的关键风险因素，建立了完整的风险评估体系。Ze Wang等[11]人结合中国山西六安集团经营的大型现代煤矿提供的危险记录数据，利用文本挖掘和复杂网络方法研究危险源之间的关系，发现风险网络之间的关联较为复杂。Donald E. Brown等[12]人从联邦铁路管理局的铁路事故报告入手，运用文本挖掘技术来自动发现事故特征，结果发现事故成本的预测精度显著提高。李珏等[13]人根据收集到的建筑施工高坠事故报告数据，使用文本挖掘的方法进行分析，得到高坠事故致因特征项。谭章禄等[14]人针对潞安集团司马煤业有限公司的安全隐患记录，利用文本挖掘对安全隐患数据进行聚类分析，分析了隐患地点、隐患对象、隐患状态与隐患程度之间的关联关系，发现安全隐患问题和隐患类型比较分散。而结构化数据与非结构化数据或半结构化数据相比，在数据结构上更加“整齐”，更易于分析利用。

随着信息资源的快速膨胀，企业的事故数据趋于多样化和复杂化，事故数据库的记录缺乏规范化会导致很多数据利用率低，尚有巨大的挖掘和利用空间。企业的安全生产信息处理能力无法与日益提高的安全信息数据量与复杂程度相匹配，传统数据分析方法不能再有效地处理这些信息，机器学习和数据挖掘等人工智能的兴起，为海量事故数据的挖掘提供了途径。

**国内外研究现状：**

在数据挖掘领域，国外针对安全生产大数据的研究已经有了很多尝试，大多是在建筑行业、煤矿行业、公路安全、道路交通方面，在船舶制造方面的安全事故研究还鲜少。Antonio trillo gabello等[15]人收集2003-2015年西班牙建筑行业的事故数据采用文本挖掘和关联规则相结合的方法探讨了各阶段的事故。Ziwei fa等[16]人以2011-2020年883份煤矿事故报告作为原始数据，借助文本挖掘和关联规则算法得出了关联关系，并得到了验证。Subasish das等[17]人收集了2014-2018年公路碰撞的事故数据进行文本挖掘，并应用关联规则算法挖掘行人碰撞的成因。Zunxiang qiu等[18]人通过对307份煤矿事故的文本挖掘，并基于关联规则算法的研究发现了不同事故类型的核心因素和关联规则以及关键环节。Erkan cakir等[19]人基于2008-2017年期间的477条拖船事故记录，采用不同的关联规则算法提取关联关系，得出船龄超过20年的拖船是严重事故的关键指标，以及船体/机械损坏和碰撞类型的事故占拖船事故的总数一半以上。Sarkar（2018）[20]则利用支持向量机、随机森林和人工神经网络等优化的机器学习方法，构建职业事故分类预测框架，论文以某钢铁企业为例，进行了实证研究，并对比分析了多种方法的效率和准确率，为机器学习技术在安全生产领域的应用提供了参考。

国内数据挖掘起步较晚，在安全事故上的应用较少，目前在建筑行业、煤矿业、公路安全、电力产业等产业有研究。李卉等[21]人针对建筑模板坍塌事故的事故致因因素采用关联规则算法进行挖掘，得到事故致因链。薛楠楠等[22]人采用文本挖掘技术对建筑施工安全事故报告进行处理，识别出8项建筑工人不安全行为及27项影响因素。黄亚春等[23]人使用文本挖掘的方法将非结构化文本处理成结构化文本，运用关联规则对关键词进行分析，挖掘出不同的风险因素对工程安全事故的影响。牛毅等[24]人基于某化工企业和浙江省高速公路的事故数据采用文本挖掘的方法对结构化数据进行挖掘分析。

国内外在安全生产事故领域都进行了事故挖掘研究，目前大多数研究只是针对单一领域安全生产事故数据的挖掘分析，运用到的数据质量各异，分析方法繁多，导致研究结果的质量参差不齐，存在较大的局限性。同时大多研究更注重某一方法的简单应用，从行业来看，研究主要集中于建筑施工行业、道路交通和煤矿行业，在其他安全生产领域中的研究还相对较少。在研究内容方面，主要分为对事故的分类、聚类、回归分析以及潜在关联规则的挖掘，在事故数据结构方面，绝大多数都是针对非结构化数据的研究。大多针对单一领域、单一数据结构或单一分析方法的简单应用，未有效分析各类事故数据结构的特征和利用现状，未根据不同数据结构特征构建相应的完整事故信息识别流程。

参考文献：

1. 中国2020年国民经济和社会发展统计公报. 国家统计局[EB]. [2021-03-04].<http://www.tjcn.org/tjgb/00zg/36540_26.html>
2. Palacios A, Mata C, Barraza C. Historical analysis of accidents in the Mexican chemical industry[J]. Revista Mexicana De Ingenieria Quimica, 2021, 20(1):113-133.
3. 中国船舶在线 [EB].[2020-07-09] <http://www.shipol.com.cn/zhzx/169ecc0b31f64375ab7fd036306dd6df.htm>
4. 央视网快看. 中国造船订单量位居全球第一[EB/OL]. [2020-12-11]<https://weibo.com/tv/show/1034:4578476439830595?from=old_pc_videoshow>.
5. Park Jonghee, KIM HANJUNE, YOON JAEJUN, et al. Development of an Ultrasound Technology-Based Indoor-Location Monitoring Service System for Worker Safety in Shipbuilding and Offshore Industry[J]. Processes, 2021, 9(2):304-304.
6. 国际船舶网. 韩国现代重工重大安全生产事故[EB/OL].[2021-03-01]<http://www.eworldship.com/html/2021/Shipyards_0301/168503.html>.
7. 船厂事故整理[EB/OL]. [2021-02-04]<https://wenku.baidu.com/view/c3a9f7239b8fcc22bcd126fff705cc1754275f05.html>
8. Zhong B T, Pan X, Love P E D, et al. Hazard analysis: A deep learning and text mining framework for accident prevention[J]. Advanced Engineering Informatics, 2020, 46,101152.
9. Kim J S, Kim B S. Analysis of Fire-Accident Factors Using Big-Data Analysis Method for Construction Areas[J]. Ksce Journal Of Civil Engineering, 2018, 22(5):1535-1543.
10. Wang Z C, Yin J B. Risk assessment of inland waterborne transportation using data mining[J]. Maritime Policy & Management, 2020, 47(5):633-648.
11. Wang Z, Li H J, Tang R W. Network analysis of coal mine hazards based on text mining and link prediction[J]. International Journal Of Modern Physics C, 2019, 30(7):
12. Brown D E. Text Mining the Contributors to Rail Accidents[J]. IEEE Transactions On Intelligent Transportation Systems, 2016, 17(2):346-355.
13. 李珏, 王幼芳. 基于文本挖掘的建筑施工高处坠落事故致因网络分析[J]. 安全与环境学报, 2020,20(04):1284-1290.

LI Jue, WANG Youfang. Causation network analysis of the construction falling or collapsing accidents based on the text mining [J]. Journal of Safety and Environment, 2020,20(04):1284-1290.

1. 谭章禄, 陈晓, 宋庆正，等. 基于文本挖掘的煤矿安全隐患分析[J]. 安全与环境学报, 2017,17(04):1262-1266.

TAN Zhanglu, CHEN Xiao, SONG Qingzheng, et al. Analysis for the potential hazardous risks of the coal mines based on the socalled text mining [J]. Journal of Safety and Environment, 2017,17(04):1262-1266.

1. CABELLO A T, MARTINEZ-ROJAS M, CARRILLO-CASTRILLO J A, et al. Occupational accident analysis according to professionals of different construction phases using association rules [J]. Safety Sci, 2021, 144(
2. FA Z W, LI X C, QIU Z X, et al. From correlation to causality: Path analysis of accident-causing factors in coal mines from the perspective of human, machinery, environment and management [J]. Resour Policy, 2021, 73(
3. DAS S, TAMAKLOE R, ZUBAIDI H, et al. Fatal pedestrian crashes at intersections: Trend mining using association rules [J]. Accident Anal Prev, 2021, 160.
4. QIU Z X, LIU Q L, LI X C, et al. Construction and analysis of a coal mine accident causation network based on text mining [J]. Process Saf Environ, 2021, 153(320-8).
5. CAKIR E, FISKIN R, SEVGILI C. Investigation of tugboat accidents severity: An application of association rule mining algorithms [J]. Reliab Eng Syst Safe, 2021, 209.
6. Sarkar S, Vinay S, Raj R. Application of optimized machine learning techniques for prediction of occupational accidents[J]. Computers and Operations Research, 2018, 37(3): 1-15.
7. 李卉. 基于关联规则的建筑模板坍塌事故致因链研究[D].华侨大学,2019.

LI Hui. Research on Cause Chain of Building Template Collapse Accident Based on Association Rules[D]. Huaqiao University,2019.

1. 薛楠楠,张建荣,张伟,赵挺生.基于文本挖掘的建筑工人不安全行为及其影响因素研究[J].安全与环境工程,2021,28(02):59-65+85.

XUE Nannan, ZHANG Jianrong, ZHANG Wei, ZHAO Tingsheng. Correlations of Construction Worker’s Unsafe Behaviors and the Influential Factors by Using Text Mining[J]. Safety and Environmental Engineering,2021,28(02):59-65+85.

1. 黄亚春. 基于自然语言处理的建筑工程安全事故报告风险研究[D].华中科技大学,2019.

HUANG Yachun. Research Based on Natural Language Processing for Risk of Construction Accident Reports[D]. Huazhong University of Science and Technology,2019.

1. 牛毅. 基于数据驱动的安全生产事故致因分析方法研究[D].中国地质大学(北京),2020.

NIU Yi. Research on the method of causation analysis of production accident based on data driving[D]. China University of Geosciences(Beijing),2020.

1. JEONG J H, WOO J H, PARK J. Machine Learning Methodology for Management of Shipbuilding Master Data [J]. Int J Nav Arch Ocean, 2020, 12(428-39).
2. 吴志强. 基于石油企业HSE大数据的隐患关联分析与可视化方法研究[D].中国石油大学(北京),2019.

### WU Zhiqiang. Research on Visualization Method and Hidden Hazard Association Analysis Based on HSE Big Data of Petroleum Enterprises[D]. China University Of Petroleum(Beijing),2019.

1. 罗玉林. 基于大数据的高速公路交通风险预测[D].长沙理工大学,2020.

LUO Yulin. Traffic risk prediction of Expressway based on big data[d]. Changsha University of Science & Technology,2020.

1. 曹帅. 基于大数据的航空安全事故数据挖掘研究[D].安徽建筑大学,2021.

CAO Shuai. Research on data mining for aviation safety accidents based on big data[D]. Anhui Jianzhu University,2021.

1. 陈晓. 基于数据挖掘的煤矿安全管理知识可视化研究[D].中国矿业大学(北京),2017.

CHEN Xiao. Knowledge Visualization of Safety Management in Coal Mines Based on Data Mining[D]. China University of Mining and Technology(Beijing),2017.

1. 张嘉瓴. 基于数据挖掘技术的道路交通事故分析[D].大连理工大学,2020.

ZHANG Jialing. Analysis of Road Traffic Accidents Based on Data Mining Technology[D]. Dalian University of Technology, 2020.

1. AYHAN B U, DOGAN N B, TOKDEMIR O B. An Association Rule Mining Model for the Assessment of the Correlations between the Attributes of Severe Accidents [J]. J Civ Eng Manag, 2020, 26(4): 315-30.
2. SANMIQUEL L, ROSSELL J M, VINTRO C. Study of Spanish mining accidents using data mining techniques [J]. Safety Sci, 2015, 75(49-55.
3. CASTRO Y, KIM Y J. Data mining on road safety: factor assessment on vehicle accidents using classification models [J]. Int J Crashworthines, 2016, 21(2): 104-11.
4. Kritika Singh, J.Maiti, Krantiraditya Dhalmahapatra. Chain of events model for safety management: Data analytics approach[J]. Safety Science, 2019, 118:568-582.
5. Ching-Wu Cheng, Chen-Chung Lin, Sou-Sen Leu. Use of association rules to explore cause–effect relationships in occupational accidents in the Taiwan construction industry[J]. Safety Science, 2009, 48(4):436-444.
6. ABDEL-BASSET M, MOHAMED M, SMARANDACHE F, et al. Neutrosophic Association Rule Mining Algorithm for Big Data Analysis [J]. Symmetry-Basel, 2018, 10(4): .