

数据驱动下地铁安全事件致因分析研究

秦强 邵明 李杰 周波 侯林 陈瀚 李有为 何金 皮小红 宋傲雪

荆州市市政建设集团有限公司 湖北荆州 434000

DOI: 10.12238/ems.v7i8.14716

[摘要] 中国城市轨道交通进入快速发展阶段, 地铁运营安全事件频发。面对海量事故文本数据, 本文提出基于文本双层网络的致因分析方法, 通过融合词网络与句网络构建 NWS 模型, 结合 RCI-NWS 致因文本识别算法, 从非结构化文本中提取关键致因因素。以武汉地铁近十年事故数据为案例, 构建致因双层网络并挖掘最大可能致因链, 发现设备故障与公共安全类事故为主要隐患, 为地铁安全管理提供数据驱动的决策支持。

[关键词] 地铁安全事件; 数据驱动; 文本双层网络; 致因分析; 事故致因链

Research on Cause Analysis of Subway Safety Incidents Driven by Data

Qin Qiang Shao Ming Li Jie Zhou Bo Hou Lin Chen Han Li Youwei He Jin Pi Xiaohong Song Aoxue

Jingzhou Municipal Construction Group Co., Ltd. Jingzhou 434000 Hubei, China

[Abstract] China's urban rail transit has entered a stage of rapid development, with frequent safety incidents in subway operations. Facing massive accident text data, this paper proposes a cause analysis method based on a text dual-layer network. By integrating word networks and sentence networks, the NWS model is constructed, and combined with the RCI-NWS cause text recognition algorithm, key causal factors are extracted from unstructured texts. Taking the accident data of Wuhan Metro in the past decade as a case, a dual-layer cause network is built to mine the most probable causal chains. It is found that equipment failures and public safety accidents are the main hidden dangers, providing data-driven decision support for subway safety management.

[Key words] subway safety incidents; data-driven; textual dual-layer network; cause analysis; accident cause chain

1. 引言

地铁作为城市交通的核心载体, 在缓解交通压力的同时, 其安全运营面临严峻挑战。2020 年“新基建”计划推动下, 中国城市轨道交通进入集中建设阶段, 但郑州地铁“7·20 事件”等事故表明, 地铁项目因参与主体多、专业复杂, 安全管控难度极大^[1]。现有研究对非结构化事故文本的致因挖掘存在信息提取不完整、致因链构建不系统等问题, 亟需数据驱动的智能分析方法^[2]。

本文提出基于复杂网络的文本双层网络模型, 通过 NWS 关键词提取与 RCI-NWS 致因识别方法, 从事故报告中构建致因链, 以武汉地铁数据验证模型有效性, 为安全风险防控提供新路径。

2. 模型构建

2.1 基于复杂网络的文本关键词提取方法

传统关键词提取忽视句子语境影响^[3], 本文提出融合词

网与句网的双层网络模型 (New Word-Sentence, NWS)。该模型通过以下步骤实现综合特征值计算:

文本预处理: 清洗事故文本, 分词并标注词性;

双层网络构建: 词网以词语为节点, 共现关系为边; 句网以句子为节点, 共享关键词为边;

特征值计算:

词网节点特征值 $WE_i = \alpha \cdot bci + \beta \cdot cci$, 其中 bci 为词频, cci 为共现强度;

句网节点特征值 $SWE_j = w_j$, 即句子包含的关键词权重和;

综合特征值 $E_i = \lambda WE_i + \gamma SWE_j$, 通过权重系数融合双层网络特征。

2.2 RCI-NWS 致因文本识别方法的建立

为了使用关键字提取方法来识别致因文本, 本章所提出的方法需要基于以下假设:

假设 1: 每条故障记录至少含一个致因文本。

假设 2: 致因文本由“名词 + 动词”或独立名词(与默认词组合)构成。

具体步骤如下:

步骤 1 字典构建: 整合文献术语(Dic1)与关键词提取结果(Dic2)

步骤 2 停用词过滤: 通过 Apriori 算法识别频率词与无意义词, 构建停用词表 ST。

步骤 3 致因匹配: 利用 $ND = Dic1 + Dic2 - ST$ 匹配致因名词, 通过依存关系寻找致因动词, 无动词时与默认词组合成致因文本。

2.3 地铁事故文本致因双层网络模型

该双层网络包含致因网络与句网络。致因网络的节点由文本提取和识别得到的致因因素构成, 若两个致因因素同处一条事故记录, 则二者之间存在一条边, 并以它们同时出现的次数为边权重^[4]。句网络由各条记录构成, 每条记录视为一个节点, 若两节点间存在共同有效词, 则认为二者相连, 共同词数量即为边权重。上下层网络构建完成后形成致因双层网络, 其仍具备网络特性^[5]。基于此, 可利用容量-负载模型的相继故障方法, 在下层致因网络中挖掘致因链。将双层网络视作特殊文本双层网络, 通过计算下层网络节点的累加综合特征值, 得到致因因素危险值, 再与边概率相乘, 从而确定节点或致因链的风险值。

3. 案例分析 (Case study)

3.1 数据采集与预处理

采集 2012-2022 年武汉地铁官方微博发布的 998 条运营记录, 清洗后获得 436 条有效数据。统计显示, 2 号线(106 起)、4 号线(98 起)事故最多, 因线路客流量大、设施老化及途经热门区域。

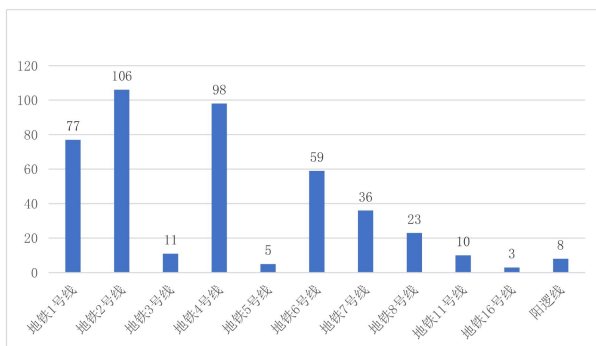


图1 武汉运营地铁线路发生的事故数(因故障晚点)(2012年3月-2022年10月)

3.2 关键词与致因文本提取

首先, 利用 NWS 关键词提取方法对原始事故文本进行清洗和筛选, 使文本长度变短, 一遍下一步进行致因文本识别, 原始记录“地铁 1 号线冲抢车门导致设备故障”提取为“冲抢车门、设备故障”;

通过 RCI-NWS 识别致因文本, 如“信号故障、失去通信”对应代码 C16、C17。

3.3 致因双层网络构建

将致因因素映射为节点, 构建边关系(如 C9 拥堵与 C15 乘客冲突共现 235 次), 通过容量-负载算法挖掘最大致因链, 主要结果包括:

最长致因链: $C9 \rightarrow C10 \rightarrow C15 \rightarrow D4$ (拥堵 $\rightarrow$ 摔伤 $\rightarrow$ 乘客冲突 $\rightarrow$ 公共安全);

高频结束节点: D3 (设备类, 占比 32%)、D4 (公共安全类, 占比 28%);

典型致因链:

$C3 \rightarrow D3$ (异物侵限 $\rightarrow$ 设备故障);

$C16 \rightarrow C17$ (信号故障 $\rightarrow$ 通信中断);

$C28 \rightarrow D5$ (自然灾害 $\rightarrow$ 自然类事故, 唯一自然类链)。

3.4 结果讨论

邀请武汉地铁运营主管与高级工程师进行会议讨论, 对爬取到的事故数据进行统计分析结果如图 2。统计分析表明, 设备类(38%)与公共安全类(32%)事故占比最高, 与致因链分析结果一致。

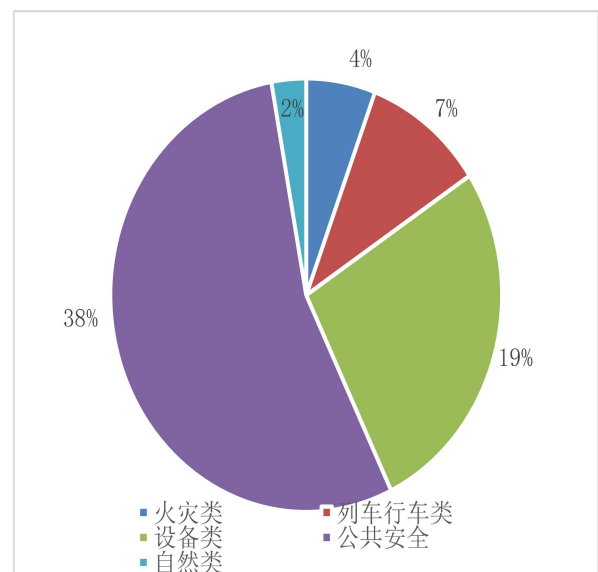


图2 不同事故故障类型占比

(1) 致因链最长为 4, 而事故记录中的最多致因因素为 5, 表明一件事故中并非所有致因要素都构成致因链的某一环节。

(2) D3 和 D4 作为最频繁作为致因链结束节点, 分别表示设备类事故和公共安全类事故较常发生。

(3) C28→D5 表示由自然灾害引起的自然类事故, 是唯一一条以 D5 结束的致因链, 说明自然类事故发生较少。

(4) 初始的节点各不相同, 各个节点为初始点的致因链唯一。

多数事故的致因链长度集中在 2 至 3, 表明事故通常由少数关键因素直接引发。D1 和 D2 作为致因链的结束节点, 分别代表火灾类和列车行车类事故, 发生频率较高。C29→D3 致因链显示操作不当是设备类事故的主要原因, 需加强培训与监督; C30→D4 揭示公共安全类事故与管理缺陷直接相关, 需强化安全管理; C31→D5 虽较少出现, 但表明自然灾害情况下应急响应不足也可能导致事故, 需提升应急处理能力。

致因链分析助力管理人员识别事故直接原因及潜在风险点和管理漏洞, 为制定预防措施提供依据。将致因链分析结果与实际统计分析结果对比, 两者基本一致, 表明设备类与公共安全类是武汉地铁最大的安全隐患。因此, 地铁运营相关人员应重视设备维修保养, 降低事故故障发生率; 故障发生后及时疏散人流, 减少公共安全问题引发的二次事故, 降低损失。武汉地铁可加强设备日常检查维护, 确保机械部件处于最佳工作状态; 在公共安全方面, 增加安全培训频率, 模拟紧急情况演练, 提高员工应急处理能力; 安装先进监控系统实时监控站内情况, 与公安机关合作打击违法犯罪, 高峰时段增派安保人员巡逻, 防范拥挤踩踏事件; 针对恶劣天气或大型活动等特殊事件, 提前制定应急预案, 确保运营连续性和安全性。

对比表明, 本文提出的关键词提取和致因文本识别方法在地铁事故致因分析领域具有可行性, 结果更精确且节省了人工处理时间, 有利于后续研究, 为识别地铁事故致因要素及致因链提供了新思路。

4. 结论与展望

本文针对地铁安全事故的非结构化文本数据, 提出一种

事故双层致因网络模型, 深入探究事故致因及其分析, 主要成果如下:

(1) 创新性地构建了融合词语与句子影响的 NWS 模型, 借助复杂网络理论计算节点综合特征值, 精准筛选出文本数据关键词, 助力地铁工作人员快速掌握事故故障文本核心信息。

(2) 鉴于 NWS 方法提取的关键词可能包含非致因文本, 提出致因文本识别方法, 依托致因字典与停用词表匹配, 精准识别事故故障文本中的致因词语, 深度服务致因分析。

(3) 利用爬虫技术获取 2012 至 2022 年武汉地铁微博运营数据, 结合 NWS 方法构建事故致因双层网络, 整合致因网络与文本句网络, 计算致因节点综合特征值确定节点危险度, 运用相继故障方法揭示事故最大可能致因链及关键因素。实验表明, 基于武汉地铁近十年数据的致因分析结果与实际情况高度契合, 未来可应用于实际地铁事故致因分析。

[参考文献]

[1] 杨赛霓, 刘悦, 刘昌新. 郑州地铁 5 号线“7·20”暴雨洪涝灾害中的人员伤亡特征与避难行为[J]. 地理学报, 2023, 78 (7): 1635-1651.

[2] XIE J, LIU B, HE L, et al. Quantitative Evaluation of the Adaptability of the Shield Machine Based on the Analytic Hierarchy Process (AHP) and Fuzzy Analytic Hierarchy Process (FAHP) [J]. Advances in Civil Engineering, 2022, 2022.

[3] 郭瑞, 李忠富. 基于 FAHP-TOPSIS 的地铁施工安全风险评价[J]. 土木工程与管理学报, 2019, 36 (2): 143-148.

[4] ZHOU Z, IRIZARRY J, ZHOU J. Development of a database exclusively for subway construction accidents and corresponding analyses[J]. Tunnelling and Underground Space Technology, 2021, 111.

[5] 张顶立, 李治国, 房倩, 等. 城市地铁工程建设风险管理体系的构建与应用[J]. 中国安全科学学报, 2013, 23 (10): 146-152.