

文章编号:1671-1637(2026)05-0260-15

基于文本挖掘的铁路沿线自然灾害网络潜在风险 路径预测

顾爽¹,程国柱^{*1},闫冬阳²

(1. 东北林业大学 土木与交通学院,黑龙江 哈尔滨 150040; 2. 青岛大学 商学院,山东 青岛 266071)

摘要:为揭示铁路沿线自然灾害链式传播机理,量化其对铁路运输安全的威胁并解析风险演化规律,构建了基于文本挖掘与复杂网络的铁路沿线自然灾害风险路径预测模型。以铁路沿线历史灾害文本数据集为基础,改进了文本挖掘技术,优化词袋模型并结合关键字提取方法,精准抽取了灾害数据中的铁路站点、致灾因子、后果及严重程度等核心要素;基于复杂网络理论,以关键词为节点(3类实体),以同事件共现语义关联为边(6类关系),建立了铁路灾害风险异构网络;设计多元路径搜索算法遍历网络拓扑,融合关联共现矩阵量化节点间风险传导强度,实现了多类型传播路径的过程化挖掘。分析结果表明:模型接收者操作特征曲线接近左上角,曲线下面积为0.938,准确率达94.873%,F1分数为0.899;量化输出节点间风险传导值,成功定位高概率灾害链为芒康站→强降雪→香格里拉站→人员受灾→电力设备损毁(风险值0.866)、江达站→贡觉站→强对流天气→人员伤亡(风险值0.841)。所得模型通过文本驱动构建灾害语义网络,可实现铁路风险路径的定量化预测与关键传导链识别,精准定位二次灾害高发链路,为复杂环境铁路的风险主动防控提供支持。

关键词:铁路运输;自然灾害;文本挖掘;潜在风险预测;异构网络;传播路径

中图分类号:U298 文献标志码:A DOI:10.19818/j.cnki.1671-1637.2026.121

Prediction of potential risk paths in natural disaster network along railway lines based on text mining

GU Shuang¹, CHENG Guo-zhu^{*1}, YAN Dong-yang²

(1. School of Civil Engineering and Transportation, Northeast Forestry University, Harbin 150040, Heilongjiang, China; 2. Business College, Qingdao University, Qingdao 266071, Shandong, China)

Abstract: To reveal the chain propagation mechanism of natural disasters along railway lines, quantify their threats to railway transportation safety, and analyze the risk evolution laws, a prediction model for natural disaster risk paths along railway lines based on text mining and complex networks was constructed. Based on the historical disaster text dataset along railway lines, the text mining technology was improved; the bag-of-words model was optimized, and combined with the keyword extraction method, core elements such as railway stations, disaster-

出版历程:2025-05-22 收稿,2025-10-28 修回,2025-11-27 录用

基金项目:中央高校基本科研业务费专项资金项目(2572023CT21);黑龙江省哲学社会科学规划年度项目(24GLC014)

作者简介:顾爽(1994-),女,黑龙江依兰人,讲师,理学博士,博士后,E-mail:s.gu@nefu.edu.cn.

*通信作者:程国柱(1977-),男,河北大成人,教授,博士生导师,工学博士,博士后,E-mail:guozhucheng@126.com.

引用格式:顾爽,程国柱,闫冬阳.基于文本挖掘的铁路沿线自然灾害网络潜在风险路径预测[J].交通运输工程学报,2026,26(5):260-274.

Citation:GU Shuang, CHENG Guo-zhu, YAN Dong-yang. Prediction of potential risk paths in natural disaster network along railway lines based on text mining[J]. Journal of Traffic and Transportation Engineering, 2026, 26(5): 260-274.

causing factors, consequences, and severity levels in the disaster data were accurately extracted; based on complex network theory, a heterogeneous network of railway disaster risks was established, with keywords as nodes (3 types of entities) and co-occurrence semantic associations in the same event as edges (6 types of relationships); a multi-path search algorithm was designed to traverse the network topology, and the associated co-occurrence matrix was integrated to quantify the risk transmission intensity between nodes, realizing the procedural mining of multiple types of propagation paths. The analysis results show that the receiver operating characteristic curve of the model is close to the upper left corner; the area under the curve is 0.938; the accuracy reaches 94.873%, and the F1 score is 0.899; the risk transmission values between node pairs are quantitatively output, and high-probability disaster chains are successfully located as: Markam Station → heavy snowfall → Shangri-La Station → personnel disaster → power equipment damage (risk value 0.866), and Jomda Station → Gonjo Station → severe convective weather → casualties (risk value 0.841). The obtained model constructs a disaster semantic network driven by text, which can achieve the quantitative prediction of railway risk paths and the identification of key transmission chains, accurately locate the high-incidence links of secondary disasters, and provide support for the proactive risk prevention and control of railways in complex environments.

Keywords: railway transportation; natural disaster; text mining; potential risk prediction; heterogeneous network; propagation path

Publication history: Received 2025-05-22; Received in revised form 2025-10-28; Accepted 2025-11-27

Funding: Fundamental Research Funds for the Central Universities (2572023CT21); Annual Philosophy and Social Science Foundation of Heilongjiang Province (24GLC014)

* **Corresponding author:** CHENG Guo-zhu, professor, PhD, E-mail: guozhucheng@126.com.

0 引 言

铁路在交通运输体系中占据核心地位,是国家和综合交通体系的重要组成部分。自 2020 年以来,中国铁路营业里程逐年增长,据 2025 年 6 月交通运输部发布的《2024 年交通运输行业发展统计公报》显示,截至 2024 年底,全国铁路营业里程 16.2 万 km,投产新线 3 113 km。然而,铁路运输作为重要的线性连续工程系统,因空间跨度极大、线路走向固定等特点,其运营安全受地质、地形、地势等因素的影响常常面临山体崩塌、滑坡、泥石流等复杂自然灾害风险,致使整条线路或区域路网的瘫痪,并伴随着对环境、经济甚至生命的巨大威胁。铁路作为中国的百年大计,必须树立防灾减灾意识才能实现可持续发展之路。如何应对自然灾害风险成为铁路运营中长期面临的难点^[1]。多年来,人们普遍认为通过相应的评估与预测等措施可以有效地降低自然灾害在铁路运营中带来的不良后果^[2]。铁路系统独特的线性脆弱性、灾害链式传播特性以及空间敏感性,使风险路径预测成为铁路运输安全领域应对自然灾害

的重要举措,可控制潜在后果和有效减小损失,防止级联失效事件发生^[3]。

中国铁路网络规模庞大、运行环境复杂多样,穿越多种气候带和地形地貌区域,使得铁路灾害的风险评估对运营安全至关重要,其超前预测能力可显著降低财产损失^[4]。当前研究主要聚焦以下两大方向。

一方面是围绕灾害发生概率、易发性与危险性展开的灾害发生机理评估。例如,Lan 等^[5]利用神经网络和模糊数学等人工智能方法,建立了多因素模式识别模型,用于矿山灾害的风险概率预测。宋哲等^[6]面向铁路局集团公司层面,识别出 35 项安全风险,通过关联关系分析明确风险间的传递关系,进而提出一种基于概率分析的铁路安全风险传递网络模型。吴晶晶等^[7]将反向传播(Back Propagation, BP)神经网络模型与地理信息系统(Geographical Information System, GIS)技术相结合,用于预测崩塌滑坡的易发性。Wang 等^[8]结合 GIS 技术建立了一个地质灾害易发性综合预测模型来保证生态环境的可持续发展。符洪恩等^[9]运用组合赋权法通过云

模型算法,对深圳市暴雨灾害进行危险性预测。宋国策等^[10]为实现铁路外部环境安全隐患的量化风险评价,通过分析铁路外部环境安全隐患导致风险事件的发生规律和事故致因机理,构建面向铁路外部环境 18 大类安全隐患的普适性风险评价指标体系。铁路沿线灾害发生机理评估是风险防控的第一道防线,其核心在于数据驱动的预测模型构建。现有研究多是单一灾害预测或仅针对特定线路开发,未有效解决铁路线性工程跨区域的场景迁移问题。

另一方面是灾损与应急需求预测。龚艳冰等^[11]提出基于高斯过程回归模型的洪涝灾害损失预测方法,并对重庆市洪灾受灾人数、农作物受灾面积和直接经济损失进行预测。Cao 等^[12]提出了一种基于 GM(1,1)模型的海洋灾害直接经济损失的预测方法。李博等^[13]以广东省台风灾害为例,利用遗传算法(Genetic Algorithm, GA)改进 BP 神经网络形成 GA-BP 神经网络模型,用于直接经济损失预测。Zhang 等^[14]针对山洪灾害的灾害机理和属性特征提出了基于径向基函数(Radial Basis Function, RBF)神经网络的洪水灾害损失评估模型。Huang 等^[15]通过数据训练,建立了地震灾民伤亡预测的极限学习神经网络结构。胡忠君等^[16]以灰色系统理论为基础,结合库存管理方法和改进 GM(1,1)动态预测模型,提出了应急物资动态需求模型。刘芳等^[17]提出了一种基于改进蚁群优化 BP 神经网络智能算法的模型,来预测洪涝灾害应急物资需求量和受灾转移人数。现有方法能有效提升精度,但对于铁路系统特有的链式致灾效应未考虑其时空异构性。

综上所述,当前铁路灾害预测面临的技术挑战主要体现在三方面:一是铁路沿线灾害数据文本化、碎片化、孤岛化现象,当前铁路灾害数据多是文本记录形式,难以从文本数据中获取有效信息及数据间的关联关系;二是模型适应性问题,通用模型在特定地理环境和铁路线网中预测精度不足;三是灾害链传导机制的瓶颈,传统方法往往忽略铁路沿线灾害风险间的潜在传播机制^[18]。风险路径是解决上述问题的一个重要突破口。挖掘潜在的风险路径,提前采取措施,可以防止灾害在铁路沿线的扩张。因此,本文针对铁路系统特点提出了一种新的基于文本挖掘技术与复杂网络理论的预测模型来识别潜在风险路径,将铁路灾害从“概率-损失”的静态评估转向“致灾机制-链式传导”的新型评估体系,以支撑铁路灾害的超前预测。主要贡献如下:①首次提出并

实践了铁路沿线自然灾害潜在风险路径的预测与分析方法。该方法基于文本挖掘技术,从灾害报告中提取关键词(如站点、致灾因子、后果等),构建了风险异构网络模型。②提出了一种多元路径度量方法,用于深入挖掘铁路灾害报告蕴含的丰富语义信息,量化风险传递关系。③设计了基于关联共现矩阵的风险路径预测模型。该模型在预测节点对时,综合考虑其多元路径及邻居节点信息,有效提升了稀疏连接场景下预测结果的准确性。

1 路径预测方法基础

复杂网络理论中的链路预测被广泛应用于各个领域来解决复杂系统中节点和边之间的相互作用、信息传递和演化规律等问题^[19],根据已知的网络节点以及结构信息等来预测网络中尚未产生连边的 2 个节点之间产生链接的可能性,即通过已知信息为一对未连接的节点对 v_i, v_j 分配分数 S_{ij} 。由于不同类型的网络具有不同的特性,其链路的连接反映不同的需求,这导致现有链路预测模型很难快速准确地评估现实网络中节点间的关系。因此,不同网络系统应根据其自身的相似性特征寻找合适的链路预测机制来满足其演化或传播规律。

针对不同目标的路径预测问题的现有通用方法可大致分为启发式方法(局部/全局算法)^[20]、图嵌入方法^[21]和图神经网络方法^[22]。本文选取表 1 所示的几种典型链路预测算法作为对比方法,并给出其中几种方法的分配分数 S_{ij} 的计算公式。其中: $\Gamma(i)$ 、 $\Gamma(j)$ 分别为节点 i 和 j 的邻居节点集,其公共邻居节点数量是 $|\Gamma(i) \cap \Gamma(j)|$, 2 个集合的所有邻居节点集是 $|\Gamma(i) \cup \Gamma(j)|$; k_i, k_j 分别为节点 i 和 j 的度; l_{ij}^+ 为拉普拉斯伪逆矩阵 L^+ 第 i 行第 j 列的元素; l_{ij} 为 $(I+L)^{-1}$ 中第 i 行第 j 列的元素, L 为拉普拉斯矩阵; q_{ij} 为随机步行者在稳态下位于节点 j 的概率。本文结合灾害链传递规律和定量研究思想,基于复杂网络理论和链路预测方法,构建铁路沿线自然灾害链网络事件的潜在风险路径预测模型。

2 铁路沿线自然灾害潜在风险路径预测模型

2.1 数据处理

铁路系统灾害数据多以文本记录的形式出现,如灾害跟踪报告、灾害库、因果分析或过程分析。这些文本数据中通常包含事件、时间、铁路站点、后果类型及严重程度等重要信息。若采用人工筛选和识

表 1 典型链路预测方法
Table 1 Typical link prediction methods

方法		分数计算公式	
启发式方法	局部算法	共同邻居(Common Neighbors, CN)	$S_{ij} = \Gamma(i) \cap \Gamma(j) $
		优先依附(Preferential Attachment, PA)	$S_{ij} = k_i k_j$
		Jaccard 系数(JC)	$S_{ij} = \frac{ \Gamma(i) \cap \Gamma(j) }{ \Gamma(i) \cup \Gamma(j) }$
	全局算法	平均通勤时间(Average Commute Time, ACT)	$S_{ij} = \frac{1}{l_{ii}^+ + l_{jj}^+ + 2l_{ij}^+}$
		矩阵森林(Matrix Forest, MF)	$S_{ij} = l_{ij}$
图嵌入方法	随机游走(Random Walk, RW)	$S_{ij} = q_{ij} + q_{ji}$	
	节点嵌入算法(Node2Vec)		
图神经网络方法	图采样聚合网络 (Graph Sample and Aggregate, GraphSAGE)		
	图卷积神经网络 (Graph Convolutional Network, GCN)		

别,无疑是一项费时费力的工作,尤其是在规模较大的铁路系统中^[23]。因此,本文采用改进的词袋(Bag-of-words, BoW)模型^[24],将铁路文本数据构建为网络模式,利用复杂网络的节点属性值替换传统词袋模型中的频率值,使向量的区分度更高,以此将铁路灾害报告中的句子分解为适当的词,并完成词分类。改进模型保留了 BoW 模型匹配并统计文档中每个元素的特点,并且通过网络化形式收集文本间的结构和语义信息,解决了 BoW 模型忽略的词语间相关性的问题,以此形成文档的向量表示。基于课题组提出的一种使用根本原因识别新词句(Root Cause Identification-new Word Sentence, RCI-NWS)关键字提取方法^[25],假设隐藏对象的重要性由文本数据中的关键词反映出来。构建文本的关键词共现网络,并在网络上生成随机游走路径。开发了一种基于卷积神经网络的隐藏信息检测模型(Convolution Neural Network for Hidden Information Detection, CHID)来识别与关键字相关的隐藏信息,使用余弦相似度来判断源文本数据中是否存在隐藏信息。从而提取出本文研究所需的铁路站点、灾害原因、后果及严重程度等关键信息。以站点、原因和后果关键词作为节点,当相同类型或不同类型的节点出现在同一起铁路灾害事件中并存在直接关联时,则形成边。同时将“轻度”“中度”“重度”等作为节点的“严重程度”属性,来一同构建铁路沿线自然灾害异质信息网络,并在网络中进行多元路径的搜索和关联共现矩阵的推导以完成铁路沿线自然灾害潜在风险路径预测。图 1 是一个铁路沿线自然灾害网络模拟。P 表示站点,C 表示原因,S 表示后

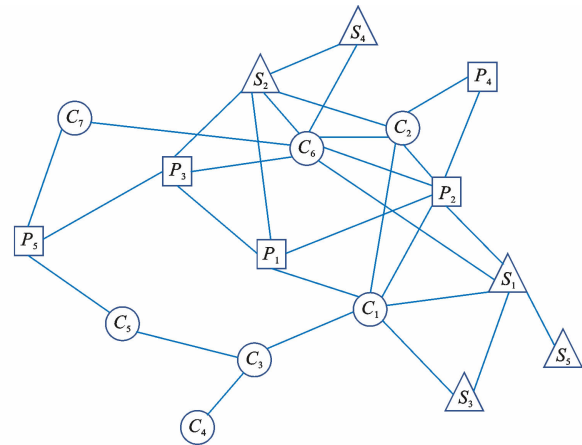


图 1 异构的灾害网络模拟

Fig. 1 Simulation of a heterogeneous disaster network

果。可以看出:模拟的网络是一个由 5 个站点节点、7 个原因节点和 5 个后果节点组成的典型异构网络,用不同下标编号表示不同的节点(例如, P_1 表示网络中第 1 个站点, C_2 表示造成铁路灾害风险的第 2 个原因, S_3 表示灾害带来的第 3 个后果等)。网络中存在 2 类构成长路径的基本元路径单元,一类是由 $P-P$ 、 $C-C$ 、 $S-S$ 组成的同质路径单元,另一类是由 $P-C$ 、 $P-S$ 、 $C-S$ 组成的异质路径单元。同质路径单元旨在挖掘风险在空间上的关联性与灾害链的演化模式。其中, $P-P$ 揭示地理上相邻或地质条件相似的路段之间存在风险关联, $C-C$ 揭示不同致灾原因之间的伴生或链式触发关系, $S-S$ 揭示不同灾害后果的并发或次生关系。异质路径单元是表达因果与风险传导逻辑的核心路径。 $P-C$ 表示“某个地理位置”与“某种致灾原因”之间的关联, $C-S$ 表示“某种原因”导致“某种后果”的因果关系, $P-S$ 表示“某个

地理位置”与“某种灾害后果”的直接关联,是一种高度概括的风险提示。

2.2 多元路径

铁路沿线自然灾害潜在风险路径预测问题是通过已知信息探究其可能在铁路沿线的风险传播路径,以确保提前采取应对措施,防止因多米诺效应造成的整条线路或区域路网的瘫痪^[26]。现有研究的解决方法通常是基于元路径,其基本思路仅是简单地计算被预测节点对之间的实际路径数。然而,这不符合铁路网络线路走向固定的特点,并且这种计算方法极有可能导致元路径上的现实语义特征被破坏。因此,本文在传统的元路径的基础上提出了多元路径的度量方法,可有效提取任意类型节点对之间的多种类型路径。假设 M 是异构网络 $G(V, E)$ 上的一条路径, $V = \{v_1, v_2, \dots, v_n\}$ 是节点类型集, $E = \{e_1, e_2, \dots, e_m\}$ 是链接类型集,代表节点类型之间的序列。不同的路径在异构网络中表达不同的语义。在本文中,一条路径代表了一次搜索所捕获的一个局部关联情境。它揭示了“哪些因素倾向于共同构成一个高风险场景”,而无需预设其间的因果方向。以图 1 展示的网络为例,节点对 (C_1, C_6) 之间的路径类型有: $C_1 \rightarrow C_2 \rightarrow C_6$ 为节点对 (C_1, C_6) 之间的一元路径, $C_1 \rightarrow P_2 \rightarrow C_6$ 和 $C_1 \rightarrow S_1 \rightarrow C_6$ 是二元路径, $C_1 \rightarrow P_1 \rightarrow S_2 \rightarrow C_6$ 则是三元路径。不同于传统的元路径方法,多元路径创新性地加入了异质节点类型,不仅能够挖掘大量路径实例,还可以实现跨语义层次的路径关联,解决了元路径方法挖掘结果“粒度混杂、难以解读”的问题。一元路径是由单一类型节点构成的序列(如 $P-P$ 、 $C-C$ 、 $S-S$),专门用于揭示风险在空间上的关联性 or 致灾因子/后果之间的伴生/演化规律;二元路径包含 2 种类型节点的序列(如 $P-C$ 、 $P-S$ 、 $C-S$),揭示了 2 类要素之间最核心直接的关联桥梁;三元路径包含全部 3 种类型节点的序列(如 $P-C-S$),代表了从“地理位置”到“致灾原因”再到“最终后果”的完整风险情景。多元路径描述节点之间通过各种节点类型和关系类型在元水平上的可达路径,反映了网络的拓扑结构和语义特征。多元路径的挖掘可以丰富网络关系,从而更准确地预测节点之间的关联性。

广度优先搜索(Breadth First Search, BFS)算法作为最常用的元路径搜索算法,可以搜索 2 个节点之间的最短路径,但对于异构网络这不能完全反映网络的特点。因此,本文在 BFS 算法的基础上增加了一些步骤来实现多元路径的搜索。利用 BFS

算法对原始网络进行搜索可得到包含:3 种节点类型的最短三元路径。对于同质路径,删除原始网络中与目标节点不同的其中一种类型的所有节点,并通过 BFS 算法对删减后的网络进行路径搜索可以得到其中一种二元路径;类似地,删除原始网络中与目标节点不同的另一种类型的所有节点,再次使用 BFS 算法对新网络进行路径搜索可以得到另一种二元路径;除此之外,同时删除原始网络中与目标节点不同的其他 2 种类型的所有节点,即只在同类节点所形成的网络中搜索形成一元路径。例:对图 1 中的 (C_1, C_6) 同质路径进行多元路径集搜索,可根据上述过程首先确定最短三元路径为 $C_1 \rightarrow P_1 \rightarrow S_2 \rightarrow C_6$ 和 $C_1 \rightarrow S_1 \rightarrow P_2 \rightarrow C_6$;在图 1 所示网络中删除 P 类型的所有节点,再次搜索得到包含 C 和 S 两种节点类型的最短二元路径 $C_1 \rightarrow S_1 \rightarrow C_6$,然后在图 1 所示网络中删除 S 类型的所有节点,可得到包含 C 和 P 两种节点类型的另一条最短二元路径 $C_1 \rightarrow P_2 \rightarrow C_6$;最后,将 P 和 S 两种类型的所有节点均删除,则得到只有 C 类型节点的最短一元路径 $C_1 \rightarrow C_2 \rightarrow C_6$;至此确定节点对 (C_1, C_6) 的多元路径集为 $\{C_1 \rightarrow C_2 \rightarrow C_6, C_1 \rightarrow P_2 \rightarrow C_6, C_1 \rightarrow S_1 \rightarrow C_6, C_1 \rightarrow P_1 \rightarrow S_2 \rightarrow C_6, C_1 \rightarrow S_1 \rightarrow P_2 \rightarrow C_6\}$ 。若是异质路径,只需要删除被预测节点对外的另一种类型的所有节点,得到最短二元路径;异质路径的多元路径集中无一元路径。例:对图 1 中的 (C_1, S_2) 异质路径进行多元路径搜索,首先由 BFS 算法得到最短三元路径 $C_1 \rightarrow P_1 \rightarrow S_2$;然后删除 P 类型节点,则得到最短二元路径 $C_1 \rightarrow C_2 \rightarrow S_2$;最后得到节点对 (C_1, S_2) 的多元路径集为 $\{C_1 \rightarrow C_2 \rightarrow S_2, C_1 \rightarrow P_1 \rightarrow S_2\}$ 。对于具有 3 种节点类型的异构网络,任意节点间的多元路径集中的元素不超过 4 个。

多元路径集中每条路径的类型、所经过的节点数以及边的权重都对预测结果有着不可忽视的影响。因此,在计算链接得分时,根据这些参数将多元路径信息量化为

$$P_{ij} = \frac{1}{n_{ij}(n_{ij} - 1)m} \cdot \sum_{y=1}^m \left[\frac{x_{ij}(y)}{y} \frac{\sum_{z=1}^{l_{ij,y}} k_{ij,y(z)}}{l_{ij,y}} \frac{\sum_{z=2}^{l_{ij,y}} k_{ij,y(z-1,z)}}{l_{ij,y} - 1} \right] \quad (1)$$

式中: P_{ij} 为节点 i, j 之间的多元路径量化值; n_{ij} 为节点 i, j 之间的多元路径数; m 为网络中共包含多元路径的条数; y 为该路径是节点 i, j 的 y 元路径, $y=1, 2$ 或 3 ; $x_{ij}(y)$ 为节点 i, j 之间是否存在 y 元

径,取值为 0 或 1; $l_{ij,y}$ 为节点 i,j 的 y 元路径所包含的节点数; $\sum_{z=1}^{l_{ij,y}} k_{ij,y(z)}$ 为节点 i,j 的 y 元路径所含节点的总和; $\sum_{z=2}^{l_{ij,y}} k_{ij,y(z-1,z)}$ 为节点 i,j 的 y 元路径所含边权的总和。

边权为数据中 2 个关键词同时出现在同一灾害事件的数量。值得注意的是,边权是经过归一化处理的。

2.3 关联共现矩阵

Pech 等^[27] 通过矩阵实现了链路预测,无论网

$$\mathbf{M} = \begin{matrix} & \begin{matrix} C_1 & C_2 & C_3 & C_6 & C_7 & P_1 & P_2 & P_3 & P_4 & S_1 & S_2 & S_3 & S_4 & S_5 \end{matrix} \\ \begin{matrix} C_1 \\ C_2 \\ C_3 \\ C_6 \\ C_7 \\ P_1 \\ P_2 \\ P_3 \\ S_1 \\ S_2 \\ S_3 \\ S_4 \end{matrix} & \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \end{matrix} \quad (2)$$

一个低秩的矩阵表明节点之间有很强的关联性。因此,矩阵的秩被作为判断 2 个节点之间是否存在联系的指标之一。若用 $r(\mathbf{M}_{ij})$ 来表示关联共现矩阵 \mathbf{M}_{ij} 的秩,则节点 i 和 j 之间的关联共现系数 C_{ij} 为

$$C_{ij} = \frac{1}{e^{r(\mathbf{M}_{ij})} r(\mathbf{M}_{ij})} \quad (3)$$

基于以上分析,对多元路径和关联共现系数 2 个指标进行归一化处理后,本文将加权平均值作为判断潜在风险路径可能性的重要依据。因此,风险路径预测(Prediction of Risk Path, PRP)模型得分 S_{ij} 为

$$S_{ij} = \frac{P_{ij} + C_{ij}}{2} \quad (4)$$

3 案例分析

3.1 高原地区铁路沿线灾害网络的构建

中国高原面积广阔,占国土陆地总面积的 22.3%,且地质地貌复杂,沿线板块构造活跃,活动

网络密集或是稀疏低秩,都表现出很好的预测性能。在自然语言处理领域也提出用共现矩阵来表达 2 个词之间的共现事件,这都与本文的预测需求高度匹配。因此,本文提出基于被预测节点对其关联节点(包括多元路径上的节点和邻居节点等)的关联共现矩阵来实现灾害网络的潜在风险预测。首先将多元路径集中的所有节点及其邻居节点提取出来作为矩阵的行,然后以 2 个被预测节点及其邻居节点为列,这些节点之间如若存在连接关系则被填充为 1,否则为 0,从而形成节点对 (i,j) 的关联共现矩阵。综上,以图 1 示例网络为例,其节点对 (C_1,C_6) 的共现矩阵 \mathbf{M} 为

断裂密集,地震烈度强烈,地形变化显著,自然灾害频繁。本文在“国家减灾网”获得了 2014 年 4 月至 2020 年 9 月的 3 650 条有关全国范围的自然灾害原始文本数据(表 2),其中有 886 条自然灾害事件发生在高原铁路沿线地区或对高原铁路沿线地区造成影响。经数据处理得到 264 个不同类型实体(即节点)的集合,包括 209 个铁路站点节点类型、18 个灾害原因节点类型、37 个灾害后果节点类型,对于单条报告,将其内部识别出的所有节点进行两两连接,生成一个全连接图;每条连接的类型由该节点对的类型决定。遍历所有历史报告,累加每种链接类型出现的次数,即该种链接在全数据集中出现的总频次。将每种链接类型的样本数量记录在表 3 中。由此,一个异构加权高原铁路沿线自然灾害网络被构建,如图 2(a)。为了便于观察网络的细节,将图 2(a)进行放大处理并截取其中一部分展示在图 2(b)。其中,蓝色节点表示铁路站点,绿色节点表示原因,黄色节点表示结果。节点大小表示节点在数据中出现的数量。边的粗细反映节点间的关联强度,由两

表2 高原铁路沿线部分自然灾害原始数据示例

Table 2 Example of original data on some natural disasters along the plateau railway

事件 ID	事件类型	公告日期	影响地区	灾情概述(节选)
HL2014000035	洪涝	2014-07-22	四川省绵阳、德阳、乐山	7月19日以来,部分地区出现大雨、暴雨过程,引发洪涝灾害,截至7月22日统计,造成绵阳、德阳、乐山3市6个县(市、区)13万人受灾,100余间房屋倒塌,700余间损坏;直接经济损失6800余万元
DF2014000031	大风	2014-07-27	西藏自治区拉萨市林周县、昌都地区察雅县、阿里地区措勤县	7月25日,拉萨市林周县、昌都地区察雅县、阿里地区措勤县遭受风雹灾害,2人因雷击死亡
DZA201400004	地震	2014-08-05	四川省凉山彝族自治州宁南县、金阳县	8月3日16:30分,云南省昭通市鲁甸县发生6.5级地震,震源深度12 km。截至8月5日初步统计,地震造成四川省凉山彝族自治州宁南县、金阳县1.16万人受灾,近500余间房屋严重损坏,2000余间一般损坏
NSL201400004	泥石流	2014-08-09	西藏自治区昌都县卡若镇	8月7日,昌都地区昌都县卡若镇发生泥石流灾害,造成1人死亡

表3 每种链接类型的样本数量

Table 3 Number of samples for each link type

路径类型	P-P	C-C	S-S	P-C	P-S	C-S	整体
样本数量	621	44	118	394	840	159	2176

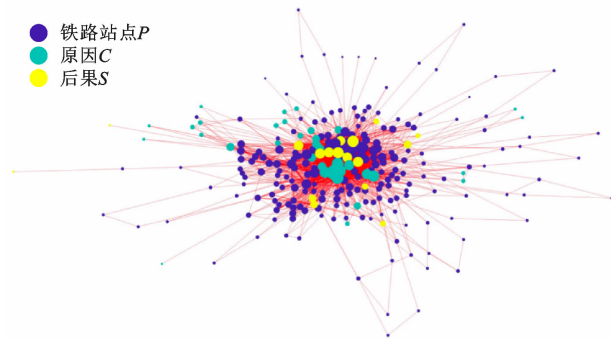
节点同时出现在同一灾害数据中的数量决定。节点布局不反映实际地理坐标,旨在展示风险路径的逻辑关联。

3.2 对比试验

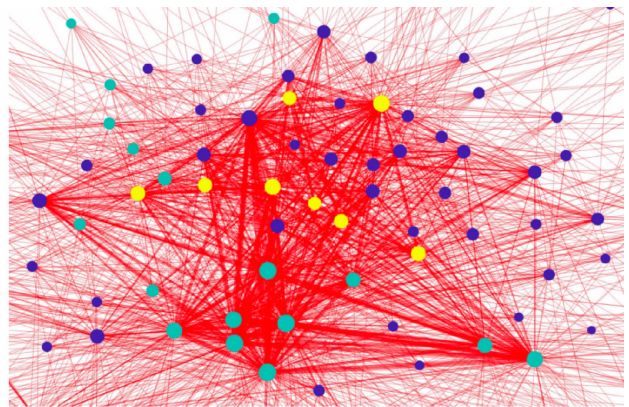
为准确评估结果的稳健型和泛化能力,采用分层交叉验证法,针对铁路灾害数据特有的站点层、原因层和后果层3级层次结构,设计了多级分层数据划分策略。以整合式分层划分为其逻辑核心,并非对每层独立划分,而是以站点层为一级划分单元,原因层和后果层作为嵌套属性,构建分层划分框架。具体实施步骤:将209个铁路站点按地理分区划分

为10组(每组20或21站),确保组间地质风险特征差异最小化,组内包含多样灾害类型;在每个站点组内按灾害原因、后果及等级分层,以保持训练/验证集的原因-后果分布一致;由此形成10个互斥子集,选择其中的8个子集作为训练集,剩下的2个子集作为测试集,重复10轮。

为了便于对比预测方法的准确性,常用指标有接收者操作特征曲线(Receiver Operating Characteristic Curve, ROC)、曲线下面积(Area Under Curve, AUC)、准确率和F1分数。ROC曲线是一个坐标图,以特异性为横轴,敏感度为纵轴,在特定的刺激条件下,由于判断标准不同而得到不同的结果;AUC是指随机选择的缺失环节比随机选择的不存在环节得分高的概率^[28]。AUC值 P_c 为



(a) 高原铁路沿线自然灾害异构网络整体



(b) 高原铁路沿线自然灾害异构网络局部放大

图2 高原铁路沿线自然灾害异构网络

Fig. 2 Heterogeneous network of natural disasters along the plateau railway

$$P_c = \frac{n' + 0.5n''}{n} \quad (5)$$

式中: n 为试验总数; n' 为得分较高的缺失链接的数量; n'' 为与不存在的链接得分相同的缺失链接数量。

准确率是指在一定试验条件下的多个测定值中,满足限定条件的测定值所占的比例,常用符合率来表示。F1 分数是统计学中用于评估二分类模型性能的指标^[29],通过计算精确率 P_r 与召回率 R_c 的调和平均数实现,特别适用于数据不均衡场景。其值 F_1 为

$$F_1 = 2 \frac{P_r R_c}{P_r + R_c} \quad (6)$$

为全面、严谨地评估 PRP 模型的预测性能与有效性,本文将其与第 1.1 节所述的典型方法进行了多维度对比分析,包括 ROC 曲线可视化(图 3~5)、关键量化指标(AUC、准确率、F1 分数,表 4)以及

统计显著性检验(t 检验,表 5、图 6)。如图 3(同质路径)、图 4(异质路径)、图 5(混合路径)所示,PRP 模型的 ROC 曲线普遍最接近坐标系左上角,直观表明其在区分风险路径与非风险路径上具有最优的排序能力。表 4 数据作为辅助严格证实了上述观察,即 PRP 在所有测试路径类型(P - P 、 C - C 、 S - S 、 P - C 、 P - S 、 C - S 以及混合路径)下的 AUC 值均显著高于其他所有对比方法。这无可辩驳地证明了 PRP 模型在核心判别指标(AUC)上具有全面且显著的优越性。PRP 模型在异质路径(P - C 、 P - S 、 C - S)和混合路径上的 AUC 优势最为突出,例如,在 C - S 路径中,PRP 模型的 $P_c = 0.943$,次优模型 GCN 模型的 $P_c = 0.915$;所有混合路径下 PRP 模型的 $P_c = 0.938$,次优模型 CN 模型的 $P_c = 0.928$,直接体现了其模型设计对复杂灾害链风险传递模式(即文本数据间的信息关联)的卓越捕捉能力。

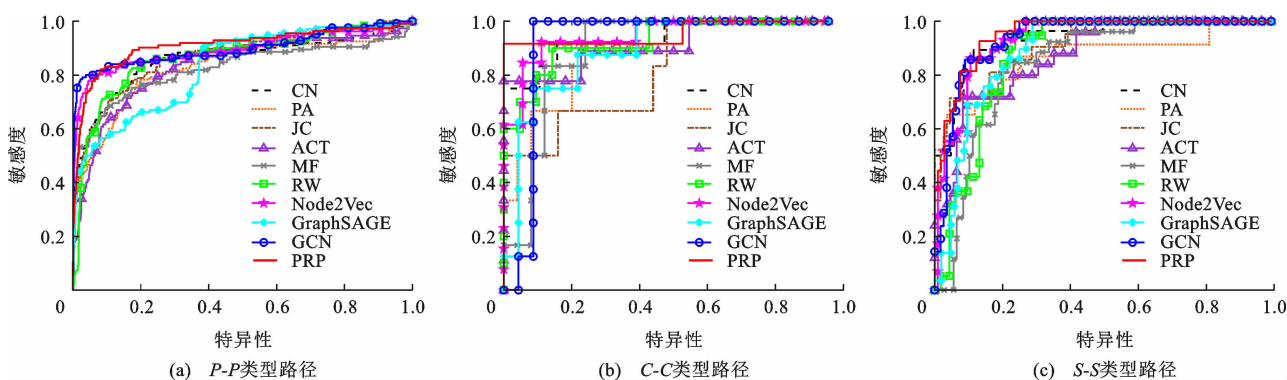


图 3 不同方法在同质路径中的 ROC 曲线比较

Fig. 3 ROC curve comparison for different methods in homogeneous paths

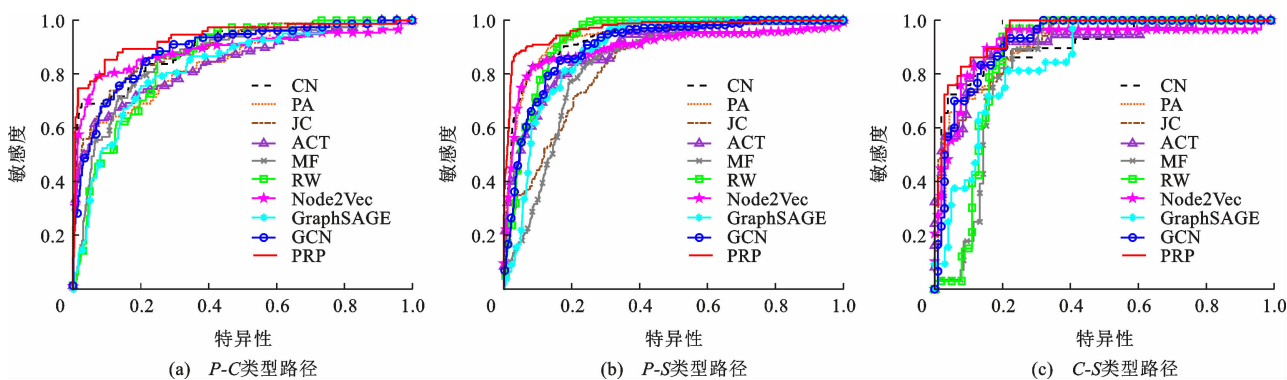


图 4 不同方法在异质路径中的 ROC 曲线

Fig. 4 ROC curve comparison for different methods in heterogeneous paths

由表 4 可知:PRP 模型的准确率在 P - P 、 S - S 、 P - C 、 P - S 路径及 F1 分数在 C - C 、 S - S 、 P - S 路径上并非单项最高,例如 P - P 路径下,GCN 准确率为 97.976%,大于 PRP 的 97.884%; P - C 路径下,ACT 准确率为 92.420%,大于 PRP 的 90.957%。然而,PRP 模型与最优准确率值和最佳 F1 分数的差距通

常很小(如 P - P 差 0.1%, S - S 差 1.5%, C - C 、 P - S 均差 0.002)。PRP 模型的准确率和 F1 分数在所有路径类型中均稳定位于所有模型的前列,且在 C - C 、 C - S 路径和混合路径上的准确率为最优,在 P - P 、 P - C 、 C - S 和混合路径上的 F1 分数为最优。综上所述,尽管在部分路径类型中 PRP 模型的准确率和

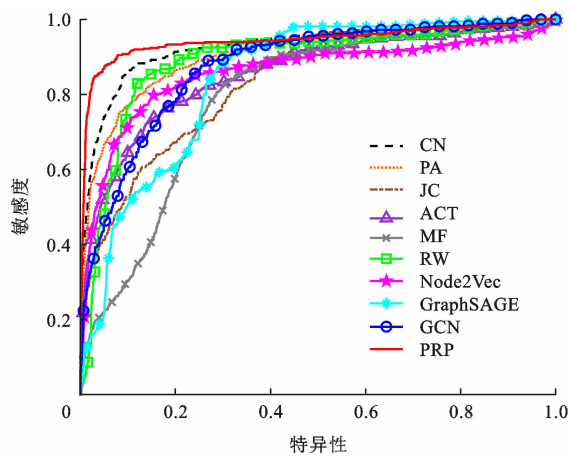


图 5 不同方法在所有混合路径中的 ROC 曲线比较
Fig. 5 ROC curve comparison for different methods in all mixed paths

F1 分数不是最优的,但结合其在更鲁棒、更核心的 AUC 指标上全面且显著领先,以及在准确率和 F1 分数上整体优异且差距微小的表现,PRP 模型的综合预测性能被确认为最优。准确率和 F1 分数的波动可能受阈值敏感性和特定路径数据分布影响,不影响 AUC 揭示的整体判别能力优势。

为定量评估 PRP 模型与各对比模型性能差异的统计显著性,进行了配对样本 t 检验。结果如表 5 所示,表中“均值差”定义为 PRP 模型性能指标减去对比模型性能指标,以 $P < 0.001$ 作为统计显著性的标准。总体而言,PRP 模型在大多数比较中展现出统计显著的优越性。具体而言,在与 CN、PA、JC、ACT、MF、RW、GraphSAGE、GCN 方法的比较中,均值差均为正且 $P < 0.001$,这表明 PRP 模型的

表 4 不同方法在不同路径类型中的 AUC 与准确率

Table 4 AUC and accuracy of different methods in different path types

方法	指标	P-P	C-C	S-S	P-C	P-S	C-S	混合路径
CN	AUC	0.858	0.886	0.927	0.901	0.937	0.904	0.928
	准确率/%	94.502	90.323	89.474	91.223	88.688	90.226	93.879
	F1 分数	0.873	0.875	0.874	0.841	0.870	0.879	0.892
PA	AUC	0.829	0.867	0.859	0.845	0.936	0.904	0.906
	准确率/%	97.608	87.097	91.729	89.096	89.465	86.466	93.692
	F1 分数	0.837	0.843	0.849	0.820	0.833	0.871	0.880
JC	AUC	0.870	0.780	0.903	0.894	0.850	0.906	0.839
	准确率/%	97.032	90.323	90.977	91.755	91.144	87.218	94.397
	F1 分数	0.860	0.864	0.875	0.863	0.825	0.854	0.884
ACT	AUC	0.835	0.869	0.866	0.839	0.891	0.899	0.859
	准确率/%	97.493	90.323	87.218	92.420	89.399	87.218	93.317
	F1 分数	0.852	0.845	0.865	0.880	0.888	0.865	0.890
MF	AUC	0.823	0.853	0.833	0.859	0.835	0.835	0.792
	准确率/%	97.562	83.871	80.451	88.697	88.623	88.948	93.692
	F1 分数	0.869	0.834	0.870	0.849	0.853	0.878	0.881
RW	AUC	0.866	0.881	0.865	0.851	0.930	0.854	0.891
	准确率/%	94.430	87.097	85.714	88.963	90.109	86.466	93.042
	F1 分数	0.891	0.853	0.893	0.892	0.850	0.889	0.886
Node2Vec	AUC	0.894	0.893	0.923	0.887	0.900	0.904	0.859
	准确率/%	97.125	90.323	87.970	88.564	90.045	88.722	93.821
	F1 分数	0.885	0.891	0.880	0.874	0.900	0.880	0.896
GraphSAGE	AUC	0.830	0.848	0.881	0.826	0.887	0.842	0.843
	准确率/%	96.940	87.419	82.707	87.766	88.559	90.451	93.184
	F1 分数	0.892	0.866	0.866	0.882	0.882	0.874	0.894
GCN	AUC	0.898	0.875	0.929	0.888	0.905	0.915	0.875
	准确率/%	97.976	93.545	90.226	91.755	90.304	87.970	94.196
	F1 分数	0.861	0.863	0.889	0.884	0.901	0.891	0.891
PRP	AUC	0.909	0.918	0.942	0.928	0.968	0.943	0.938
	准确率/%	97.884	93.548	89.474	90.957	89.528	90.977	94.873
	F1 分数	0.896	0.889	0.885	0.897	0.899	0.897	0.899

表 5 PRP 模型与其他对比方法的 t 检验结果Table 5 The t -test results of PRP model compared with other methods

对比方法	均值差	t 统计量	P 值	95%置信区间	结论
PRP-CN	0.013 7	25.973 7	<0.001	[0.012 7,0.014 8]	显著不同
PRP-PA	0.008 0	14.844 4	<0.001	[0.007 0,0.009 1]	显著不同
PRP-JC	0.078 6	52.807 8	<0.001	[0.075 7,0.081 5]	显著不同
PRP-ACT	0.038 0	55.874 2	<0.001	[0.036 7,0.039 4]	显著不同
PRP-MF	0.063 0	129.819 1	<0.001	[0.062 0,0.063 9]	显著不同
PRP-RW	0.233 1	67.683 0	<0.001	[0.226 3,0.239 8]	显著不同
PRP-Node2Vec	-0.004 4	-14.200 8	<0.001	[-0.005 0,-0.003 8]	显著不同
PRP-GraphSAGE	0.250 3	95.019 1	<0.001	[0.245 1,0.255 5]	显著不同
PRP-GCN	0.115 7	62.523 8	<0.001	[0.112 1,0.119 3]	显著不同

性能显著优于这些对比方法。然而,在与 Node2Vec 的比较中,结果呈现出不同的模式。均值差为较小的负值(PRP-Node2Vec 为 -0.0044),且其 $P < 0.001$ 。这表明,尽管 PRP 在多数比较中占优,但在此特定比较中,Node2Vec 的表现略优于 PRP 模型。为了更直观地展示性能差异的分布情况与稳定性,绘制了 PRP 模型与各对比方法性能指标的箱线图,并叠加了所有数据点的散点图,如图 6 所示,箱线图的结果与 t 检验的结论高度一致。

3.3 铁路沿线灾害风险路径结果评价

使用 PRP 模型预测高原铁路沿线自然灾害间的风险传播路径。图 7 直观展示了预测结果的可视化。其中,用不同粗细和不同颜色的实线表示不同传播路径的连接概率,深蓝色细线表示低概率传播路径,其在 PRP 模型的预测结果中较低 ($0 < S_{ij} \leq 0.3$);浅蓝色、绿色细线和黄色粗线是 $0.3 < S_{ij} < 0.7$ 的传播路径;橙色和红色粗线是 $S_{ij} \geq 0.7$ 的传播路径。

表 6 展示了 PRP 模型预测出的高风险路径示例及其评分(风险值)。这些案例不仅验证了模型的预测能力,更揭示了高原铁路沿线潜在的复杂灾害链风险模式,为精准防控提供了决策依据。

3.3.1 高频同质风险路径分析

(1)空间节点风险关联,识别跨区域威胁。在 $P-P$ 类型路径中 PRP 模型预测的高风险节点对分别是贡觉-芒康(风险值为 0.917)、江达-贡觉(风险值为 0.911)、昌都-甘孜(风险值为 0.910),这些被预测为高概率互相影响站点。由此可知,地理邻近性是风险传导的关键因素。昌都与甘孜位于省界区域,模型也成功捕捉到跨行政区风险传递的可能性。当灾害发生在这些高关联区域时,必须立即启动邻近区域及跨省区域的应急预案,防止灾害蔓延造成

更大范围损失。这体现了 PRP 模型在空间风险协同防控上的预警能力。

(2)灾害链触发关系,揭示次生灾害机制。在 $C-C$ 类型路径中 PRP 模型预测降雪 \rightarrow 泥石流(风险值为 0.951)、山洪 \rightarrow 滑坡(风险值为 0.940)、风雹 \rightarrow 山体崩塌(风险值为 0.931)具有高预测分数。这直接反映了原生灾害诱发次生灾害的经典链式反应为严重降雪 \rightarrow 融雪/饱和土壤 \rightarrow 泥石流、山洪 \rightarrow 冲刷/侵蚀坡体 \rightarrow 滑坡、风雹 \rightarrow 强风/冲击 \rightarrow 山体崩塌。模型预测为灾害链的主动阻断提供关键时间窗口,例如:发生降雪(原生灾害)时,应重点监测和防范泥石流(次生灾害);处理山洪时,必须同步评估滑坡风险;风雹天气时,需提前加固易崩塌山体。PRP 模型通过量化这种链式关系,提升了灾害链风险的前瞻性管理能力。

(3)灾害后果关联,预判综合损失场景。在 $S-S$ 类型路径中 PRP 模型预测的高风险后果对为人员受灾 \leftrightarrow 电力设备损毁(风险值为 0.945)、房屋倒损 \leftrightarrow 电力设备损毁(风险值为 0.906)、房屋倒损 \leftrightarrow 供水设备损毁(风险值为 0.902)。模型捕捉到符合物理规律的后果耦合:房屋倒塌常连带摧毁电线杆/水管(空间共毁);电力中断直接影响居民生活(功能依赖)。当灾害导致房屋倒损时,救援需同步部署电力、供水抢修力量,并优先疏散安置受影响人员,防止次生灾害人员伤亡。这体现了 PRP 模型在多维度损失协同评估与应急资源优化配置上的应用价值。

3.3.2 高频异质风险路径分析

(1)地点-灾害关联,定位高风险区域与灾害。PRP 模型预测的高风险节点-灾害对($P-C$ 类型路径)是甘孜-风雹(风险值为 0.918)、雅安-地震(风险值为 0.916)、眉山-洪涝(风险值为 0.911)。甘孜邻

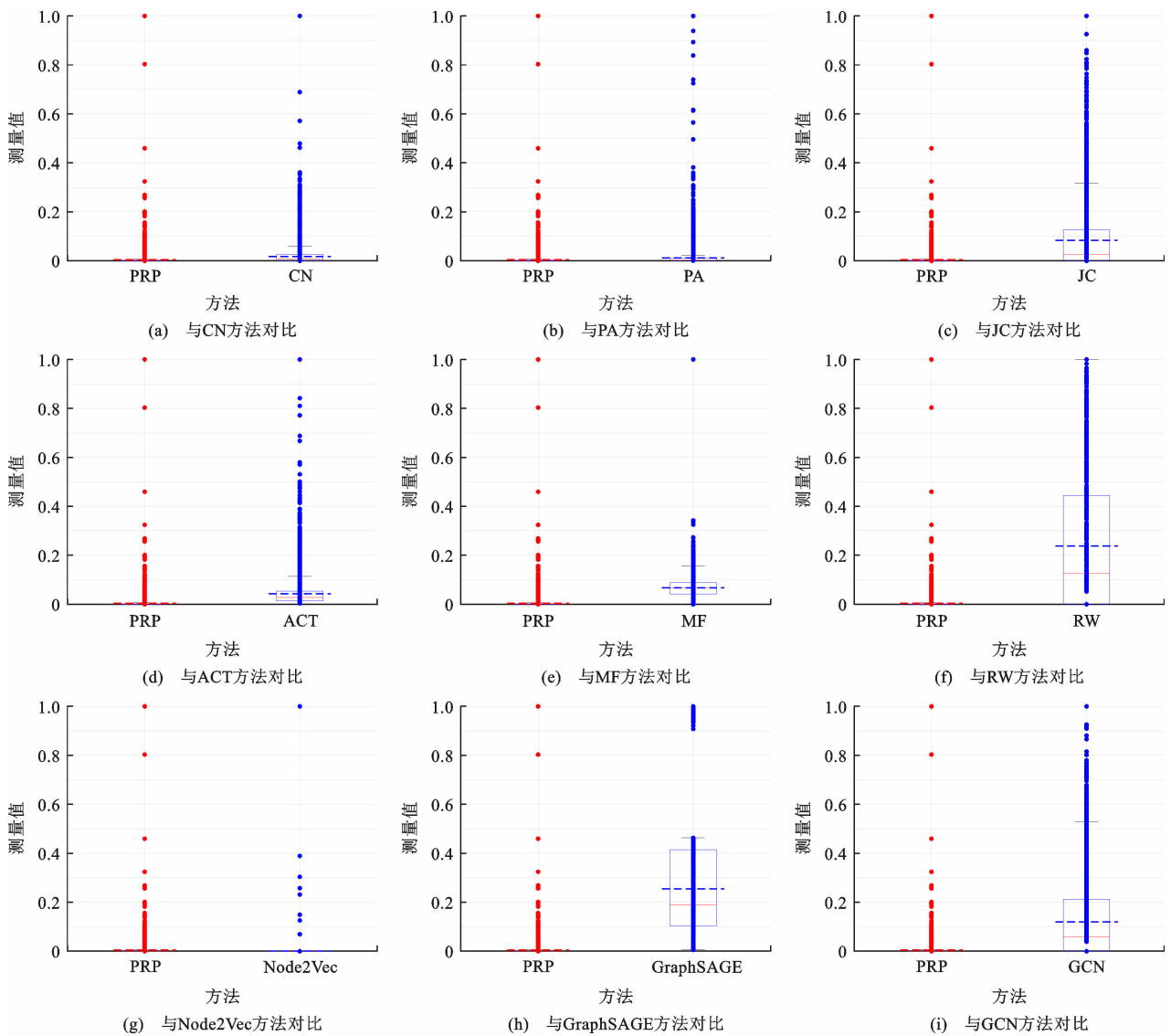


图6 PRP模型与其他对比方法t检验的箱线图与数据点叠加图

Fig. 6 Box plot and data point overlay for t-test of PRP model and other comparative methods

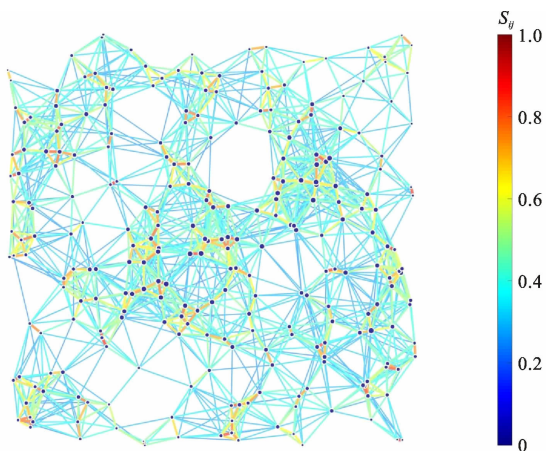


图7 PRP模型对高原铁路沿线自然灾害风险路径预测结果的可视化

Fig. 7 Visualization of PRP model for predicting natural disaster risk paths along the plateau railway

近风雹多发的昌都、凉山(空间传播风险),雅安位于深源地震活跃的板块交界带(地质特征),眉山周边成都、乐山、雅安洪涝频发(区域水文风险)。模型精准识别了特定地区的最高发/最威胁灾害类型,指导区域差异化防控重点,即甘孜需强化风雹预警防御,雅安需提升抗震设防与应急能力,眉山需重点加强防洪排涝工程。

(2)地点-后果关联与灾害-后果关联,预判损失热点。PRP模型预测出的P-S类型路径的高风险关联节点对是自贡-经济损失(风险值为0.955,原因是邻近灾害多发区眉山、雅安、广安,灾害易蔓延至此造成损失)、香格里拉-人员受灾(风险值为0.953,原因是受高原高频灾害波及)、洛隆-房屋倒塌(风险值为0.951)。C-S类型路径的高风险节点对为强对流天气→人员伤亡(风险值为0.929)、大

表 6 不同路径类型的预测结果
Table 6 Prediction results of different path types

路径类型	预测结果	分数
P-P	贡觉-芒康	0.917
	江达-贡觉	0.911
	昌都-甘孜	0.910
C-C	降雪-泥石流	0.951
	山洪-滑坡	0.940
	风雹-山体崩塌	0.931
S-S	人员受灾-电力设备损毁	0.945
	房屋倒损-电力设备损毁	0.906
	房屋倒损-供水设备损毁	0.902
P-C	甘孜-风雹	0.918
	雅安-地震	0.916
	眉山-洪涝	0.911
P-S	自贡-经济损失	0.955
	香格里拉-人员受灾	0.953
	洛隆-房屋倒损	0.951
C-S	强对流天气-人员伤亡	0.929
	大风-道路损毁	0.913
	滑坡-桥梁损毁	0.908

风/滑坡→道路损毁/桥梁损毁(风险值分别为 0.913、0.908)。数据证实存在强对流引发滑坡/泥石流(二次灾害)导致人员伤亡和基础设施损毁的案例。模型揭示了特定地点最易遭受的损失类型以及特定灾害最可能导致的严重后果。这提示:在自贡,需加强经济韧性建设与灾害保险;在香格里拉,需完

善人员疏散避难体系;在洛隆,需提升房屋抗震抗灾标准;面对强对流天气时,需提前转移高危区人员并巡查交通干线;发生滑坡时,需快速评估和抢通道路/桥梁。

3.3.3 链式连锁传导风险路径分析

基于模型输出的任意两点间风险值,可构建完整的潜在灾害风险传播路径。为了量化这类链式风险并对其机理进行深入分析,引入路径风险传导强度 R_{path} 指标,计算每条关键路径的路径风险值。这个值不是简单的静态边得分 S_{ij} 的叠加,而是通过考虑路径上节点间的发生概率或风险放大效应来计算。整条路径的强度 R_{path} 将由链式法则计算,例如:对于一条路径 $A \rightarrow B \rightarrow C$, $R_{path} = S_A P(B|A) \cdot P(C|B)w_C$ 。其中: S_A 为初始风险; $P(B|A)$ 为基于网络数据计算时给定 A 发生时 B 发生的概率,即 S_{AB} ; w_C 为节点 C 的严重程度权重。这能直观展示风险从 A 到 C 的传导与放大过程。

以预测出的 2 条高频风险链为例深入解构其传导过程,图 8(a)、(b)分别展示了路径 1、2 的风险传导过程模拟,路径 1 为芒康(P_1)→降雪(C_1)→香格里拉(P_2)→人员受灾(S_1)→电力设备损毁(S_2),路径 2 为江达(P_2)→贡觉(P_4)→强对流天气(C_2)→人员伤亡(S_3)。图 8 中蓝色节点表示铁路站点 P ,绿色代表原因节点 C ,黄色节点表示后果 S 。节点的大小表示节点初始风险值的大小,箭头的粗细反映了 2 个节点之间的发生概率即预测分数。

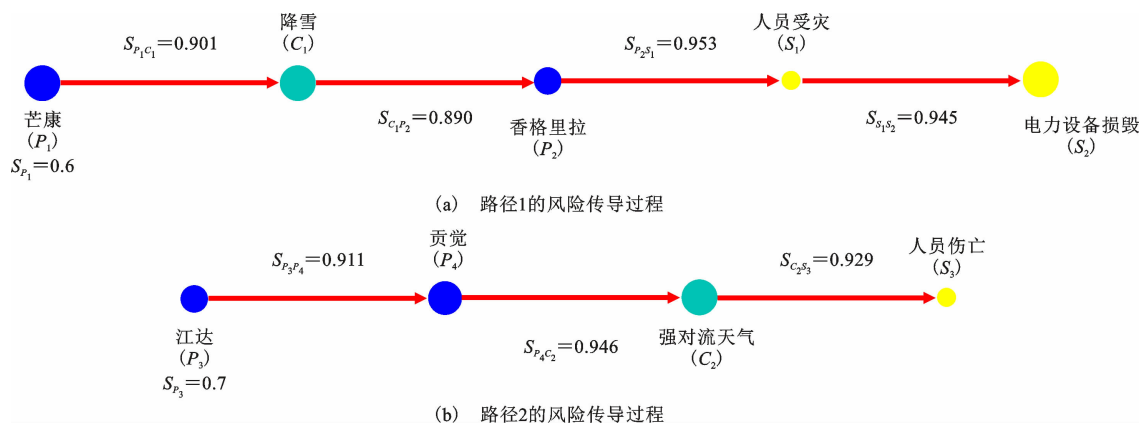


图 8 风险传导路径可视化

Fig. 8 Visual of risk transmission path

表 7 精确地展示风险传导每一步的量化计算过程,为图 8 风险传导路径可视化提供数据支撑。严重程度权重 w 根据数据提取阶段的“轻度”“中度”“重度”关键词分别设定为 1.0、1.5、2.0。以“芒康

(P_1)→降雪(C_1)→香格里拉(P_2)→人员受灾(S_1)→电力设备损毁(S_2)”这条路径为例,深入分析其链式传导过程。

(1)初始风险:节点 P_1 的固有风险得分为 $S_{P_1} = 0.6$ 。

表7 风险传导过程量化分析

Table 7 Quantitative analysis of risk transmission process

路径	传导步骤	当前节点(风险值)	预测分数	传导至下一节点风险值	结论
1	起点	$P_1(S_{P_1}=0.6)$			风险放大:灾害 后果严重
	第1跳	$P_1(S_{P_1}=0.6)$	$S_{P_1C_1}=0.901$	$0.6 \times 0.901=0.541$	
	第2跳	$C_1(S_{C_1}=0.541)$	$S_{C_1P_2}=0.890$	$0.541 \times 0.890=0.481$	
	第3跳	$P_2(S_{P_2}=0.481)$	$S_{P_2S_1}=0.953$	$0.481 \times 0.953=0.458$	
	第4跳	$S_1(S_{S_1}=0.458)$	$S_{S_1S_2}=0.945$	$0.458 \times 0.945=0.433$	
	后果与放大	$S_2(S_{S_2}=0.433)$	$w_{S_2}=2.0$	$0.433 \times 2.0=0.866$	
2	起点	$P_3(S_{P_3}=0.7)$			风险放大:灾害 后果严重
	第1跳	$P_3(S_{P_3}=0.7)$	$S_{P_3P_4}=0.911$	$0.7 \times 0.911=0.638$	
	第2跳	$P_4(S_{P_4}=0.638)$	$S_{P_4C_2}=0.946$	$0.638 \times 0.946=0.603$	
	第3跳	$C_2(S_{C_2}=0.603)$	$S_{C_2S_3}=0.929$	$0.603 \times 0.929=0.560$	
	后果与放大	$S_3(S_{S_3}=0.560)$	$w_{S_3}=1.5$	$0.560 \times 1.5=0.841$	

(2)第1跳($P_1 \rightarrow C_1$):该站点对降雪高度敏感,根据PRP模型计算出的 P_1 、 C_1 的得分 $S_{P_1C_1}=0.901$,传导至此的风险值为0.541。

(3)第2跳($C_1 \rightarrow P_2$):计算得出 $S_{C_1P_2}=0.890$,降雪扩散到此站点的风险值为0.481。

(4)第3跳($P_2 \rightarrow S_1$):发生人员受灾后果,由 $S_{P_2S_1}=0.953$ 可得此步风险值为0.458。

(5)第4跳($S_1 \rightarrow S_2$):电力设备损毁, $S_{S_1S_2}=0.945$,风险传导至此为0.433。

(6)最终后果:由数据可得“电力设备损毁”的严重程度为重度,即 $w_{S_2}=2.0$,风险值跃升至0.866。

这些推演出的高概率、多环节灾害链路径是PRP模型最核心的创新应用价值。其超越了传统单灾种或单环节分析,实现了灾害从源头(地点/灾害)到终端损失(后果)的完整传递链条可视化。在识别到链首风险时,可针对性地在关键传递节点提前部署措施,从根本上阻断或减轻级联失效。为制定“靶向性”应急预案、优化资源投放提供直接、量化的依据。

4 结 语

(1)构建了基于文本挖掘技术的铁路沿线自然灾害风险异构网络。基于改进的词袋模型和关键字提取方法(RCI-NWS)从2014年4月至2020年9月的886条有关高原铁路沿线地区或对高原铁路沿线地区造成影响的自然灾害原始文本数据中,分解并提取出了209个铁路站点、18个灾害原因和37个灾害后果关键词;以得到的264个关键词为节点,以发生在同一事件中存在直接关联关系的2176组节点对为边,形成了一个涵盖3种节点类型和6种

边类型的铁路沿线自然灾害风险异构网络。

(2)提出了一种新的铁路沿线自然灾害网络的潜在风险路径预测模型。应用改进的BFS算法得到了一个包括一元路径、二元路径和三元路径在内的元路径集合,实现了多元路径的综合度量,可有效提取任意类型节点对之间的多种路径,从不同角度反映网络特征,避免了元路径上的现实语义特征被破坏;此外,本文提出了包含被预测节点对、多元路径集中的节点和邻居节点等要素在内的关联共现矩阵,从多角度实现了灾害网络中潜在风险路径的综合预测。

(3)以高原铁路沿线自然灾害网络为例的试验表明,本文提出的潜在风险路径预测(PRPP)模型的ROC曲线和AUC值都优于其他方法。结果显示,贡觉与芒康、江达与贡觉、昌都与甘孜是最有可能扩散灾害风险的站点;降雪与泥石流、山洪与滑坡、风雹与山体崩塌具有较高的分数;后果之间较强的潜在影响关系出现在人员受灾与电力设备损毁、房屋倒损与电力设备损毁、房屋倒损与供水设备损毁之间。此外,本文还预测了交叉型链接之间的潜在关系。在P-C中,甘孜受到风雹灾害影响、雅安受到地震灾害影响和眉山受到洪涝灾害影响的概率很大。对于P-S来说,自贡与经济损失之间具有很高得分,香格里拉地区存在人员受灾风险,洛隆地区易发生房屋倒损的后果。在C-S的预测结果中,强对流天气导致人员伤亡的灾害是高概率事件。最后,结合不同路径类型的预测结果,本文给出了2条潜在灾害风险路径及其链式连锁传导过程。

(4)以上研究是基于高原铁路沿线自然灾害事件数据进行的,研究结果具有一定的局限性,模型的

普适性后续还需在其他地区和不同地理范围内应用后进行修正。未来的研究工作需在不同城市或全国范围内扩展该方法,甚至收集世界范围内的铁路灾害事件,探索具有重要指导意义的潜在风险路径,以及不同国家铁路灾害事件中潜在风险路径的共性和差异。此外,本文正在推动与铁路运营商、建设单位的合作,以获得更多非公开数据来支持此结论。例如,将研究对象由自然灾害扩充至工程建设中的其他灾害事件。

参考文献:

References:

- [1] LU C F, CAI C X. Challenges and countermeasures for construction safety during the Sichuan-Tibet Railway project [J]. *Engineering*, 2019, 5: 833-838.
- [2] 李远远,梅红波,任晓杰,等.基于确定性系数和支持向量机的地质灾害易发性评价[J].*地球信息科学学报*,2018,20(12):1699-1709.
LI Yuan-yuan, MEI Hong-bo, REN Xiao-jie, et al. Geological disaster susceptibility evaluation based on certainty factor and support vector machine[J]. *Geo-Information Science*, 2018, 20(12): 1699-1709.
- [3] 夏溪蔓,孟学雷,程晓脚,等.基于内外双视角的高速铁路风险预测[J].*铁道科学与工程学报*,2025,22(7):2921-2931.
XIA Xi-man, MENG Xue-lei, CHENG Xiao-qing, et al. High-speed railway risk prediction based on internal and external perspectives [J]. *Journal of Railway Science and Engineering*, 2025, 22(7): 2921-2931.
- [4] 王浙涛,陈如海,魏国俊.地质灾害危险性评估在铁路建设中的重要性[J].*甘肃科学学报*,2003,15:72-75.
WANG Zhe-tao, CHEN Ru-hai, WEI Guo-jun. Significance of risk assessment of geological hazards in railway construction [J]. *Journal of Gansu Sciences*, 2003, 15: 72-75.
- [5] LAN T W, FAN C J, LI S, et al. Probabilistic prediction of mine dynamic disaster risk based on multiple factor pattern recognition[J]. *Advances in Civil Engineering*, 2018, 2018: 7813931.
- [6] 宋哲,郭湛,习年生,等.基于概率分析的铁路安全风险传递网络模型研究[J].*铁道运输与经济*,2024,46(4):142-152.
SONG Zhe, GUO Zhan, XI Nian-sheng, et al. Research on railway safety risk transfer network model based on probability analysis [J]. *Railway Transport and Economy*, 2024, 46(4): 142-152.
- [7] 吴晶晶,江思义,吴秋菊,等.基于GIS与BP神经网络的崩塌滑坡地质灾害易发性预测[J].*资源信息与工程*,2021,36(4):100-107.
WU Jing-jing, JIANG Si-yi, WU Qiu-ju, et al. Landslide geological hazard vulnerability evaluation based on GIS and BP neural network [J]. *Resource Information and Engineering*, 2021, 36(4): 100-107.
- [8] WANG X D, ZHANG C B, WANG C, et al. GIS-based for prediction and prevention of environmental geological disaster susceptibility: From a perspective of sustainable development [J]. *Ecotoxicology and Environmental Safety*, 2021, 226: 112881.
- [9] 符洪恩,高艺桔,冯莹莹,等.基于GA-SVR-C的城市暴雨洪涝灾害危险性预测——以深圳市为例[J].*人民长江*,2021,52(8):16-21.
FU Hong-en, GAO Yi-ju, FENG Ying-ying, et al. Hazard prediction of urban rainstorm and flood disasters based on GA-SVR-C model: Case study of Shenzhen City [J]. *Yangtze River*, 2021, 52(8): 16-21.
- [10] 宋国策,王高磊,卢大玮,等.基于ISM-ANP的铁路外部环境安全隐患风险评价模型及应用[J].*铁道运输与经济*,2024,46(6):161-168.
SONG Guo-ce, WANG Gao-lei, LU Da-wei, et al. Risk evaluation model of safety hazards in railway external environment based on ISM-ANP and its application [J]. *Railway Transport and Economy*, 2024, 46(6): 161-168.
- [11] 龚艳冰,向林,刘高峰.基于高斯过程回归模型的洪涝灾害损失预测研究——以重庆市为例[J].*长江流域资源与环境*,2019,28(6):1502-1510.
GONG Yan-bing, XIANG Lin, LIU Gao-feng. Study on flood disaster loss prediction based on gaussian process regression model: A case study of Chongqing city [J]. *Resources and Environment in the Yangtze Basin*, 2019, 28(6): 1502-1510.
- [12] CAO Y, YIN K D, LI X M. Prediction of direct economic loss caused by marine disasters based on the improved GM(1,1) model [J]. *Journal of Grey System*, 2020, 32(1): 133-145.
- [13] 李博,冯俏彬,戚克维.基于改进的神经网络模型预测气象灾害经济损失——以广东省台风灾害为例[J].*重庆理工大学学报(自然科学)*,2021,35(4):247-253.
LI Bo, FENG Qiao-bin, QI Ke-wei. Economic loss prediction of meteorological disaster based on improved neural network model: Taking the Guangdong typhoon as an example [J]. *Journal of Chongqing University of Technology (Natural Science)*, 2021, 35(4): 247-253.
- [14] ZHANG Y, HAO Y H. Loss prediction of mountain flood disaster in villages and towns based on rough set RBF neural network [J]. *Neural Computing and Applications*, 2022, 34(5): 2513-2524.
- [15] HUANG X, SONG J Y, JIN H D. The casualty prediction of earthquake disaster based on extreme learning machine method [J]. *Natural Hazards*, 2020, 102: 873-886.
- [16] 胡忠君,刘艳秋,李佳.基于改进GM(1,1)的洪涝灾害应急物资动态需求预测[J].*系统仿真学报*,2019,31(4):702-709.
HU Zhong-jun, LIU Yan-qiu, LI Jia. Dynamic demand forecast of emergency materials for flood disasters based on improved GM(1,1) model [J]. *Journal of System Simulation*, 2019, 31(4): 702-709.

- [17] 刘芳,冯丹,宫雪然. 基于IACO-BP算法的洪涝灾害应急物资需求预测[J]. 沈阳工业大学学报, 2019, 41(3): 332-338.
LIU Fang, FENG Dan, GONG Xue-ran. Demand predicting of emergency supplies for flood disaster based on IACO-BP algorithm[J]. Journal of Shenyang University of Technology, 2019, 41(3): 332-338.
- [18] HOU L, WU X G, WU Z, et al. Pattern identification and risk prediction of domino effect based on data mining methods for accidents occurred in the tank farm [J]. Reliability Engineering and System Safety, 2020, 193: 106646.
- [19] 许佳,丁超,张昕禹,等. 基于复杂网络的城市内涝灾害应急管理研究综述[J]. 自然灾害学报, 2024, 33(3): 1-16.
XU Jia, DING Chao, ZHANG Xin-yu, et al. Review of complex network-based urban waterlogging disaster emergency management research [J]. Journal of Natural Disasters, 2024, 33(3): 1-16.
- [20] 马超群,张爽,陈权,等. 客流特征视角下的轨道交通网络特征及其脆弱性[J]. 交通运输工程学报, 2020, 20(5): 208-216.
MA Chao-qun, ZHANG Shuang, CHEN Quan, et al. Characteristics and vulnerability of rail transit network based on perspective of passenger flow characteristic[J]. Journal of Traffic and Transportation Engineering, 2020, 20(5): 208-216.
- [21] LÜ L Y, ZHOU T. Link prediction in complex networks: A survey [J]. Physica A: Statistical Mechanics and its Applications, 2011, 390: 1150-1170.
- [22] 户佐安,邓锦程,韩金丽,等. 图神经网络在交通预测中的应用综述[J]. 交通运输工程学报, 2023, 23(5): 39-61.
HU Zuo-an, DENG Jin-cheng, HAN Jin-li, et al. Review on application of graph neural network in traffic prediction[J]. Journal of Traffic and Transportation Engineering, 2023, 23(5): 39-61.
- [23] 赵晓丽,苏筠. 基于微博文本数据的灾害信息公众关注分析——以河南暴雨洪涝事件为例[J]. 自然灾害学报, 2024, 33(1): 89-98.
ZHAO Xiao-li, SU Yun. Research on public concern of disaster information based on Weibo data: Take the rainstorm flood event in Henan Province as an example[J]. Journal of Natural Disasters, 2024, 33(1): 89-98.
- [24] YAN D Y, LI K P, GU S, et al. Network-based bag-of-words model for text classification[J]. IEEE Access, 2020, 8: 99.
- [25] YANG L, LI K P, ZHAO D, et al. A network method for identifying the root cause of high-speed rail faults based on text data[J]. Energies, 2019, 12(10): 1908.
- [26] 李曼,刘贵源,王艳辉,等. 风险演化研究综述及其在轨道交通领域的应用与探讨[J]. 铁道运输与经济, 2025, 47(1): 13-30.
LI Man, LIU Gui-yuan, WANG Yan-hui, et al. Review of risk evolution research and its application and discussion in rail transit[J]. Railway Transport and Economy, 2025, 47(1): 13-30.
- [27] PECH R, HAO D, PAN L M, et al. Link prediction via matrix completion[J]. Europhysics Letters, 2017, 117(3): 38002.
- [28] LI K P, GU S, YAN D Y. A link prediction method based on neural networks[J]. Applied Sciences, 2021, 11: 5186.
- [29] 潘黎明. 复杂网络上的链路预测及信息传播研究[D]. 成都: 电子科技大学, 2019.
PAN Li-ming. Studies on link prediction and information spreading in complex networks [D]. Chengdu: School of Computer Science & Engineering, 2019.