

面向城市轨道交通列车运维安全的事理图谱构建研究^{*}

樊茜琪^{1,3} 左建勇¹ 龚明² 贾波³ 李政江¹

(1. 同济大学交通学院, 201804, 上海; 2. 中车工业研究院有限公司, 100160, 北京;

3. 上海轨道交通检测认证(集团)有限公司, 201804, 上海)

摘要 [目的]城市轨道交通运维系统庞大而复杂,使用传统数据准确分析运维风险具有一定难度,将大数据分析纳入风险管理是一种趋势。通过构建城市轨道交通列车运维安全事理图谱,为城市轨道交通安全管理和风险防控提供新的视角和工具,从而为提升城市轨道交通系统的安全性提供理论基础和方法支持。[方法]聚焦于列车运维安全事理图谱的构建,对构建事理图谱的3个关键步骤进行了优化。事件抽取环节构建了城市轨道交通专业词典并设计了触发词规则;事件关系抽取环节结合了语义依存与顺序关系;事理对齐环节采用优化的聚类算法来定义主要风险因素。通过对传统知识图谱方法、通用事理图谱方法以及本文方法的对比评价,验证本文方法在事理图谱构建中的准确性和有效性。最后通过 Gephi 软件实现了事理图谱用数据存储和可视化。[结果及结论]传统的基于数理统计的故障分析和依赖人工经验的风险分析在预见性、全面性和及时性等方面存在一定的局限性;展示了事理图谱在运维风险管理中从理论到应用的完整路径,并实现了数据分析与可视化表示的综合方法论。随着方法论的不断完善和实证研究的深入,这套城市轨道交通领域事理图谱抽取范式将能进一步促进城市轨道交通安全管理水平的提升。

关键词 城市轨道交通; 列车运维安全; 事理图谱; 风险分析

中图分类号 U298

DOI:10.16037/j.1007-869x.2025.04.007

Research on the Construction of Event Evolutionary Graph for Urban Rail Transit Train Operation and Maintenance Safety

FAN Qianqi^{1,3}, ZUO Jianyong¹, GONG Ming², JIA Bo³, LI Zhengjiang¹

(1. College of Transportation Engineering, Tongji University, 201804, Shanghai, China; 2. CRRC Academy, 100160, Beijing, China; 3. Shanghai Railway Certification (Group) Co., Ltd., 201804, Shanghai, China)

Abstract [Objective] As the urban rail transit operation and maintenance system is large and complex, it is somewhat difficult to accurately analyze the operation and maintenance risks using traditional data. Incorporating big data analysis into risk management is a development trend. By constructing an event evolutionary graph for urban rail train operation and maintenance safety, new perspectives and tools are provided for urban rail transit safety management, risk prevention and control, thereby providing a theoretical basis and methodological support for improving the safety of urban rail transit systems.

[Method] Focusing on the construction of event evolutionary graph for urban rail train operation and maintenance safety, three key steps for constructing the event evolutionary graph are optimized. In the event extraction stage, a professional dictionary of urban rail is constructed and trigger word rules are designed; in the event relationship extraction stage, semantic dependency and sequential relationship are combined; in the event evolutionary alignment stage, an optimized clustering algorithm is used to define the main risk factors. Through comparative evaluation of traditional knowledge graph methods, general event evolutionary graph methods and the above proposed method, accuracy and effectiveness of the last one in the construction of event evolutionary graph are verified. Finally, the data storage and visualization of the event evolutionary graph are realized through Gephi software.

[Result & Conclusion] Traditional fault analysis based on mathematical statistics and risk analysis relying on manual experience have certain limitations in terms of predictability, comprehensiveness, and timeliness. The complete path from theory to application of event evolutionary graph in operation and maintenance risk management is presented, and a comprehensive methodology for data analysis and visual representation is realized. With the continuous improvement of methodology and the deepening of empirical research, it is believed that this paradigm for extrac-

^{*} 国家重点研发计划项目(2020YFB1600704)

ting the theoretical graph of urban rail transit field will further promote the improvement of urban rail transit safety management level.

Key words urban rail transit; train operation and maintenance safety; event evolutionary graph; risk analysis

城市轨道交通(以下简称“城轨”)列车运维系统因技术密集、种类繁多、各系统存在相互交叉等特点,使得借助大数据分析运维风险成为一种新模式。本文旨在通过事理图谱技术构建一个全新的城轨运维风险分析框架,通过事理图谱识别和表征复杂系统中的风险因素及其相互作用,将散落的风险事件和因素以图谱的形式直观展现。

1 事理图谱

1.1 研究现状

事理图谱已在深度学习、金融经济、神经科学、生物信息学、社会科学和航空航天等领域得到广泛应用^[1]。在城轨领域,文本数据通常缺乏明确的规则性关系,这使得直接应用传统的规则抽取技术难以达到高准确性,特别是在列车运维安全这一子领域,相关的研究尤为稀缺,导致可用于构建事理图谱的规则和模型不够完善。例如,城轨领域的术语、专业表达方式以及操作流程具有其特定性和复杂性,如果不进行定制化处理,很容易导致信息抽取的准确率低、相关性差。

1.2 构建关键技术

在构建、推理及应用事理图谱的过程中,核心是事件抽取技术、事件关系推断技术和自然语言处理技术^[1-3]。事件抽取技术指从非结构化文本中识别出事件相关元素(如事件触发词、事件论元等)的过程;事件关系推断技术指利用文本中的信息来推断事件之间的关系;自然语言处理技术指自然语言处理模型通过在大量文本上的预训练,学会了丰富的语言表示,能够理解复杂的词汇依赖和语境信息,是实现事件识别和事件关系抽取的有效工具。

2 城轨列车运维安全事理图谱规则定义

有效构建城轨列车运维领域事理图谱的过程主要分为3个关键步骤:事件抽取、事件关系抽取和事理对齐,如图1所示。

2.1 事件抽取

事件抽取任务要素的定义包含事件触发词、事

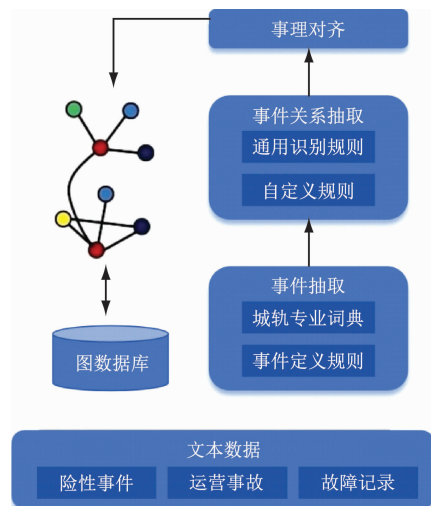


图1 城轨列车运维安全事理图谱构建流程

Fig. 1 Construction process of event evolutionary graph for urban rail train operation and maintenance safety

件类型、事件元素以及事件元素角色,如表1所示。

表1 事件抽取任务要素定义

Tab. 1 Definition of event extraction task elements

要素	定义
事件触发词	表明事件发生并决定事件类型的核心词汇,通常为动词和名词
事件类型	城市轨道交通事件类型,可分为险性事件、运营事故、状态改变(故障)、信息传递等
事件元素	涉及或参与事件的实体,比如地点、人物、组织、设备等
事件元素角色	在事件中,各个事件元素所承担的具体角色或作用

本文运用基于机器学习的事件抽取方法,将事件抽取任务分为事件触发词检测与事件论元识别两个步骤。使用自然语言处理技术对文本进行分词等预处理,通过构建事件触发词来识别与风险事件直接相关的事件,再通过句法分析识别其余事件的论元及角色,进行语义角色标注,抽取可以表征事件的短语或短句作为识别结果。重复进行上述抽取流程,将符合事件要求的结果保留。

2.2 事件关系抽取

事件关系抽取是获取城轨列车运维风险事件发展过程的必要手段,是以事件对的形式从事件的描述文本中将存在联系的前置事件和后续事件表示出来。为精准识别城轨领域事件关系,提出了基于语义关系的显性关系抽取和基于叙事顺序的隐性关系抽取规则,如图2所示。

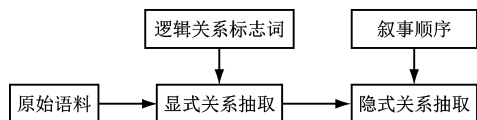


图2 事件关系抽取规则

Fig. 2 Event relation extraction rules

1) 基于语义关系的显性关系抽取。其依赖于对文本语义特征的分析,对两个事件是否存在直接演化关系进行判定,并提取相关联事件对。以“原因是”“结果是”“导致”“造成”“后”等为代表的逻辑关系标志词支撑了语义关系的识别。根据风险事件演变路径研究的需要,本文主要关注施事关系、当事关系、成事关系、源事关系、依据角色、结局角色、等同关系、先行关系、顺承关系、因果关系、结果关系,来描述事件的演变,如表2所示。

表2 城轨领域文本语料典型关系

Tab.2 Typical relationships of text corpus in urban rail transit field

关系	定义	示例
施事关系	指示谁是行动或事件的发起者	〈检修工〉检查列车系统
当事关系	指示事件或行动的直接对象或主题	检修工检查〈列车系统〉
成事关系	表示某个动作导致了一个新状态的出现	技术更新提高了〈列车运行效率〉
源事关系	标示出事物的起源或来源	因为〈配件磨损〉,列车出现故障
依据角色	指出某一行动或判断的根据	根据〈定期维护计划〉,检修工开始检查
结局角色	事件的结果或后果	检查后,列车系统故障〈得到解决〉
等同关系	表示两个概念等同或同义	安全检查〈即是〉预防措施
先行关系	表示时间或序列上的先后关系	〈完成安全培训〉后,新员工开始上岗
顺承关系	表示事件或行为的自然顺序或逻辑上的后续	列车出站〈后〉,进入下一个检修周期
原因关系	指出造成某结果的原因	〈操作失误〉导致列车延误
结果关系	展示了一个事件或行动的直接结果	操作失误导致〈列车延误〉

注:〈〉表示在文本中识别出的对应语料关系示例。

2) 基于叙事顺序的隐性关系抽取。在城轨领域的记录文本中,大量事件记录缺少逻辑标志词,而是直接按照事件发生的先后顺序进行记录。在基于加载的城轨专业词典基础上进行分词和词性

识别,去除事件、地点、数量等词后重新进行自然语言处理,得到语义角色标注结果和语义依存分析结果。如:语料“07:11 10426202 次(车底 06)在出段过程中车辆屏显示高速断路器断开、网压为零;组织该车换端回段,热备车上线替开。故障原因为受电弓升弓高度不足未与接触网接触受流,重新调整受电弓升弓高度后恢复受流。造成晚点 1 列,晚点 5 min 40 s。”先通过语义识别得到“故障原因为”和“造成”为关键词的原因和结果辨识,然后选用叙事顺序进行描述,则可得到的事件关系为“车辆屏显示—高速断路器断开—网压为零—换端回段—热备车上线替开”,这样得到的事件关系更加符合事理顺序。

2.3 事理对齐

2.3.1 知识融合

将上述识别和抽取出来的事件及其关系对齐到具体的知识框架中,从而构建出完整的事理图谱。这需要预先定义一个知识框架,包括一系列的事件类型和事件之间可能的关系类型。在城轨领域,这个框架可能包括的事件类型如“脱轨”“碰撞”“晚点”等,以及关系类型如“导致”“之后发生”“为了”等。对齐的过程可能需要人工干预,以确保图谱的准确性和一致性。在深入分析和整理已经抽取得到的事件后,这一环节可视作是对事件信息的进一步聚类 and 标准化处理,使之符合同一规范事件的描述要求。在文本聚类处理中,基于势能优化的 CFSFDP(快速密度峰值搜索)算法来刻画聚类中心,再利用 K-Means(K-均值聚类)算法进行迭代聚类,以确定最大 k 值 k_{\max} 及最佳 k 值 k_{opt} ,以此自动选择初始聚类中心^[4]。其原理是由于聚类中心点与其他任何比其密度大的点之间有相对较大的距离,那么那些具有较大距离值的点以及较大密度值的点就有更大的可能成为初始聚类中心。计算不同 k 值时的 Davies-Bouldin 指标(简称 DB 指标)值,指标最小时的 k 值即为最佳聚类数。最佳聚类数时的聚类结果即可作为文本聚类结果使用。

2.3.2 事件分类

为了对城轨领域的风险事件进行更深层次的探索与研究,本研究进一步定义了一系列典型事件,并通过人工干预来约束和优化最终希望得到的目标事件集合。在具体分析城轨列车的运维风险时,主要风险源被识别为设备老化或设计缺陷、人为错误、环境因素、管理缺陷,如表3所示。

表3 城轨列车运维主要风险

Tab.3 Main risks in urban rail train operation and maintenance

风险分类	风险事件
设备老化或设计缺陷	车体、转向架、车辆连接装置、制动装置、受流装置、车辆设备、车辆电气系统故障或存在设计缺陷
人为错误	操作人员疏忽或技能不足(如乘务人员故障判断不准确、司机缺乏经验等事件)、维护工作不足
环境因素	自然灾害(如下雨、下雪、地震、洪涝等事件)、异物侵限
管理缺陷	预防性维护计划不到位、应急响应程序更新不及时、安全管理制度不足

通过前述的事件聚类、标准化描述以及风险源及后果的详细分类分析,深入挖掘运作机制、问题成因和可能的解决策略等方面的信息,从而构建一个全面、准确且实用的城轨领域事理图谱。

3 城轨列车运维安全事理图谱构建

3.1 数据源与预处理

本文以城轨列车为例进行研究,数据源于互联网新闻报道、交通部发文、运营企业运营数据等。

1) 行业事故/险性事件。此数据集共包含 56 条记录,均为城轨领域运营期间的安全事件,内容为完整的事事件调查报告^[5]。

2) 某城轨运营企业运营事故报告。此数据集共包含 275 起记录,包括该企业关于清客、退出正线运营、晚点 5 min 及以上等事故事件的记录数据,包含事故描述、后果、处理措施及原因分析等板块。

3) 某城轨运营企业车辆系统故障维修记录。此数据集共包含 3 603 起记录。本文针对的是城轨列车运维安全,因此只关注风险事件的演变路径,即在故障记录中只提取故障发生的设备类型、故障描述、故障后果、故障类型、维修措施、原因分析等相关数据,并将其形成新的连续的故障描述。

3.2 事件及事件关系抽取

3.2.1 基于城轨事理图谱构建规则的事件及事件关系抽取

基于前述设定的事件和事件关系抽取规则进行城轨列车运维安全事理图谱的构建。主要使用 Python 语言,借助 LTP(语言技术平台)模型进行分词、词性标注、语义角色标注、语义依存分析等操作,并将设定的规则加入其中。本文由于设定规则的特殊性以及不具有明显的逻辑联系词,认为抽取

的事件关系均属于因果关系(或时序关系),存储时以二元组的形式,即“事件 A-事件 B”的形式进行提取存储。经过城轨事理图谱构建规则处理后的结果如表 4 所示。

表4 城轨事理图谱构建规则下的处理结果

Tab.4 Processing results under construction rules of urban rail event evolutionary graph

字段	内容
语料 1	2021 年 4 月 3 日,某地地铁 X 号线甲站至乙站下行区间发生弓网故障,造成甲站至乙站下行区间接触网塌网,导致乙站至丙站区间单向中断行车约 5 h 58 min,直接经济损失约 17 万元。本次事件的直接原因是列车受电弓拉杆与上臂杆的球头连接部位在质保期内疲劳断裂,致使扭臂尾端抬升并侵入设备限界。设计、制造单位对刚性接触网与受电弓在复杂运行工况下的受力和失效机理认识不到位,运营单位对受电弓关键部件风险认识不足,也是造成本次事件的原因
结果 1	弓网故障发生→下行区间接触网塌网→中断行车。 本次事件的直接原因是列车受电弓拉杆与上臂杆的球头连接部位在质保期内疲劳断裂,致使扭臂尾端抬升并侵入设备限界。 列车受电弓拉杆与上臂杆的球头连接部位疲劳→列车受电弓拉杆与上臂杆的球头连接部位断裂→扭臂尾端侵入设备限界。 设计、制造单位对刚性接触网与受电弓在复杂运行工况下的受力和失效机理认识不到位,运营单位对受电弓关键部件风险认识不足是造成本次事件的原因→造成本次事件
语料 2	07:11 10426202 次(车底 06)在出段过程中车辆屏显示高速断路器断开、网压为零;组织该车换端回段,热备车上线替开。故障原因为受电弓升弓高度不足未与接触网接触受流,重新调整受电弓升弓高度后恢复受流。造成晚点 1 列,晚点 5 min 40 s
结果 2	车辆屏显示高速断路器断开、网压为零→组织该车换端回段→热备车上线替开。 故障原因为受电弓升弓高度不足未与接触网接触受流,重新调整受电弓升弓高度后恢复受流→造成晚点,晚点受电弓升弓高度不足→未接触→未受流→调整受电弓升弓高度→恢复受流

从语料抽取结果来看,存在重复抽取现象,长句可能作为事件被抽取的同时,其内部也被进一步细化进行抽取;存在描述不准确现象,比如“造成晚点,晚点”,实际是原语料用来描述列车数量和时间的两个部分。尽管有以上不足存在,但基本抽取了记录中的因果关系、时序关系等事件,实现了相关记录从非结构化数据向结构化数据的转变,有助于图谱的建立,减少了人工辅助识别的工作量。

3.2.2 事件关系抽取效果评价

为量化评估本研究构建的城轨列车运维安全事理图谱规则,随机抽取数据库中 50 条数据作为评

价对象。利用知识图谱方法、通用事理图谱方法和本研究方法进行数据处理,将专家人工识别的事件及事件关系作为验证数据库,以验证本研究方法的优化性能,评价指标包括精确率、召回率、F1 分数(通常用来评价数据不平衡情况下的分类器性能)。分别计算不同方法下的各个指标值,结果如表 5 所示。

表 5 不同方法下的城轨列车运维安全事件抽取效果评价表

Tab.5 Evaluation table of urban rail train operation and maintenance safety events extraction effect using different methods

方法	精确率/%	召回率/%	F1 分数/%
知识图谱方法	78.75	37.06	50.40
通用事理图谱方法	83.19	55.29	66.43
本研究方法	56.86	85.29	68.24

本研究方法的精确率为 56.86%,这表明虽然该方法可能标记更多的样本为正例,但其准确性较低。经过进一步分析可以看出本研究方法抽取结果中不仅包含了与风险事件发生直接相关的事件,

也包含了在抢修过程中一系列检测、测试的过程性活动,尽管对于描述风险事件发生来说可能是噪声数据,但对于指导运维工作来说也是重要的组成数据。本研究方法的召回率最高,为 85.29%,这表明其在不遗漏正类样本的能力上表现最好。本研究方法的 F1 分数最高,为 68.24%。本研究方法在模型表现上总体具有较为优秀的表现。

3.3 事理对齐及选择

事理对齐步骤是对图谱中的原始事件和结果进行归纳和抽象,以归一化的风险事件表示来消除具体事件中的非本质差异,增强研究的泛化能力。这一过程通常依赖于聚类算法,根据事理之间的相似性进行匹配,从而形成了一组包含共性并且能够代表广泛情形的抽象风险事件。进一步地,剔除无影响的中间事件,将风险事件根据前述研究中的风险分类进行类别划分。为保证研究的普适性和可靠性,设置了一定的筛选条件,有助于排除那些偶发的、孤立的风险事件,保证分析焦点集中在具有普遍关联能量和潜在影响力的事件上。得到的部分事件关系对如表 6 所示。

表 6 事件关系对

Tab.6 Event relationship pairs

事件源	目标对象	事件源	目标对象	事件源	目标对象
CCTV(视频监控)系统故障	退出正线运营	设计缺陷	退出正线运营	贯通道异常	退出正线运营
安全管理制度不足	维护工作不足	受流器磨损或损坏	牵引供电故障	其他	退出正线运营
车辆控制系统故障	车门故障	维护工作不足	牵引供电故障	电气装置故障	牵引供电故障
车辆控制系统故障	乘客信息系统故障	维护工作不足	缓冲装置故障	电源模块故障	退出正线运营
车辆控制系统故障	广播故障	维护工作不足	制动装置故障	电源模块故障	车门故障
车体故障	退出正线运营	异物侵限	退出正线运营	预防性维护计划不到位	缓冲装置故障

值得注意的是,这种基于文本数据构建的事理图谱可能无法完全捕捉到所有具有实际意义的事件关系,这主要受限于两方面因素。一是筛选标准的建立,使一些条目由于参与事件的数量过少而被排除在外;二是原始语料库的质量和完整性。

3.4 事理图谱存储与可视化

为深入解析这一图谱中所蕴含的复杂事件关系,揭露事件之间的关系布局与特征,本文选用了 Gephi 软件进行可视化分析^[6]。通过节点来表示风险事件,同时利用箭头指出事件之间的因果关系,并通过箭头的粗细来体现特定路径上事件发生的频次,可以更清晰地识别出影响城轨系统稳定性和效率的关键因素。图 3 是原始数据以脱轨为中心事

件的相关事件关系图。通过聚类进行图谱剪枝化简,能够对发生频次较高的共性风险事件进行重点研究。这些事件对的关系直观表现如图 4 所示。

4 结语

本文探讨了面向城轨列车运维安全的事理图谱构建过程,并以此作为一个理论和实践的框架,旨在为城轨领域事理图谱抽取设立标准化的工作流程和方法。

为了充分适应城轨领域的专业特性和需求,本文依据城轨领域丰富的文本语料,制定了句子结构+触发词的事件识别规则,以及语义关系+叙事关系的事件关系抽取规则,并通过事件聚类、标准

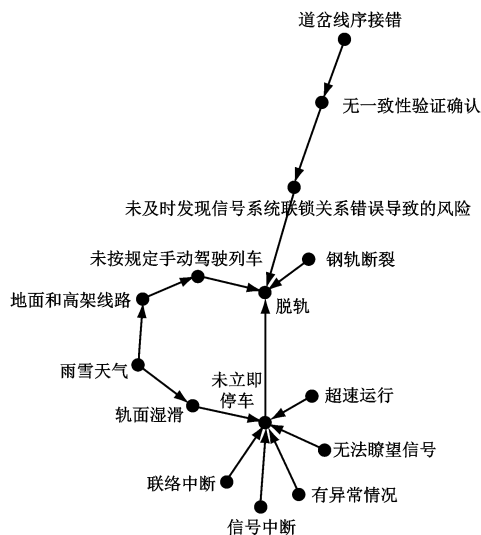


图3 以脱轨为中心事件的相关事件关系图

Fig. 3 Relationship diagram of related events with derailment as the central event

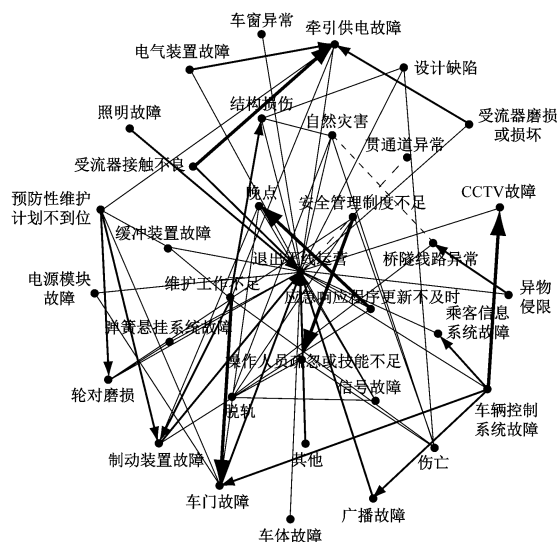


图4 发生频次较高的风险事件

Fig. 4 Risk events with higher frequency of occurrence

化描述以及人工干预进行事理对齐,确保了图谱构建过程的专业性和准确性。通过实例分析验证了规则集的有效性,构建的城轨列车运维安全事理图谱能反映文本中记录的各种事理关系,并描绘出风险事件之间的依存和转化关系。可视化的风险演

变路径,有助于更加深入地理解和把握城轨安全风险体系的动态特性。

在未来的工作中,需要进一步优化数据收集和处理方法,提升语料库的质量和覆盖面,探索更灵活的筛选标准和分析方法,以适应不同类型和规模的风险事件分析。

参考文献

- [1] 胡志磊, 靳小龙, 陈剑赞, 等. 事件图谱的构建、推理与应用[J]. 大数据, 2021, 7(3): 80.
HU Zhilei, JIN Xiaolong, CHEN Jianyun, et al. Construction, reasoning and applications of event graphs[J]. Big Data Research, 2021, 7(3): 80.
- [2] 赵军, 刘康, 何世柱, 等. 知识图谱[M]. 北京: 高等教育出版社, 2018.
ZHAO Jun, LIU Kang, HE Shizhu, et al. Knowledge graph[M]. Beijing: Higher Education Press, 2018.
- [3] 项威. 事件知识图谱构建技术与应用综述[J]. 计算机与现代化, 2020(1): 10.
XIANG Wei. Reviews on event knowledge graph construction techniques and application[J]. Computer and Modernization, 2020(1): 10.
- [4] 李敏. K-means 算法的改进及其在文本聚类中的应用研究[D]. 无锡: 江南大学, 2017.
LI Min. Improvement of K-means algorithm and its application in text clustering research[D]. Wuxi: Jiangnan University, 2017.
- [5] 马子彦. 轨道交通运营事故案例分析[M]. 北京: 北京交通大学出版社, 2013.
MA Ziyang. Case analysis of rail transit operation accident[M]. Beijing: Beijing Jiaotong University Press, 2013.
- [6] 邓君, 马晓君, 毕强. 社会网络分析工具 Ucinet 和 Gephi 的比较研究[J]. 情报理论与实践, 2014, 37(8): 133.
DENG Jun, MA Xiaojun, BI Qiang. Comparative study of the social network analysis tools: Ucinet and Gephi[J]. Information Studies (Theory & Application), 2014, 37(8): 133.

· 收稿日期:2024-11-20 修回日期:2025-02-07 出版日期:2025-04-10

Received:2024-11-20 Revised:2025-02-07 Published:2025-04-10

· 通信作者:樊茜琪,高级工程师,fanqianqi@126.com

· ©《城市轨道交通研究》杂志社,开放获取 CC BY-NC-ND 协议

© Urban Mass Transit Magazine Press. This is an open access article under the CC BY-NC-ND license

欢迎访问《城市轨道交通研究》网站

www. umt 1998. tongji. edu. cn