

基于知识-数据双驱动事件图谱的 配网智能决策关键技术研究

技术报告

2024 年 6 月

目 录

一、国内外研究现状	1
1.1 知识图谱研究现状	1
1.1.1 知识图谱概念定义	1
1.1.2 领域知识图谱研究现状	1
1.1.3 知识图谱研究综述	3
1.2 事件图谱研究现状	3
1.2.1 事件图谱概念发展	3
1.2.2 领域事件图谱研究现状	4
1.2.3 事件图谱研究综述	5
1.3 电力领域图谱研究现状	5
1.3.1 电力领域图谱总体研究现状	6
1.3.2 配电网领域图谱研究现状	6
1.3.3 电力领域图谱研究综述	7
1.4 综合评述	7
二、配网事件抽取关键技术	8
2.1 基于事件模板框架的配网事件知识抽取技术	9
2.2 基于 BERT-BiLSTM-CRF 的配网设备实体识别技术	10
三、配网事件图谱构建及可视化方案报告	12
3.1 时序-因果双类型关系配网事件图谱构建原理	13
3.1.1 事理图谱模块	13

3.1.2 时序事件链模块	14
3.1.3 设备关联图模块	16
3.1.4 时序-因果双类型关系配网事件图谱示意图	17
3.1.5 时序-因果双类型关系配网事件图谱本体示意图	18
3.2 基于原图谱本体的新图谱 Schema 设计	19
3.3 图谱可视化设计方案	20
四、配网事件图谱推理关键技术研究报告	22
4.1 基于贪心前向匹配算法的自动词库构建技术	22
4.2 基于事件图谱和图关联分析的推理技术方案	24
五、配网事件图谱故障定位归因及辅助决策应用方案	25
5.1 基于 TCEKG 的故障诊断模型	25
5.2 基于 TCEKG-FDM 的故障定位	27
5.2.1 DMS 告警解析	27
5.2.2 故障类型识别	28
5.3 基于 TCEKG-FDM 的故障归因	29
5.3.1 故障原因预测	30
5.3.2 相似事件推荐	30
5.4 基于 TCEKG-FDM 的故障辅助决策	32
5.4.1 规程知识库构建	32
5.4.2 故障辅助决策推荐	34
参考文献	35

一、国内外研究现状

事件图谱起源于知识图谱，知识图谱和事件图谱目前在特定领域中得到广泛的应用。本文面向配电网故障处置领域进行事件图谱的构建，亟待了解知识图谱、事件图谱以及它们在电力领域中的研究现状。因此，本文将对知识图谱、事件图谱和电力领域图谱的国内外研究现状进行综合评述。

1.1 知识图谱研究现状

知识图谱（Knowledge Graph, KG）由谷歌公司于 2012 年正式提出，目的是实现搜索引擎的智能化，提供用户友好的搜索结果。知识图谱概念正式提出后，受到学术界和业界的广泛关注，并被应用于知识问答、智能搜索、欺诈检测等多个领域。本文将从知识图谱概念定义和领域知识图谱研究现状展开。

1.1.1 知识图谱概念定义

当前，学术界和业界对知识图谱尚无明确的定义，不同研究方向的学者从不同角度对知识图谱进行定义。其一，从组成要素角度定义。漆桂林等（2017）认为知识图谱是一个由实体或概念作为结点、实体/概念之间的各种语义关系作为边的，具有有向图结构的知识库；Ji S X 等（2021）将知识图谱定义为一种由实体、关系和语义描述组成的结构化事实表示形式。其二，从表示方式角度定义。刘峤等（2016）认为知识图谱是由“实体-关系-实体”三元组以及实体及其相关属性-值对相互联结构成的网状知识结构。其三，从功能用途角度定义。李娟子等（2017）认为知识图谱将互联网信息表示成更接近人类认知世界的形式，提供了一种更好地组织、管理和理解互联网海量信息的能力；李涛等（2017）将知识图谱定义为一种帮助用户迅速、准确地查询到所需信息的智能、高效的知识组织方式；Hogan 等（2021）将知识图谱定义为旨在积累和传递现实世界知识的数据图。

1.1.2 领域知识图谱研究现状

知识图谱可根据覆盖范围分为通用知识图谱（General-purpose Knowledge

Graph, GKG) 和领域知识图谱 (Domain-specific Knowledge Graph, DKG)。GKG 知识规模庞大, 注重知识的广度, 目前很多研究机构和公司都发布了 GKG, 主要应用于面向互联网的搜索、推荐、问答等业务场景, 其中 Freebase、Wikidata、DBpedia、YAGO 这 4 个项目处于 GKG 的核心地位; 相较于强调知识广度的 GKG, DKG 面向专业领域, 注重行业知识的深度, 用于特定领域或业务场景。

DKG 作为相对较新的研究领域, 还没有一个完善、一致、全面的定义。Fan Y 等 (2017) 认为 DKG 是一种特殊类型的知识图谱, 用于对特定领域的实体之间的关系进行建模; Kejriwal (2019) 从本体论角度, 认为 DKG 是充实领域本体的结果; Abu-Salih 等 (2021) 给出更为通用的定义, 即 DKG 是对高层次主题领域及其具体子领域的明确概念化, 并用语义相关的实体和关系表示。

DKG 在学术界被广泛研究。Abu-Salih 等 (2021) 总结了文献中曾研究过的七大主题领域的 DKG, 包括医疗、教育、信息和通信技术、科学和工程、金融、社会和政治、旅游。具体来说, 在医疗领域, Zhang Y 等 (2020) 提出健康知识图谱 HKGB, 用于从多个数据源构建特定疾病、大规模和跨语言的健康知识图, 为整合异质数据资源和增强基于知识的服务提供了理想的技术手段; 在教育领域, Shi D Q 等 (2020) 提出基于多维知识图谱框架的网络学习路径推荐模型, 可以生成并推荐合格的个性化学习路径, 以改善网络学习者的学习体验; 在工程领域, Zhu Y Q 等 (2017) 设计了地质数据的知识图谱框架, 构建了地质数据的知识图谱应用系统, 并相应完成了地质信息的动态更新。

此外, DKG 在业界也受到广泛关注, 目前在各个行业中也有相关产品的出现。在生活服务领域, 美团构建了生活服务标签图谱, 更好地连接了商户供给和用户意图, 并在美团知识问答、搜索和推荐中广泛应用。阿里巴巴构建了面向直播的新零售知识图谱 AliMe MKG, 结合多模态技术, 构建以内容为中心的商品素材知识图谱, 提供商品的认知画像, 帮助消费者在直播间内进行消费决策; 在医疗健康领域, 医渡云构建了真实世界的医疗知识图谱及临床事件图谱, 为医疗机构提供了信息检索、事件搜索、诊断推荐、智能问诊、智能生成病历等功能。腾讯在医疗知识图谱方向也进行了实践, 通过构建医疗知识图谱, 实现了辅助诊断、辅助理解、数据治理、用药助手等功能; 在机器翻译领域,

vivo 利用知识图谱辅助机器翻译，应用跨语言知识图谱增强机器翻译表现，利用多维度信息辅助翻译理解；在电子竞技领域，小米通过构建游戏知识图谱，帮助应用方更好地理解游戏和用户意图。

1.1.3 知识图谱研究综述

综上所述，目前知识图谱的定义尚不明确，国内外学者主要从组成要素、表示方式、功能用途等不同角度对知识图谱进行定义。此外，知识图谱可按照知识范围分为 GKG 和 DKG，其中 DKG 发展更加多元化。

知识图谱的研究现状可以归纳为以下两点：

（1）尽管知识图谱的定义不尽相同，但知识图谱本质上是一个大规模语义网络，网络中的节点是概念、实体或属性，网络中的边是概念、实体或属性之间的语义关系。

（2）知识图谱可分为 GKG 和 DKG 两大类型。其中，GKG 注重知识的广度，以百科、常识、概念知识图谱为主，主要服务于互联网搜索、推荐、问答等业务场景；DKG 注重知识的深度，主要和具体的业务场景相结合，从而为特定领域的应用提供技术支持。

1.2 事件图谱研究现状

知识图谱主要关注静态实体及其之间的语义关系，并以语义网络（Semantic Network）的形式组织起来。但知识图谱只关注到实体及其关系中包含的静态知识，并不能很好的处理具备因果、时序性的动态知识，因此引发了对事件图谱的研究。本节将从事件图谱概念发展和领域事件图谱研究现状展开。

1.2.1 事件图谱概念发展

Glavas 等（2014）首次提出事件图谱（Event Graph）的概念，它作为一种新颖的基于事件的文档表示模型，可以对文本中描述的事件信息进行过滤和结构化，以满足高效检索和简明呈现事件相关信息需要；Rospocher 等（2016）从知识图谱角度衍生出以事件为中心的知识图谱（Event-Centric Knowledge Graph），其中节点是由 URI 和实体标识的事件，而边是事件-实体关系、事件-事件关系以及关于实体的一般事实。

之后，随着事件预测、对话系统等下游任务的出现和发展，事件的演变和发展变成事件图谱的重点研究方向。Li Z Y (2017) 提出了和事件图谱相似的事件演化图 (Event Evolutionary Graph, EEG)，但事件演化图相比于事件图谱，其中的事件节点是由抽象的、概括的、语义完整的动词短语表示的，它进一步考虑了事件之间的因果关系，并揭示了现实世界事件的进化模式和发展逻辑；Gottschalk 等 (2018) 提出了以事件为中心的时序知识图谱 (Event-Centric Temporal Knowledge Graph)，其中事件、实体和关系被表示为节点，以促进对网络、新闻和社交媒体上有关当前和历史事件的信息进行语义分析；Ding X (2019) 提出由事件演化图发展而来的事件逻辑图 (Event Logic Graph, ELG)，其中节点是抽象的、概括的、语义完整的事件三元组。

此外，随着其他人工智能技术的不断发展，事件图谱在形式上更加多样化。Ma Y B 等 (2022) 在多模态技术发展的背景下，提出了大规模的多模态事件知识图谱 (Multi-Modal Event Knowledge Graph, MMEKG)，它以事件作为基本构件，通过连接、补充和消除不同模态的知识，以便更好地理解或推理知识。

1.2.2 领域事件图谱研究现状

自事件图谱概念提出后，国内外学术界和业界都基于不同类型的事件图谱在特定领域或业务场景进行了广泛的研究和应用，目前已经在智能搜索、旅游、医疗健康、公共安全、图书情报等领域取得了一定的进展。

Rudnik 等 (2019) 开发了一个面向事件的知识图谱和基于事件的搜索引擎，实现了知识图谱结构化数据和新闻文章非结构化数据的查询，提高了搜索引擎的检索能力；Wu J 等 (2020) 提出了一个以事件为中心的旅游知识图谱，它从以事件为中心的旅游知识图谱中获得的游客行为模式，以便更有效地捕捉用户与兴趣点之间的关系，实现了对用户兴趣点的推荐；刘忠宝等 (2020) 面向《史记》中的历史事件和元素构建了《史记》事理图谱，发现了《史记》中蕴含的鲜为人知的知识，促进了文献学、历史学和社会学的研究。李纲等 (2021) 基于国家安全事件态势感知的情报需求，提出国家安全大数据事件图谱，能够更好地服务于国家安全事件态势的全面监控与感知，从而为国家安全事件管理决策提供情报支撑；张海涛等 (2021) 提出了重大突发事件领域的事理图谱，并以 2019 冠状病毒疫情为例，研究了事理图谱对重大突发事件的演变逻辑与规

律的有效性，从而可以更加科学准确的应对和治理重大突发事件；Bajracharya 等（2022）提出了基于图数据库存储的临床数据的实体-事件知识图，它将病人数据与知识库联系起来，实现了健康信息学的强大用例分析。

此外，业界也基于事件图谱开发了相关的产品。百度提出了百度大脑事件图谱，该图谱通过整合客观世界复杂多样的事件数据，将实体图谱在时间、空间、动作多个维度展开，目前已应用到搜索、信息流等百度内部的产品中；京东提出了京东科技事理图谱，该图谱共有 100 万多节点，覆盖了 45 个产业链的行业知识。目前，在金融领域，京东科技事理图谱已经可以实现为金融舆情系统提供多维度的连接关系、为智能投研系统自动发现投资逻辑以及为金融预测模型提供丰富的潜在特征等。在智慧研报领域，京东科技事理图谱已经可以实现对研报资讯相关文章的推荐等。

1.2.3 事件图谱研究综述

综上所述，国内外对于事件图谱的定义相较于知识图谱的定义更不明确，目前事件图谱类型多样，且很难统一，学术界和业界也没有将知识图谱和事件图谱严格区分开，可以认为事件图谱是知识图谱的一种变体。此外，国内外对于事件图谱在特定领域的研究和应用较少，国内主要聚焦于公共安全方向，国外研究方向比较分散、资料较少，大多和具体领域相结合。

事件图谱的研究现状可以归纳为以下两点：

（1）目前学术界和业界出现的事件图谱类型多样，但尚未出现统一的定义和分类标准。事件图谱本质上是包含事件节点的语义网络，可以认为是知识图谱在事件方向的衍生概念。

（2）事件图谱由于包含事件节点，大多和具体领域相结合。不同业务场景下的事件图谱侧重点不同，在新闻领域更注重事件的查询，在舆情和重大突发事件领域更注重事件的演变，在旅游领域更注重事件的模式。

1.3 电力领域图谱研究现状

目前我国能源互联网建设正处于逐步推进的阶段，电力系统作为能源互联网建设的一大重点，其知识体系涉及发、输、变、配、用等多个领域，领域知识中的设备、人员、事件等存在复杂的关联性。知识图谱概念提出后，国内外

学者在电力领域基于知识图谱技术进行了相关研究。本文将从电力领域图谱总体研究现状和配电网领域图谱研究现状展开。

1.3.1 电力领域图谱总体研究现状

吴克河等（2019）提出新一代电力系统可以通过人工智能技术进行综合调控，使得系统能够对海量的信息进行萃取，从而更好地进行感知、分析、决策和控制。随着知识图谱概念的提出，近年来，国内外专家学者在电网拓扑结构、电网故障处置、电力设备管理等方向提出了基于领域知识图谱的电力系统应用。

乔骥等（2020）提出了面向电网调度故障处理的知识图谱总体框架，将操作规程、处置预案、调度细则等电网调度数据凝练为电力调度领域知识图谱，实现了基于知识图谱的故障信息解析判别、智能辅助决策和多维度人机交互；谭刚等（2020）针对问答系统中存在的中文电力问题表达模糊等问题，提出了一种融合领域特征的知识图谱智能电网客服问答系统 HDKG-QA，实现了以较低的响应延迟提供高准确率的回答；刘广一等（2021）从能源供应链协调优化控制问题出发，需要对能源互联网各个环节、要素、主体的数据进行联合建模，从而提出了“发输变配电网一张图”、“能源互联网一张图”的概念；王渊等（2022）提出了一种基于全业务统一数据中心的知识图谱构建方法，解决了国网公司业务数据跨专业贯通、数据资源的智能分析和管理工作；李旭斌等（2023）在其电力设备新型健康管理和智能运维体系架构中，提出了一种电力设备健康管理及智能运维知识图谱构建总框架，对提高电力设备健康管理水平，实现运维的精细化、智能化具有重要的促进作用。

此外，业界也对电力领域图谱进行了广泛的应用。善智互联在电力领域应用知识图谱、自然语言处理、深度学习、图计算等技术，整合电力设备信息、电力调度知识库等数据，提供设备智能运检、电力智能调度等解决方案；星环科技围绕电力设备、电力作业和电力员工三大核心构建电力知识图谱，提供基于多维源数据分析的设备运行及作业风险系统、基于工业互联网平台的新能源智能知识云图系统等解决方案。

1.3.2 配电网领域图谱研究现状

配电网作为电力系统中面向终端用户的最终环节，配电网智能化运维理论

上能够提升电力系统正常运转的能力，降低电网故障所带来的经济损失。目前，配电网领域结合图谱进行智能化升级已经从理论走向实践。

在配电网拓扑结构方向，高泽璞等（2020）通过知识图谱融合多个低压配电网信息系统中的数据，构建出低压配电网拓扑结构，实现了对低压配电网信息系统中的户变关系的辨识。在配电网故障处置方向，王骏东等（2021）针对配电网调度规则、故障预案等数据，提出了一种基于知识图谱的配电网故障辅助决策总体框架，并利用该图谱研发了故障辅助决策系统，实现了故障调控的快速、智能、准确的响应；叶欣智等（2022）更为细致地将配电网故障处置知识图谱构建框架分为设备拓扑、缺陷库、故障预案、调度规程四类图谱，并通过实体链接将四类图谱融合成整个框架。在配电网电压控制方向，姜磊等（2021）提出了一种基于电压控制策略知识图谱地检索方法及交互策略，有效地提升了配电网无功电压控制策略生成效率，提高了在不同场景的适用性。

1.3.3 电力领域图谱研究综述

综上所述，目前国内的电力系统已经朝向智能化方向发展，学术界研究方向包括智能调度、智能问答、智能设备管理等多个领域，并且在业界也已经有相关产品的出现，国外学术界和业界对电力领域图谱的研究比较少见。

电力领域图谱的研究现状可以归纳为以下两点：

（1）目前电力领域图谱在拓扑辨识、故障处置、设备管理等方向均有研究，但无论是整个电力领域还是聚焦于配电网领域，领域图谱大多基于知识图谱技术，而事件图谱在电力领域的研究较少。

（2）电力领域结合知识图谱的研究目前以技术框架和方法论为主，具体实现部分描述较少，缺乏电力领域图谱构建的具体案例。

1.4 综合评述

综上所述，随着知识图谱概念的提出，学术界和业界对知识图谱的研究从服务于智能搜索的通用知识图谱，逐渐下沉到细粒度、知识专业且相对封闭的领域知识图谱和事件图谱，并在特定领域中取得了一定的研究进展。

（1）国内外学术界和业界对知识图谱和事件图谱还尚无明确的定义，两者之间的界限也比较模糊。其中，知识图谱在结构上可以统一为三元组组成的网

状结构。而事件图谱作为知识图谱的衍生概念，尽管底层结构仍然是三元组组成的网状结构，但由于事件表示方法、事件粒度和等因素的不同，目前发展出各类不同的事件图谱形式，尚未出现明确的分类依据。

（2）目前在特定领域多见基于知识图谱的应用，而基于事件图谱的应用相对较少。其中，领域知识图谱更多结合静态实体，并在推荐系统、知识查询、问答系统等方向广泛应用；领域事件图谱更多结合动态事件，并在模式发现、事件预测、事件查询等方向应用。

（3）电力领域已在拓扑识别、故障处置、优化调度等方向将知识、经验等整合成领域知识图谱。其中，配电网作为电力系统用户接入端的重要环节，由于其业务流程和知识、经验相关性较高，更符合图谱的应用场景，因此提出的图谱构建技术框架和方法论较多。

但目前的研究方向还存在一些不足之处，具体可以分为以下四点：

（1）事件图谱本身定义和分类的模糊性给事件图谱的构建带来困难。

（2）知识图谱和事件图谱大多结合医疗、金融、旅游、公共安全等领域，电力领域图谱的相关研究还比较少。

（3）电力领域图谱大多以知识图谱作为底层技术支撑，在配电网故障处置这种事件性强的方向上缺乏以事件图谱作为底层技术的研究。

（4）配电网领域图谱大多围绕其构建和应用的技术框架和方法论，缺乏结合具体数据的案例描述和分析。

二、配网事件抽取关键技术

事件抽取概念比较宽泛，本项目在该阶段除事件抽取技术之外，一系列知识抽取技术还包括实体抽取（也称实体识别）、属性抽取（即事件论元抽取）、关系抽取等。针对目前拿到的故障单结构化数据，总体技术框架如下：

表 2-1 本项目知识抽取系列技术框架

技术名称	技术框架
事件抽取技术	事件模板框架
实体抽取技术	深度学习技术
属性抽取技术	事件模板框架
关系抽取技术	预定义事件图谱模式

其中，事件抽取和属性抽取技术通过事件模板框架可以一次性提取并整理成事件子图，在上阶段性汇报已提交，本次将在第一节简单说明；实体抽取技术选用 BERT-BiLSTM-CRF（基于双向 Transformer 编码器的表示模型-双向长短期记忆网络-条件随机场）模型；关系抽取技术通过预定义事件图谱模式（Schema）完成。

2.1 基于事件模板框架的配网事件知识抽取技术

目前数据为结构化数据，需要根据结构化数据中各字段的实际含义人工规定事件类型，包括故障原因事件、故障出现事件和故障结果事件。针对这三类事件设计事件模板框架如下：

表 2-2 故障原因事件框架模板

事件类型	故障原因事件	子事件类型	配网设备故障、用户内部故障等
论元角色	对应概念	角色类型	备注
故障设备	故障设备	参与者	无
故障现象	故障现象描述	参与者	无
责任原因	责任原因细类	参与者	无
关联线路	无	参与者	从“原因分析”中抽取
日期	无	属性	从“故障发生时间”中抽取
天气	天气	属性	无
气温	气温	属性	无

表 2-3 故障出现事件框架模板

事件类型	故障出现事件	子事件类型	线路开关跳闸、母线失地等
论元角色	对应概念	角色类型	备注
所属地区	所属地区	参与者	无
所属辖区	所属辖区	参与者	无
维护班组	维护班组名称	参与者	无
抢修班组	抢修班组	参与者	无
关联线路	故障概述	参与者	从“故障概述”中抽取
故障级别	故障级别	属性	无
故障发生时间	故障发生时间	属性	无
保护动作情况	保护动作情况	属性	无
重合闸情况	重合闸情况	属性	无
失地情况	失地情况	属性	无
日期	无	属性	从“故障发生时间”中抽取
天气	天气	属性	无
气温	气温	属性	无

表 2-4 故障结果事件框架模板

事件类型	故障结果事件	子事件类型	主干线全线停电、分支线停电等
论元角色	对应概念	角色类型	备注
关联线路	停电线路	参与者	无
停电公变数	停电公变数	属性	无
停电专变数	停电专变数	属性	无
影响低压户数	影响低压户数	属性	无
开始停电时间	开始停电时间	属性	无
恢复送电时间	恢复送电时间	属性	无
预计送电时间	预计送电时间	属性	无
日期	无	属性	从“故障发生时间”中抽取
天气	天气	属性	无
气温	气温	属性	无

其中，由于已经规定好了具体的事件、子事件、事件参与者、事件属性（论元角色），直接提取结构化数据中某些单元即可完成事件抽取、属性抽取以及部分实体抽取。

2.2 基于 BERT-BiLSTM-CRF 的配网设备实体识别技术

由于基于事件模板框架构建事件图谱的局限性，难以寻找事件间的时序关系，难以利用图结构算法，因此设计了一套全新的融合时序-因果双类型关系的事件图谱，本节将侧重于该图谱中最重要的部分，即设备实体识别技术。

本任务使用 BERT-BiLSTM-CRF 技术代替原有 BiLSTM-CRF 技术，加入预训练模型为词向量进行初始化，能够更好地将设备实体语义嵌入到向量空间中。模型架构见图 2-1。

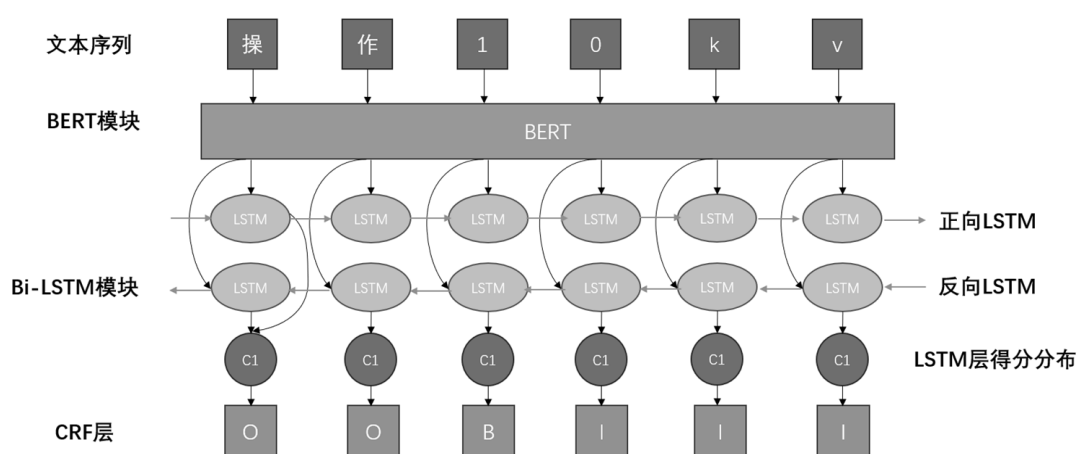


图 2-1 BERT-BiLSTM-CRF 模型架构

本任务实现的技术细节如下：

随机抽取获得 500 条合并整理好的数据，导入开源标注平台 Doccano，并

设置故障设备实体标签 LINE，对文本进行标注，标注准则为：标注的实体中应至少包括“馈线”和“具体位置”两部分，如“10kV 口星线#024 杆”，如果涉及更具体的设备应也进行标注，如“10kV 口星线#024 杆开关”。完成标注后，进行基于 BERT-BiLSTM-CRF 模型的配网故障设备实体识别操作。

利用 BERT-BiLSTM-CRF 三层结构进行文本标注，这三层的作用分别为：

①借助预训练模型 BERT 的特征抽取能力，经由“预训练+fine-tune”的经典范式，获取单词的嵌入，获得充足的上下文信息；②BiLSTM 再次捕捉语义信息，并输出单词对应每一类实体的得分；③CRF 考虑标签间的相互作用，从全局角度进行调优，并最终输出标注结果。具体包含以下步骤：

①新建 NerDataset 类对训练数据进行封装，主要用于将长文本拆分为多个短文本，并加入[CLS]、[SEP]、<PAD>标签，方便 BERT 模型进行学习训练。

②使用预训练的 BERT 模型来学习大规模文本数据的语言表示。BERT 以无监督的方式学习了丰富的上下文信息，这使得它在序列标注任务中能够更好地理解词语之间的关系。经过这一步骤，可以将单词转化为嵌入向量，这些嵌入向量保留了上下文信息和语义特征，使得模型能够更好地理解输入文本。此处预训练 BERT 模型的隐藏层维度为 768，dropout 率为 0.1，隐藏层层数为 12，注意力头数为 12，词表大小为 21128。

③在 BERT 层之后添加一个双向长短时记忆网络（BiLSTM）层。BiLSTM 能够捕捉序列中的上下文信息，因为它同时考虑了单词的前后文信息。这有助于模型更好地理解序列中的依赖关系和结构。LSTM 网络的核心在于通过门控机制实现长短期记忆，进而作用于序列信息捕捉。BiLSTM 是从前向和后向连个方向获得输出并进行连接，这种手段使 LSTM 网络具有了同时处理上下文信息的能力。

④使用条件随机场（CRF）层，这是一个标准的序列标注工具。CRF 层在整个标注序列中建模标签之间的依赖关系，确保生成的标注序列是一致的。这有助于解决标注错误传播的问题，提高模型的性能。

⑤模型充分利用有标签的数据进行训练，使用损失函数来衡量模型的标注与真实标签的差距，并使用反向传播算法来更新模型参数。

⑥计算模型的精确度、召回率和 f1 值，并记录，如果 f1 值超越历史最高

f1 值，则存储为最佳模型 `best_module`，否则继续重复步骤 A2，直到完成 50 个 epoch 的训练。实验结果表明，模型可以在已标注的小数据集上获得 0.9837 的 f1 值。

⑦使用训练好的模型对其他未标注数据进行实体识别，输出格式为：先输出文本段落，再输出列表格式的标注结果。以便对结果进行人工查验。人工查验的方法为：从原始数据集中随机筛选 100 条数据，与输出结果进行比对，确保模型在文本顺序、实体范围等方面没有明显问题，并且能取得质量较高的标注结果。

⑧移除一部分存在明显错误的实体，需要移除的实体包括：①结构不完整的实体，实体应包括“馈线”和具体“位置”，存在缺失的实体，如“开关 A 相”，“10kv 大甲线”，应该进行移除；②存在明显错误的实体，如地区名“莆田”“厦门”，无意义实体“110kv”等。为了保证格式统一，同时移除实体中多余的符号和重复字段。将处理之后的实体以列表格式保存输出，完成本次实体识别任务。通过本次标注，共获得了 191579 个标注结果，合计 163923 个实体。

总体来说，相比不使用 BERT 进行词向量初始化，使用后的模型 F1 值提高近 20%，效果能达到 98.37%。结果表明，使用 BERT-BiLSTM-CRF 可支撑目前新型事件图谱构建使用。

三、配网事件图谱构建及可视化方案报告

配网事件图谱构建的大量数据来自于故障单结构化数据，但大多数字段取值来源于系统内置表单选项，已经无法进行进一步处理（如：实体消歧、实体对齐等），但其优势在于数据量大，概率统计置信度高。

之前图谱构建研究方向使用基于事件框架模板的领域事件图谱构建技术，其核心思想在于：其一，一次性使用全量数据构建一个大规模事件图谱；其二，本体构建以二维表表头为重要参考依据；其三，由于在本体构建时考虑实例图的连通性，因此形成了需要考虑语义的复杂异质图。

但其缺陷在于，难以发挥事件图谱的优势。具体来说：该图谱的连通性依靠概念节点和线路节点来提供，导致节点抽象程度不同，直接应用图算法会忽略语义关系。此外，图谱中事件链固定，后续知识更新难度大。

因此，为了构建能够充分发挥事件图谱优势的图结构，需要开展新型事件图谱构建方式研究，新型图谱应满足如下两条优势：其一，表现形式有优势，能够表示事件的演化/发展模式或逻辑关系；其二，在不和 EKG 节点语义起冲突的条件下，能够应用图算法。基于以上背景和分析，提出问题及解决思路。

表 3-1 新型事件图谱构建问题及解决思路

问题	解决思路
事件链固定	因果固定，时序可扩展
复杂异质图难以在考虑节点语义的条件下应用图算法	降低图的复杂性
一次性构建大而全的图谱容易受到表头的限制且图规模过大	分而治之

由表 3-1 可知，新图谱的构建思路将采用“分而治之”的思想，并将不同类型图的复杂度降低，并针对以前因果事件链固定的缺陷重新设计时序可扩展链。本报告将聚焦于新图谱构建的原理、构建及可视化方案。

3.1 时序-因果双类型关系配网事件图谱构建原理

新型配网事件图谱将综合考虑时序和因果双类型关系，将图谱分为连通的三个关键模块——事理图谱模块（因果关系图）、时序事件链模块和设备关联图模块，具体说明见表 3-2。

表 3-2 时序-因果双类型关系配网事件图谱模块说明

图谱模块	模块说明	模块备注
事理图谱模块	小规模图，仅一个，带概率，提供全局因果信息	
时序事件链模块	小规模链，多个，不带概率，提供关联馈线视角时序信息和因果信息	事理图谱模块和时序事件链模块中的因果事件链存在包含关系，两者依靠事件链接相互关联
设备关联图模块	大规模图，仅一个，提供设备关联信息	时序事件链模块和设备关联知识图谱依靠设备相互关联

3.1.1 事理图谱模块

分析故障单结构化数据后，发现故障原因、故障类型、影响范围三列数据取值固定、具有因果关系。该部分仍然按照之前人工定义的因果事件链，即故障原因事件→故障出现事件→故障结果事件，见图 3-1。

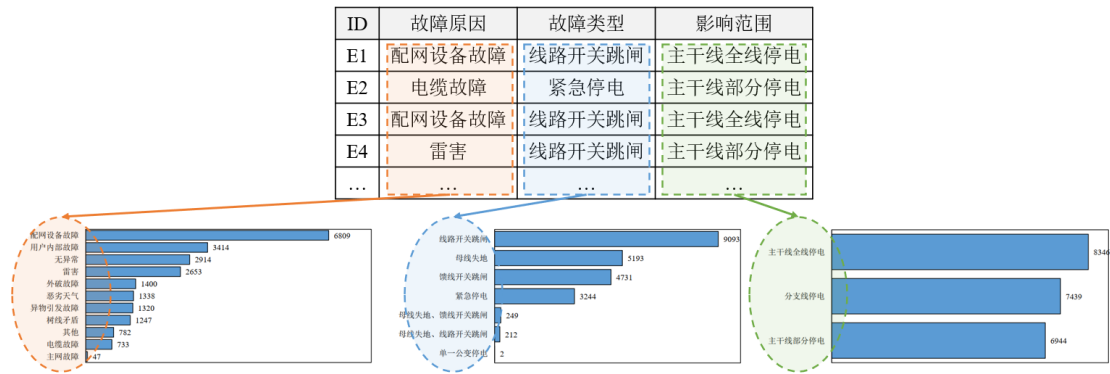


图 3-1 因果事件分析示意图

因此，原图谱中事件体系部分即为本图谱中事理图谱模块，且本模块中事理之间的转移可通过大数据的统计分析得到置信度高的概率信息，模块见图 3-2。

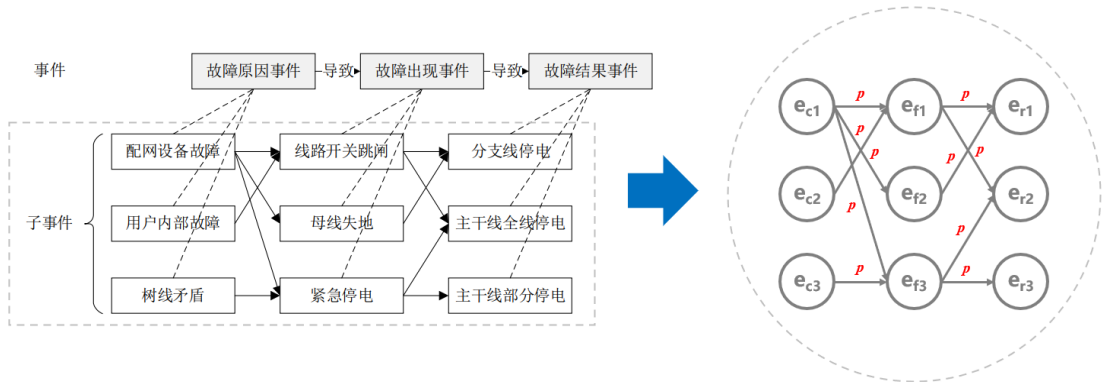


图 3-2 事理图谱模块示意图

其中，上图左侧为原图谱中构建的事件体系，各子事件间通过因果关系连接；右侧是新图谱中的事理图谱模块，本质上只是在原事件体系的基础上加入概率信息，没有对之前事件体系做变动。二者之间的关系为，左侧事件体系在原图谱中不出现，它是事件图谱本体的一部分；右侧事理图谱模块在新图谱中将出现在图谱中，它不是事件图谱本体的一部分。

3.1.2 时序事件链模块

以前的事件链基于人工定义的因果关系，链路节点难以扩展，后续分析挖掘工作较难进行。如图 3-3 所示，右侧绿色节点为左侧绿框中的三类因果事件，右侧四条事件链中在原图谱中，由于左侧绿框中因果事件模式的固定，只能包含 3 个绿色节点，而无法再扩展出其他红色节点。

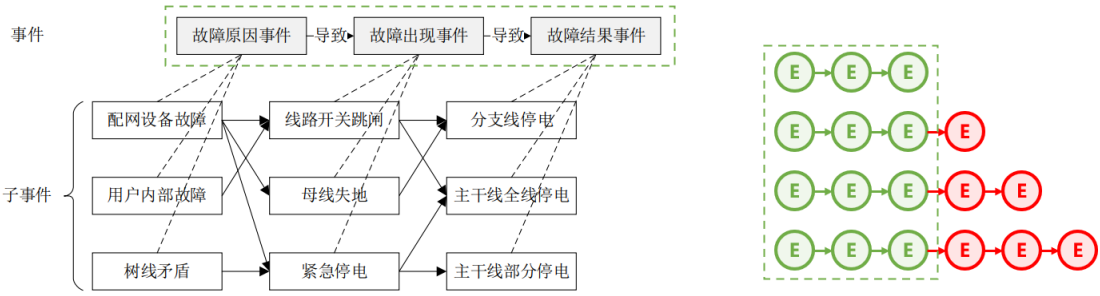


图 3-3 链路节点扩展示意图

因此，本研究中试图寻找一种不被结构化数据限制的新的事件链构建方式。具体来说，即从关联馈线视角出发，每条关联馈线可能在不同时间都会发生故障，按照关联馈线发生故障的时间顺序，将每条关联馈线发生的故障事件构成链，可以解决链路节点不可扩展的问题，最后再将这些时序链构成图，示意图见图 3-4。

ID	故障时间	故障原因	故障类型	影响范围	关联线路
E1	2022/9/13	配网设备故障	线路开关跳闸	主干线全线停电	A线
E2	2022/9/14	电缆故障	紧急停电	主干线部分停电	B线
E3	2022/11/10	配网设备故障	线路开关跳闸	主干线全线停电	A线
E4	2023/1/6	雷害	线路开关跳闸	主干线部分停电	A线
E5	2023/1/7	电缆故障	线路开关跳闸	主干线全线停电	B线
...

A线 E → E → E

B线 E → E

图 3-4 时序链构建示意图

按照时序关系可以将同一关联馈线对应的故障事件 E 连接起来，每个 E 都会包括三类因果子事件 e，因果子事件 e 之间构成因果事件链，示意图见图 3-5，左侧灰色结构是单个事件 E，对应故障单中一行数据，右侧彩色结构是时序事件链，每条事件链对应故障单中多行数据。

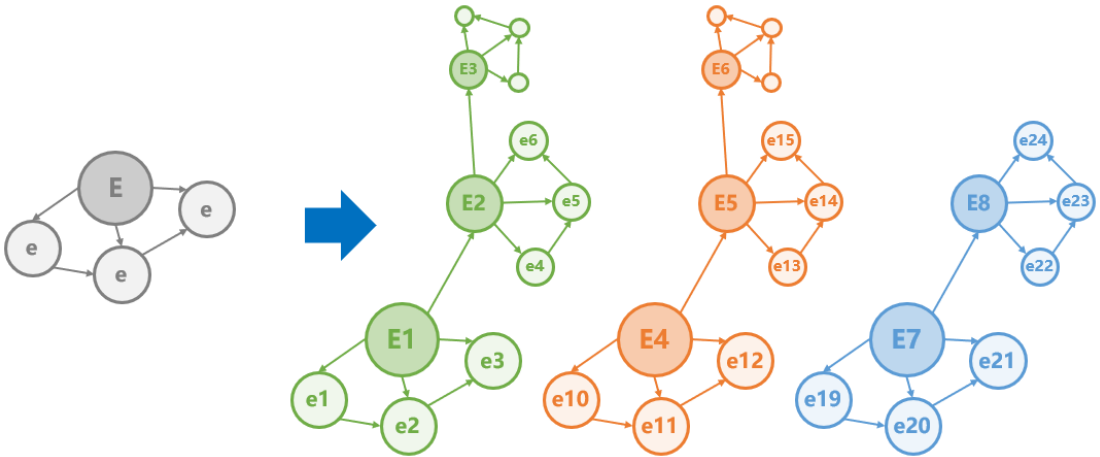


图 3-5 时序链模块示意图

其中，子事件 e 之间构成的因果事件链属于事理图谱的一部分，可以进行事件链接，建立起时序事件链模块和事理图谱模块之间的联系，见图 3-6。

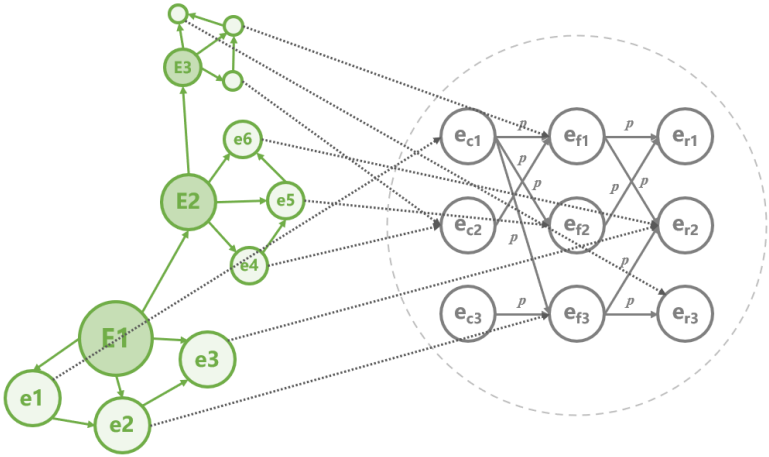


图 3-6 时序链模块和事理图谱模块联系示意图

3.1.3 设备关联图模块

事件链间的信息交互应该来自于结构化数据中不同行之间的数据联系。具体来说：联系产生自：概念实体（确定的，抽象的，无法更新的）和设备实体（不确定的，具体的，可更新的）。而概念实体由于其确定性（拿到结构化数据就能统计出有多少概念实体），将有很多其他实体或事件被归类（连接）到它们之上，因此度很大，只会形成树，难以形成图。设备实体由于其可抽取，且数量非常多，任何设备实体之间都可能产生关联，因此更有可能形成图。二者的比较见下表 3-3。

表 3-3 概念实体和设备实体比较

	概念实体	设备实体
实体类型角度	更像是其他实体或事件的类别，抽象程度较高	都是比较具体的线路名称
关系类型角度	概念实体和其他实体或事件连接时，是“从属关系”	设备实体和其他设备实体连接时，是“关联关系”

有表可知，用设备实体进行信息交互更为合适。从数据的角度考虑，设备实体也是唯一可能从结构化数据中找到不同行记录中各种关联关系，且存量较大，更有意义进行挖掘的一种实体类型（上篇报告中使用了 BERT-BiLSTM-CRF 进行

设备实体抽取就是为了解决本模块的构建)。

为了让时序事件链模块和设备关联图模块产生联系，需要从数据角度考虑，每行记录中都会包括一些非结构化数据（文本数据），从文本数据中抽取的设备实体之间可能会多次出现在不同行记录中，从而将事件链关联起来。设备关联图模块和时序事件链模块关联示意图见图 3-7。

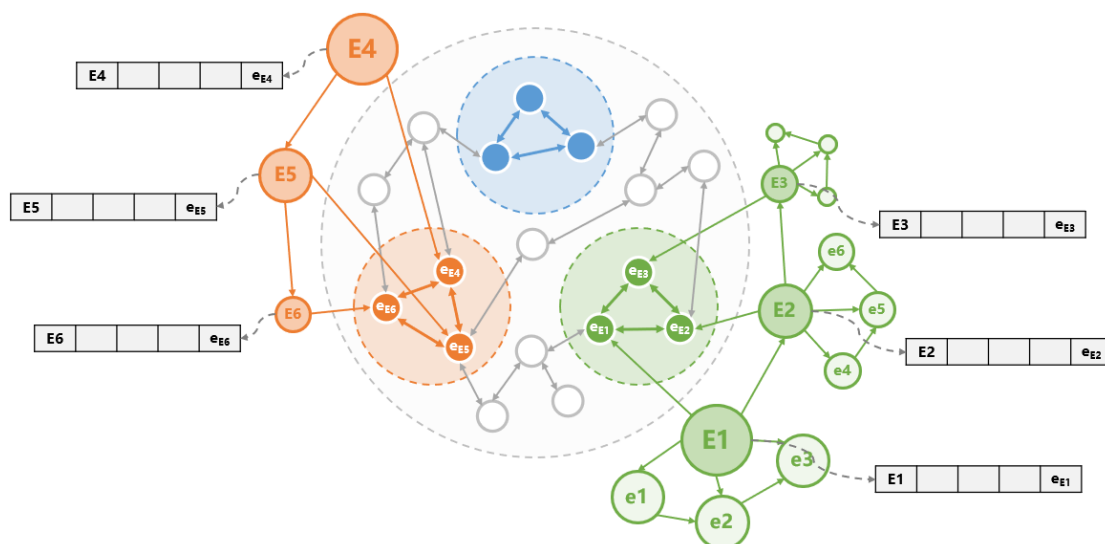


图 3-7 设备关联图模块和时序链模块联系示意图

其中，灰色圆形内部为设备关联图模块，游走在设备关联图模块周围的彩色链为时序事件链模块，同颜色的时序事件链和设备实体之间相互链接，由于不同事件可能连接到同一个设备实体，进而会将不同的事件链连接到一起。

3.1.4 时序-因果双类型关系配网事件图谱示意图

见图 3-8。

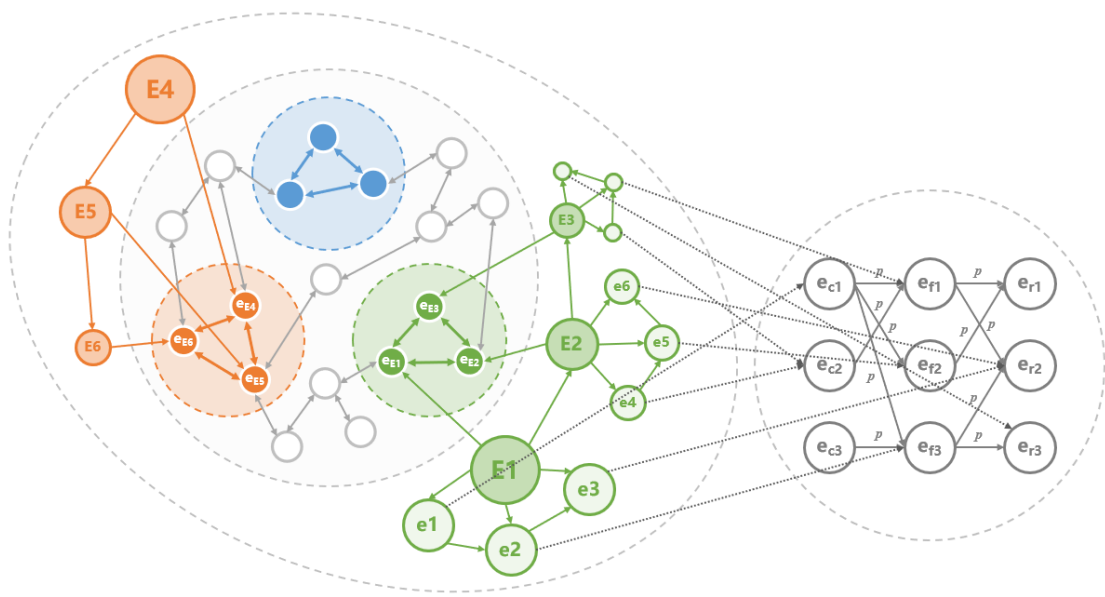


图 3-8 时序-因果双类型关系配网事件图谱示意图

3.1.5 时序-因果双类型关系配网事件图谱本体示意图

见图 3-9。

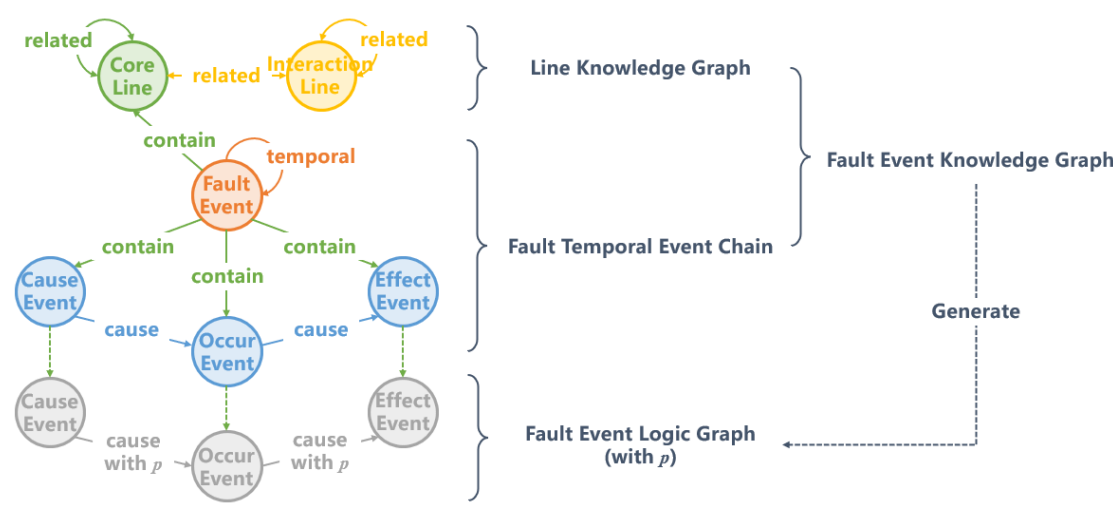


图 3-9 时序-因果双类型关系配网事件图谱本体示意图

由上到下为三个模块，设备实体关联图模块（Line KG）、时序事件连模块（Temporal Event Chain）和事理图谱模块（Event Logic Graph）。其中，前两个模块构建好之后，可以通过统计分析和图聚合生成第三个模块。

3.2 基于原图谱本体的新图谱 Schema 设计

新图谱 Schema 的设计参考原图谱事件模板框架部分，因此变动不大，具体见下表 3-5~3-6，最终 Schema 见图 3-11~3-12。

表 3-5 新图谱事件模板框架

事件类型	论元角色	对应概念	角色类型
故障事件	所属辖区	所属辖区	实体
	维护班组	维护班组	实体
	抢修班组	抢修班组	实体
	故障时间	故障发生时间	属性
	天气	天气	属性
	气温	气温	属性
故障原因事件	故障设备	故障设备	实体
	故障现象	故障现象描述	实体
	责任原因	责任原因细类	实体
	故障原因	故障原因	实体
故障出现事件	故障级别	故障级别	属性
	保护动作情况	保护动作情况	属性
	重合闸情况	重合闸情况	属性
	失地情况	失地情况	属性
故障结果事件	关联线路	停电线路	实体
	停电公变数	停电公变数	属性
	停电专变数	停电专变数	属性
	影响低压户数	影响低压户数	属性
	开始停电时间	开始停电时间	属性
	恢复送电时间	恢复送电时间	属性
	预计送电时间	预计送电时间	属性

表 3-6 新图谱关联三元组

头实体	尾实体	关系类型	关系描述
责任原因细类	责任原因	从属	属于
故障现象描述	故障现象	从属	属于
故障设备	故障设备分类	从属	属于

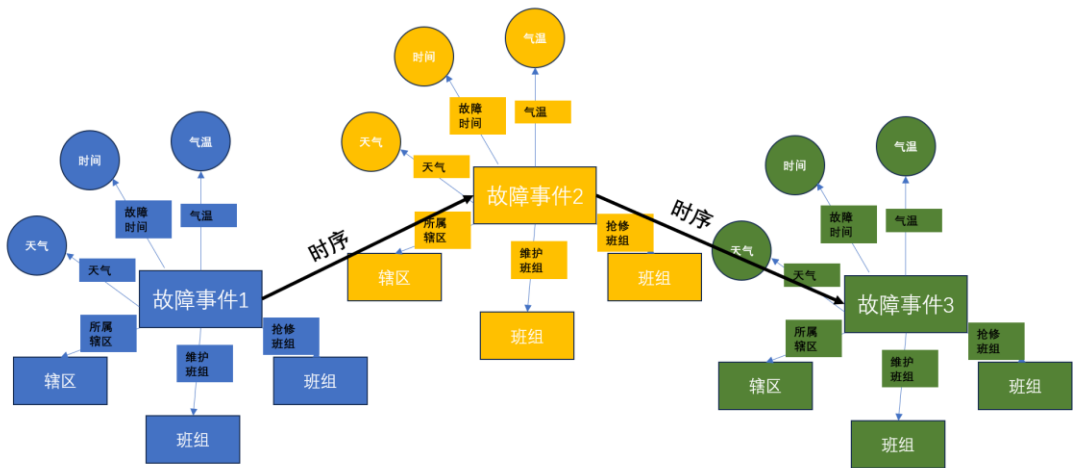


图 3-11 时序事件链部分本体图

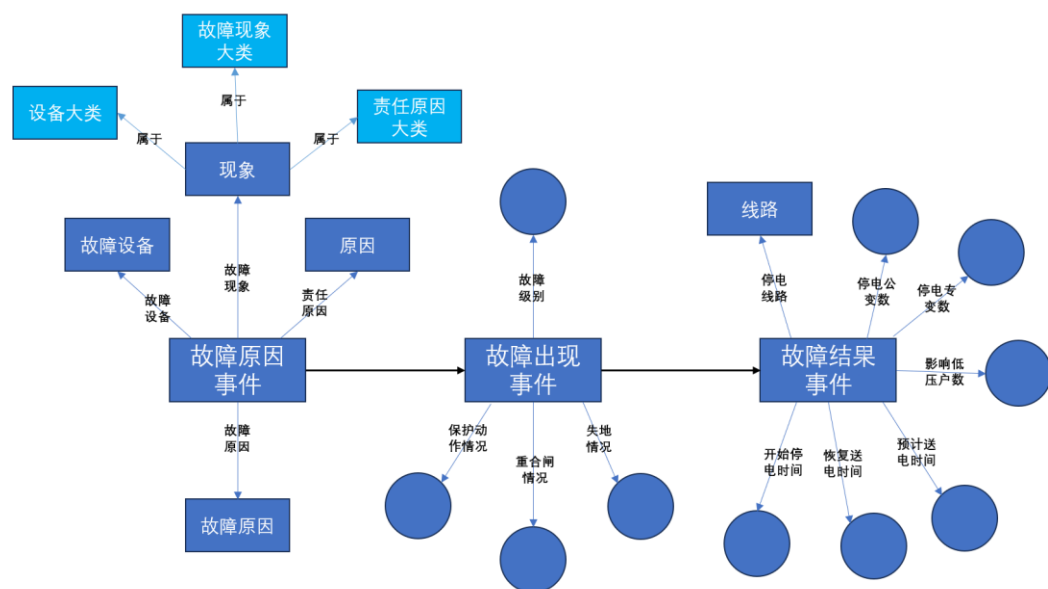


图 3-12 因果事件链部分本体图

3.3 图谱可视化设计方案

图谱可视化设计方案分为：图数据库可视化、图谱页面可视化。其中，图数据库可视化由于采用 Neo4j 作为本项目数据库选型，直接使用其自带 Web 界面进行可视化即可；图谱页面可视化需要结合最终项目的软件原型，根据需求进行可视化，这里主要描述技术选型。

（一）图数据库可视化

将图谱数据存入 Neo4j 图数据库中，通过图数据库查询语言直接进行查询，

其可视化效果见图 3-13~3-14。

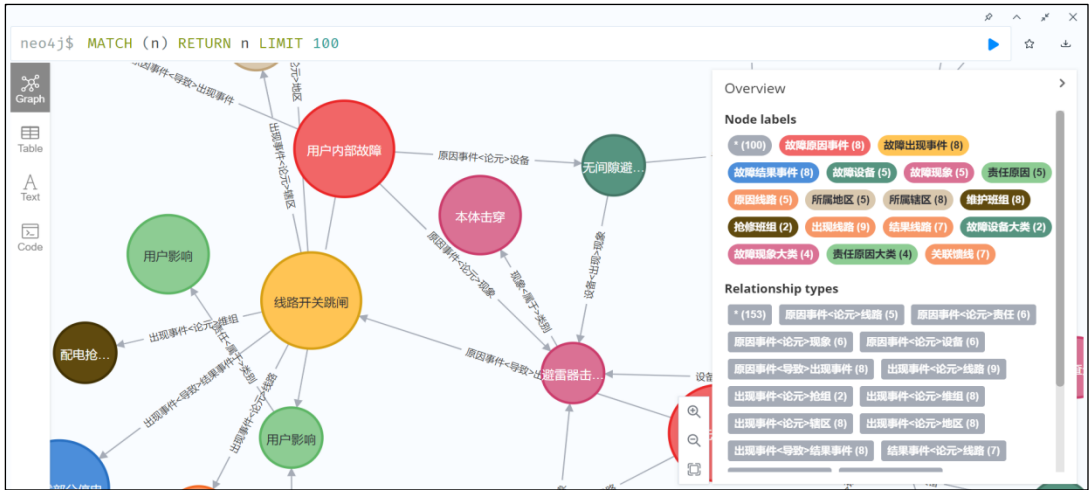


图 3-13 图数据库可视化案例-1

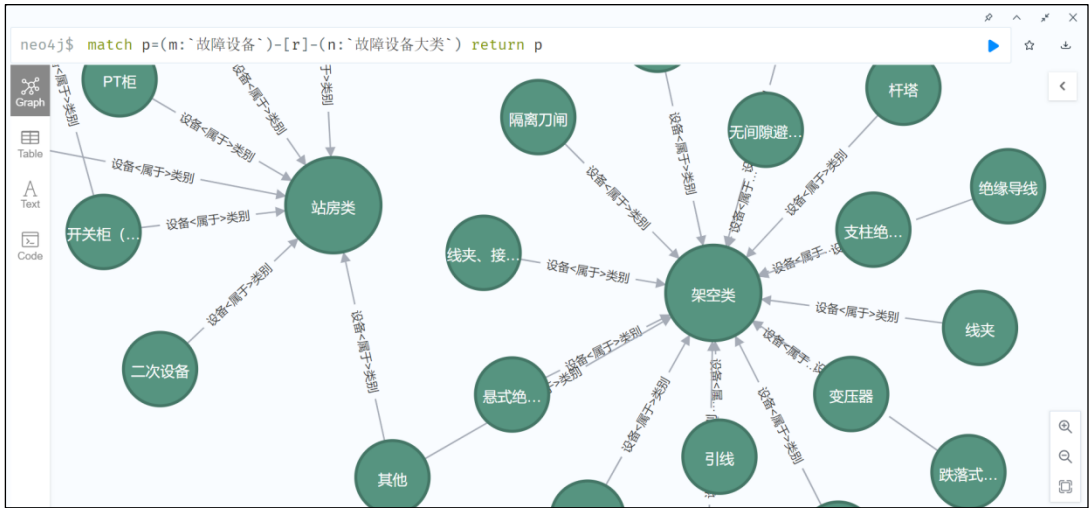


图 3-14 图数据库可视化案例-2

（二）图谱页面可视化

将图谱数据存入 Neo4j 图数据库中，后端编写图数据库查询语句并将算法封装进 Flask API 中，前端通过 Axios 异步技术请求 API 获取数据，并通过 ECharts 可视化技术进行力导向图可视化，其可视化效果见图 3-15。

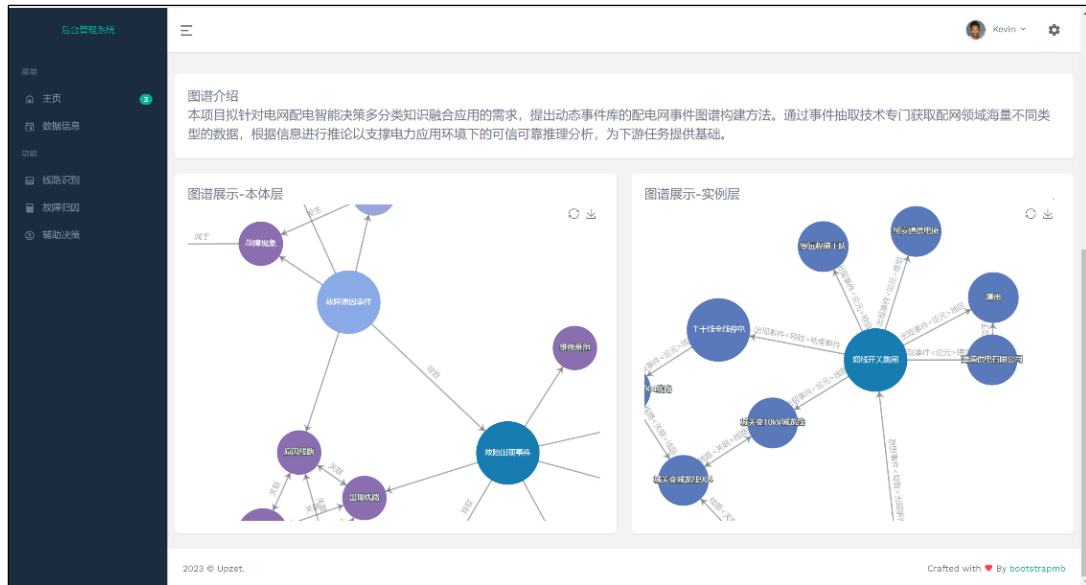


图 3-15 图谱页面可视化案例

四、配网事件图谱推理关键技术研究报告

目前存在的事件子图间连通性不强，但后续故障归因、定位、辅助决策都需要依靠不同事件案例作为历史数据做支撑，因此设备实体（原线路实体）之间需要产生联系。联系的产生分为两种：（1）属于同一记录的设备实体有联系；（2）不同记录之间可能存在语义相似的实体。本质上本任务首先要解决的是（2），因此需要研究设备实体对齐方法。

本研究报告包括目前事件图谱推理关键技术的设备实体对齐技术和基于事件图谱的配网故障归因定位及辅助决策方案（图关联分析和大模型生成）。

4.1 基于贪心前向匹配算法的自动词库构建技术

经过模式提取后的设备实体仍然关联性较差，这是由于实体识别时可能存在的误差，如识别出“大甲线*杆”和“大甲线线路*杆”，在语义上等价，但在做实体对齐时无法直接链接。通过 BERT、M3E 等预训练模型得到的词向量同样存在语义相似度低的情况，原因在于本研究的实体粒度较细，而预训练模型却是在实体粒度较粗的文本上进行训练的。

综上，本研究选用了准确度更高、更可解释的设备词库技术，将每个实体

进行分解得到子词（如“A线B支线C配变”将被分解为“[A线, B支线, C配变]”），再根据子词间的连接关系提高图连通性，进而提高事件关联性。

该方法能够通过多轮迭代的方式自动获取设备子词表和冗余信息表。其中，设备子词表可以对设备实体进行分解，冗余信息表可以去除深度学习实体识别过程中得到的多余字词，提高子词质量。具体来说，基于贪心前向匹配算法的自动词库构建算法如下：

（1）去除开头的“#”（理由：10kV 对于后续任务没用，且 kV 信息本质上被保存在文字中）；

（2）按“#”切割，取第一个切割单元；对所有实体（DATA1）进行如上操作得到词库 V1；

（3）统计 V1 中词频，绘制柱状图，在突变点截断（或阈值截断），得到词库 V2（如果|V2|比较小，人工检查并筛选提高词库质量），保存 V2（为提高词库质量，用最小长度的词来匹配最大长度的词，把大词切成小词（防止出现如“A线”和“A线B分线”出现次数都很多的情况），得到 V3，保存 V3）；

（4）贪心前向匹配算法，设置最大子串长度为“词库中最大长度词的长度”，当第一次匹配到该词就在该词后面加“#”并删除该词，对所有实体（DATA1）进行如上操作得到处理后实体并保存为 DATA2；

（5）用 DATA2 作为初始数据，重复 1~4，得到 V2'（V3'），直至 DATA 中无设备或设备较少，得到停用词库 VS，VS；

（6）合并 V2、V2' ...（V3、V3' ...），得到词库 V。

如下图 4-1 所示是用词库进行子词分解的结果，画红框处为停用词。

['上港线', '溪庄支线', '杆开关', '生源服装支线'],	
['上港线', '溪庄支线', '杆开关', '台支线', '熔断器', '跌落式'],	
['上港线', '溪庄支线', '杆刀闸', '段母线', '生源服装支线'],	
['上港线', '溪庄支线', '杆刀闸', '石古线', '枕头山变'],	
['上港线', '溪庄支线', '杆刀闸', '石古线', '枕头山变'],	
['上港线', '溪庄支线', '杆开关', '石古线', '分支线'],	
['溪庄支线', '杆刀闸', '石古线', '分支线'],	
['上港线', '溪庄支线', '变低压', '杆开关'],	
['上港线', '溪庄支线', '普贤变'],	
['上港线', '溪庄支线', '段母线', '普贤变'],	
['上港线', '溪庄支线', '池园线', '普贤变'],	
['上港线', '溪庄支线', '池园线', '普贤变'],	
['上港线', '溪庄支线', '盛达线', '普贤变'],	
['上港线', '溪庄支线', '盛达线', '普贤变'],	
['九一线', '溪庄支线', '田攸线', '普贤变'],	

图 4-1 基于设备词库的设备子词分解结果

4.2 基于事件图谱和图关联分析的推理技术方案

基于事件图谱进行推理，从而实现故障定位归因、辅助决策的方案要借助图关联分析，选择多个相关联的事件案例，后进行时序分析、规则挖掘、数据挖掘等，得到最终分析结果。下图 4-2 以“10kV 大甲线 05 杆”为例，介绍推理方案。

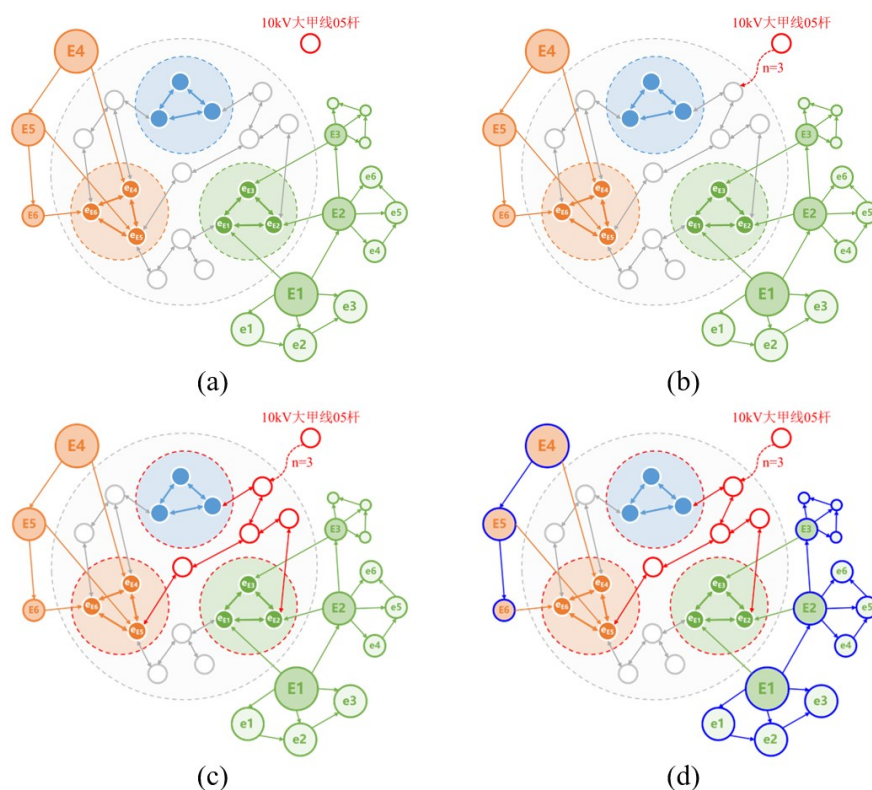


图 4-2 基于事件图谱和图关联分析的事件推理案例

假设目前“DMS 研判 10kV 大甲线 05 杆开关跳闸”，事件推理方案如下：

(1) 图 4-2(a)：对自然语言进行实体识别，确定“什么设备/线路出现故障/停电”（案例中识别出“10kV 大甲线 05 杆”），完成定位任务第一步；

(2) 图 4-2 (b)~图 4-2 (c)：将“10kV 大甲线 05 杆”进行基于启发式规则的模式提取、基于设备词表的子词分解，得到查询实体“大甲线”，在图谱中查询该线路，设置 $n=3$ ，查询 3 跳邻居，选择相关联的实体（图中红色边框），确定“什么设备/线路还可能出现故障/停电；什么设备/线路导致本设备/线路出现故障/停电”，完成定位任务第二步。此时定位任务结束。

(3) 图 4-2 (d)：根据选择到的各个实体确定多个事件案例（图中蓝色边

框)。故障归因任务可通过事件案例得到带概率的事理图谱子图(局部信息)和带概率的事理图谱(全局信息)完成,此外还可以对原因描述进行文本分析得到词云。辅助决策任务可以通过关联实体故障类型的统计分析进行推理,确定故障类型后通过“故障类型-处理方法”规程文档完成辅助决策。

此外,由于时序链的可扩展性优势,还可以进行事件链的模式挖掘,分析如“下次可能在什么时间出现什么类型的故障?”的问题,如果经统计“某线路多次发生故障”,可以生成分析报告,从而提前研判、辅助决策。

五、配网事件图谱故障定位归因及辅助决策应用方案

配电网络故障诊断对于确保电网平稳运行非常重要。配电网线路级故障诊断技术几乎不存在,一些使用知识图谱的技术也仅限于一些静态实体,忽略了更具体的线路和事件演化的动态逻辑。目前,国内广泛使用配电管理系统(DMS),其中存在海量的故障单数据。本节基于 3.1 节提出的一种实用的配电网故障诊断事件知识表示法——时序-因果双类型关系配网事件图谱,实现了配网事件图谱故障定位和归因应用。具体而言,我们在 3.3 节设计了一系列抽象模型来组织故障单,并构建了时序和因果事件知识图谱(TCEKG),该图谱可记录时序因果信息,并可与国内的 DMS 系统很好地集成。此外,我们还设计了两个基于 TCEKG 的故障诊断模型。为了使故障诊断模型更加关注近期事件,我们设计了一种用于过滤事件的时间衰减机制。在四个真实世界数据集上进行的广泛实验和演算研究表明,我们基于 TCEKG 的故障诊断模型可以有效地利用 TCEKG 完成配电网故障诊断任务,并且在单次推理中效率很高。

5.1 基于 TCEKG 的故障诊断模型

目前的配电网电力故障诊断方法都是基于现场勘查,基于 KG 的 FDM 也很少考虑线路级故障诊断。配电网线路之间存在天然的复杂相关性,在复杂相关性组成的图结构中可以获得更多相关信息,用于故障诊断。因此,我们首先提出了基于 TCEKG 的配电网故障诊断方法,并与国内的 DMS 进行了很好的集成,利用 TCEKG 实现了配电网线路级故障原因预测。

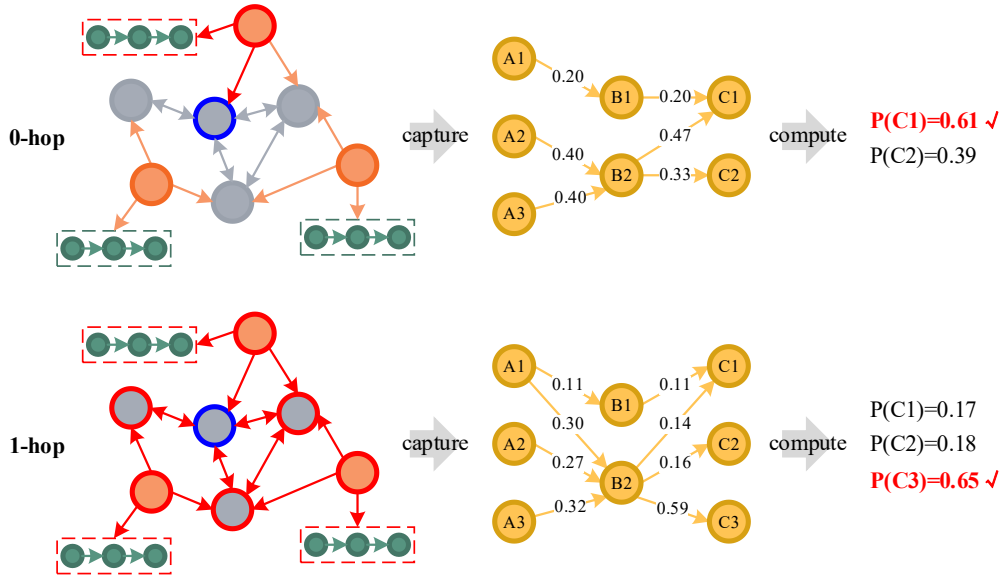


图 5-1 非 TCEKG 方法（上）和基于 TCEKG 的方法（下）

如图 5-1 所示，非 TCEKG 方法相当于 0 跳查询方法，而基于 TCEKG 的方法相当于 k 跳（ $k>0$ ）查询方法。由于多跳查询的时间成本较高，我们设置 $k=1$ 。我们的方法可以捕获更多的历史信息（即更多的 CEM），包括捕获更多的原因事件类型和更全面的概率信息，从而完成更准确的故障诊断。此外，考虑到事件之间的时间特性，我们在捕获信息的过程中引入了时间衰减机制，该机制将作用于每个 CEM。时间衰减函数如下：

$$\tau(t) = \exp \left\{ -\lambda \cdot \frac{|t - t_*|}{C} \right\}$$

具体来说，基于 TCEKG 的故障诊断模型（下称 TCEKG-FDM）是配网智能决策系统中的关键模块，TCEKG 通过图的形式组织故障单数据，形成结构化的配网故障事件库，通过命名实体识别模型识别问题中的线路实体，根据被识别的线路实体在图谱中寻找对应的锚点实体，并根据该锚点实体进行多条查询，收集历史信息，计算条件概率，从而完成故障原因的诊断。基于 TCEKG 的方法包括两个阶段，即捕获阶段和计算阶段。捕获阶段根据 TCEKG 执行多跳查询，捕获与特定线路相关的多个事件记录，计算阶段根据捕获的事件记录进行计算，最后输出预测结果。具体来说，计算阶段有两个可用的模型，即 TCEKG_{BASE} 和 TCEKG_{PROB}。TCEKG_{BASE} 将使用所有捕获的事件记录，通过时

间衰减函数为每个事件记录分配权重，并直接根据因果事件对最终结果进行分组和加权。TCEKG_{PROB} 将首先根据权重计算“结果事件-故障事件-原因事件”事件链的条件概率，然后进行带有时间衰减的分组加权求和，得出最终结果。

5.2 基于 TCEKG-FDM 的故障定位

配网故障定位任务是电力系统维护中的关键环节。它旨在迅速、准确地确定配电网络中故障的位置，以便及时采取维修措施，减少停电时间和影响。项目中故障定位任务被定义为 DMS 告警解析和故障类型识别。配网中的故障类型包括短路、断线、接地故障等。识别故障类型是定位故障的前提，通常通过故障指示器、保护装置和监控系统的报警信息来判断。

5.2.1 DMS 告警解析

DMS 告警解析是指基于自定义词库进行研判告警解析，旨在通过高质量线路词库，精准识别告警信息中存在的线路信息，便于后续在图谱中进行实体链接，从而完成故障类型识别及归因。DMS 告警解析主要由以下几个模块构成：

（1）数据录入模块

数据录入模块是用户输入 DMS 告警信息的输入框，可以通过直接输入或集成 DMS 系统。本项目选择前者，即“直接输入”的方案。该模块提供一个输入框，用户可以通过直接输入告警信息来进行数据录入。相比集成 DMS 系统的方案，直接输入的方式具有灵活性高、实施难度低、部署快捷等优点。



图 5-2 数据录入模块

（2）预定义词库模块

预定义词库模块预定义了一个通过 4.1 节技术构建的高质量线路词库，用于识别查询中的线路名称，确保告警信息解析的准确性和一致性。高质量词库共包含 10794 个通过命名实体识别模型提取、清洗的高质量线路词汇。

表 5-1 高质量词库示例（部分）

首台支	新县线	港南支	西畲支	龙华变
-----	-----	-----	-----	-----

清水线	竹岱线	竹坑线	上聂支	霞州线
梧塘变	秦溪线	赛专线	龚源支	中山变
松山变	庵洋线	大嶝变	湫口线	榕东线
贵后线	勋村支	河乾支	大溪变	山顶线

(3) 线路匹配模块

线路匹配模块根据高质量词库对查询中的线路信息进行精准匹配，便于后续获取 TCEKG 内的事件信息。

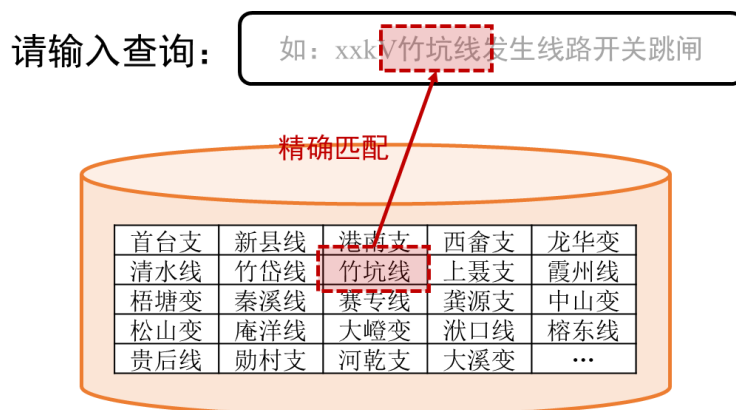


图 5-3 线路匹配模块

(4) 实体链接模块

实体链接模块根据识别出的线路对齐到 TCEKG 中，完成 DMS 告警文本数据到图数据的映射。在此过程中，由于高质量词库依赖于 TCEKG，因此通过高质量词库识别出的线路一定可以对应于 TCEKG 中设备关联图的某个实体上。

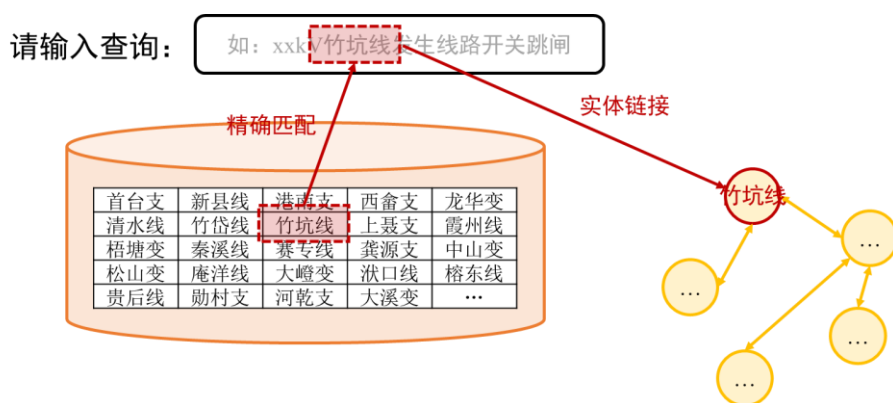


图 5-4 实体链接模块

5.2.2 故障类型识别

故障类型识别是指基于 TCEKG-FDM 进行故障类型识别，需要根据 DMS

告警解析结果，获取 TCEKG 中对应实体的 k 跳邻居，同时获取邻居上对应的事件链和事件信息，事件链中包含的对应邻居曾发生过的所有故障的原因、出现及结果事件，以及各个事件发生的时间；收集到这些信息后，会将其处理成可输入模型的数据格式，并输入到 TCEKG-FDM 中进行概率计算和结果预测。故障类型识别主要由以下几个模块构成：

（1）邻居捕获及数据生成模块

邻居捕获模块是根据对齐到 TCEKG 中的实体，获取该实体的关联实体，即 k 跳邻居，项目中考虑到配网故障响应的实时性，将 k 定为 1，即捕获 1 跳邻居。捕获邻居的同时将会把邻居对应的事件链上的信息加载到本地缓存中，并将其整理成便于模型读入的数据格式，该数据格式是一种级联图的形式，图 5-5 中 A 代表结果事件、B 代表故障事件、C 代表原因事件，“结果-故障-原因”节点之间的连接是一种级联图形式。

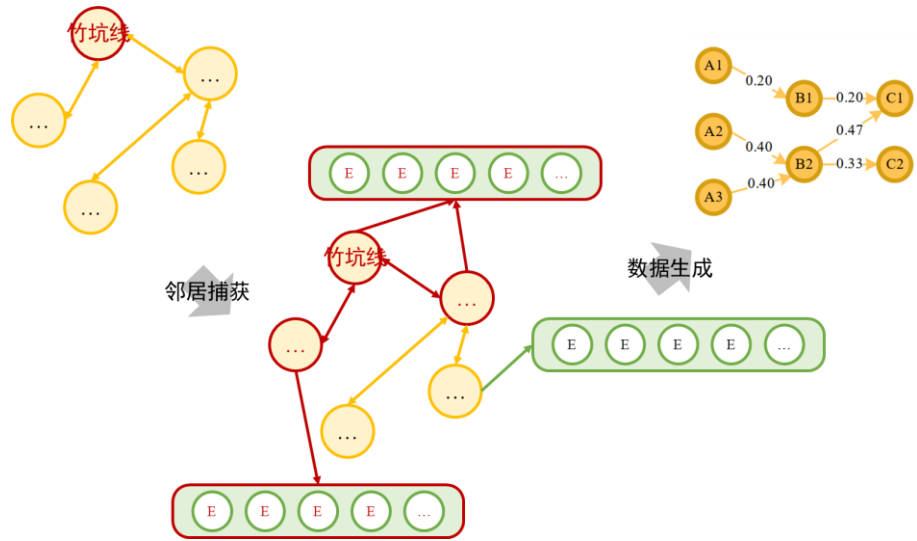


图 5-5 邻居捕获及数据生成模块

（2）TCEKG-FDM 故障类型预测模块

TCEKG-FDM 故障类型预测模块是通过将（1）中得到的数据输入到 5.1 节设计的模型中，通过概率计算、时间衰减等模块后预测故障类型的模块。模型架构及预测方式具体见 5.1 节。

5.3 基于 TCEKG-FDM 的故障归因

配网故障归因任务旨在通过深入分析和确定配电网络故障的根本原因，从

而制定相应的预防措施，提高配电网的整体可靠性和稳定性。这一任务在配网管理系统中至关重要，因为它直接影响到故障响应的效率和有效性。项目中故障归因任务被定义为故障原因预测和相似事件推荐。

5.3.1 故障原因预测

TCEKG-FDM 故障原因预测模块是通过将 5.2.2（1）中得到的数据输入到 5.1 节设计的模型中，通过概率计算、时间衰减等模块后预测故障类型的模块。模型架构及预测方式具体见 5.1 节。

5.3.2 相似事件推荐

配网故障诊断中的相似事件案例推荐是一种利用历史故障数据和先进的分析技术，来帮助快速定位和解决当前故障的有效方法。该功能通过识别和推荐与当前故障情况类似的历史事件，为运维人员提供有价值的参考信息，提升故障处理的效率和准确性。相似事件推荐主要由以下几个模块构成：

（1）TCEKG 图谱模块

TCEKG 图谱模块是 3.1 节构建好的图谱，存放在图数据库中。图谱中存放了历年的故障单数据，支持百万节点级别图数据库存储及流转，为了提高图谱数据检索的效率，我们将更为具体的事件信息当作属性节点存放在数据库中，其中节点及属性解析分析维度不低于 10 项。TCEKG 图谱的 Schema 是经过详细的分析设计后得到的，便于后续标签化模块进行检索。

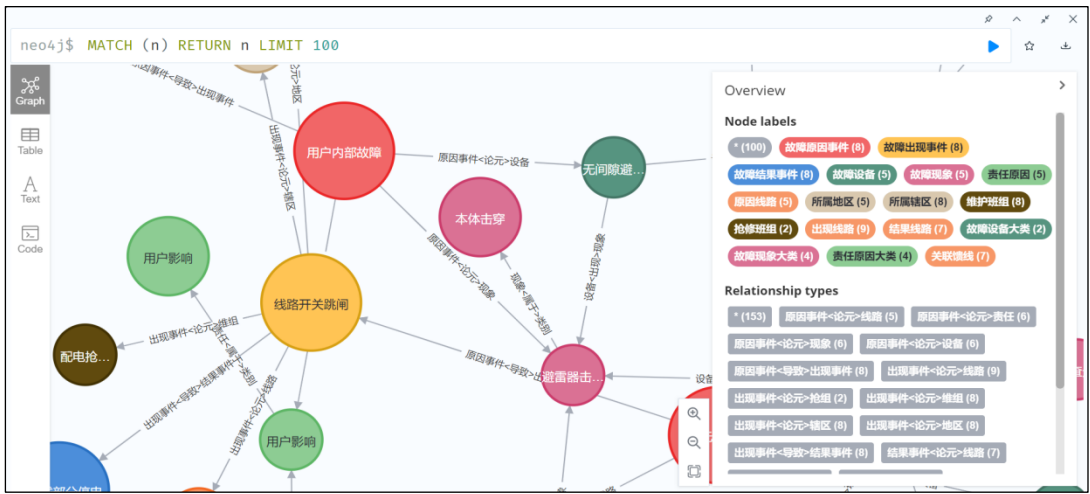


图 5-6 TCEKG 图数据库 Schema 及实例化

(2) 标签化模块

标签化模块将根据 DMS 告警信息解析后得到的线路和时间戳，进行标签化操作。标签化是指对线路、时间戳进行解析，解析粒度为年、月、日、季节、具体线路等，用于后续的具体线路级别的多维度匹配和推荐生成。具体的：

a. 数据解析：采用文本解析技术或正则表达式等方法，从 DMS 告警信息解析结果中提取线路和时间戳信息。

b. 时间转换：将时间戳转换为年、月、日等时间维度，或判断所属季节。

c. 标签生成：根据解析和匹配结果，生成相应的标签信息，标签信息会对应到图数据库的表模式中，以便于构建便于查询的标签化数据。

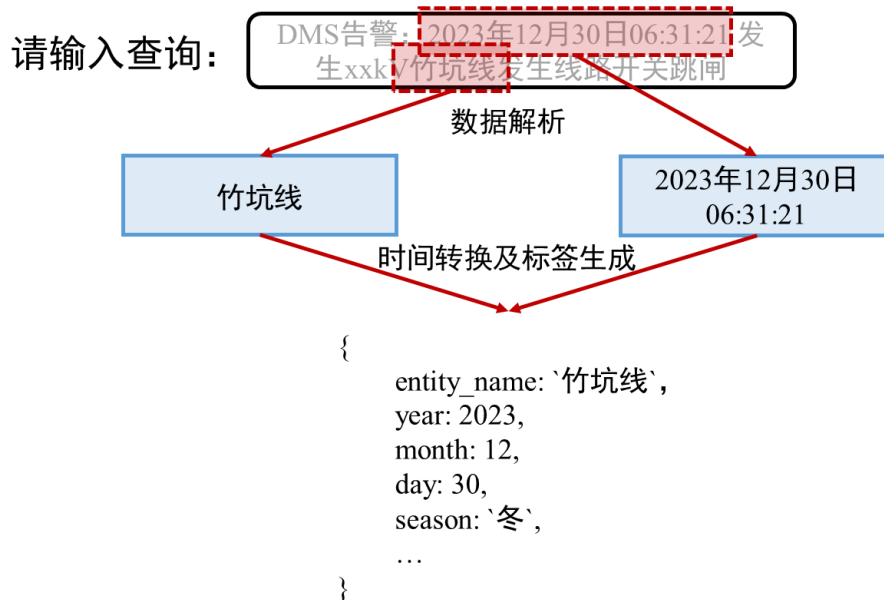


图 5-7 标签化模块

(3) 多维度匹配和推荐生成模块

多维度匹配的评分机制和推荐生成是配网故障诊断系统中关键的功能之一，它们帮助系统根据历史数据和实时信息，为当前故障事件生成合适的推荐方案。首先，针对每个历史故障事件和当前故障事件，计算它们之间的得分。得分计算可以基于以下维度进行：

a. 线路得分：对比当前故障的线路信息与历史故障事件的线路信息，计算线路得分；b. 时间得分：比较当前故障发生的时间与历史故障事件的发生时间，计算时间得分；c. 季节得分：判断当前故障发生的季节与历史故障事件的季节

是否相同，计算季节得分；d. 其他特征得分：根据需要，考虑其他特征（如环境条件、设备类型等）的得分。

根据得分计算结果，对每个历史故障事件进行加权评分。可以设置不同维度的权重，例如，线路得分权重较高，时间得分次之，季节得分等等。对加权评分进行排序，选取评分最高的若干个历史故障事件作为推荐的候选方案。根据多维度匹配的评分结果，生成推荐结果。推荐结果应包括以下内容：a. 历史故障事件描述：对于每个推荐的历史故障事件，提供详细的事件可视化子图；b. 推荐原因解释：解释为什么推荐这些历史故障事件作为当前故障的得分参考。

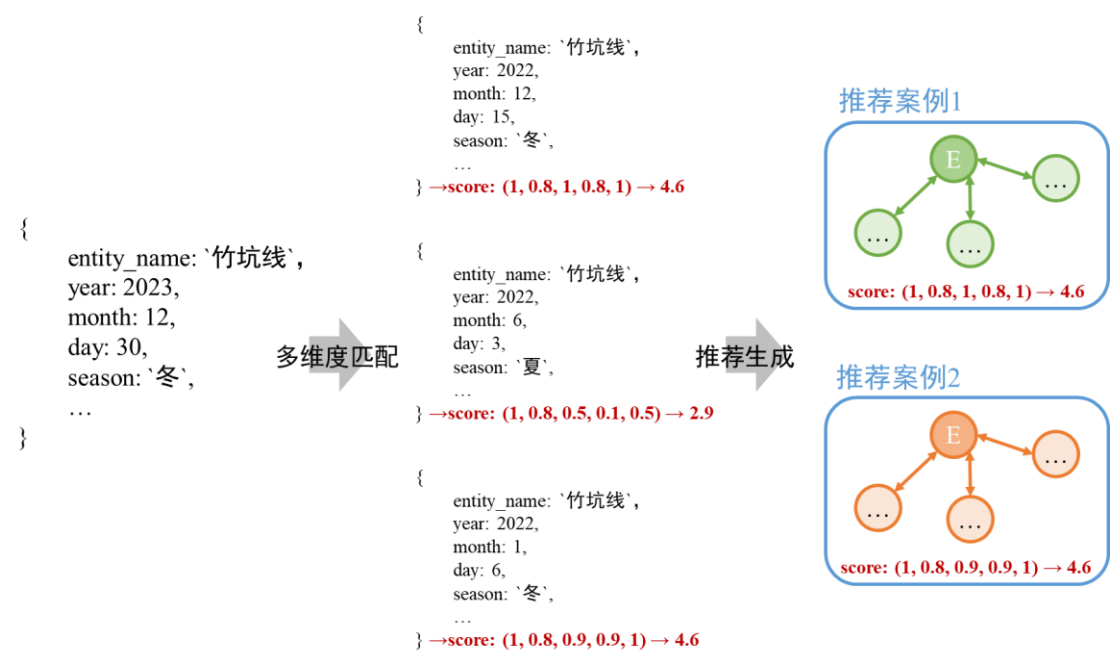


图 5-8 多维度匹配和推荐生成模块

5.4 基于 TCEKG-FDM 的故障辅助决策

配网故障辅助决策任务是指利用数据分析、模型计算和智能推荐等技术手段，辅助运维人员在配电网故障处理过程中进行决策，提高决策的准确性和效率。项目中故障归因任务被定义为规程知识库构建和故障辅助决策推荐。

5.4.1 规程知识库构建

规程知识库构建是按照 RAG (Retrieval-Augmented Generation) 中的知识库构建方式来构建的，可以采用分块、索引等技术来提高检索效率和推荐准确

性。规程知识库构建主要由以下几个模块构成：

（1）规程文档分块

规程文档分块是指将规程文档按照其中描述的故障类型及其对应的原因结构进行划分，使得每个分块包含一个完整的故障类型及其对应的处置方法。下面详细描述规程文档分块的过程：例如，可以按照以下结构进行分块：

a. 设备故障类：

- 电缆老化导致的故障：描述电缆老化的现象、影响和处理方法。
- 开关器件损坏引起的故障：描述开关器件损坏的原因和处理流程。

b. 线路故障类：

- 线路过载导致的故障：介绍线路过载的原因、检测手段和处理策略。
- 线路短路引起的故障：说明线路短路的可能性、影响和应对措施。

c. 环境因素类：

- 恶劣天气导致的故障：描述恶劣天气对配电网的影响和应急处理措施。

d. 人为因素类：

- 操作失误导致的故障：说明操作失误的原因、预防措施和纠正方法。
- 维护不当引起的故障：介绍维护不当的表现、影响和规范操作流程。

（2）文本块向量化

文本块向量化是将文本块转换为向量表示的过程，使得计算机可以理解和处理文本数据。下面详细描述使用 S-BERT 模型和 BM25 进行文本块向量化的过程，以及向量数据库的使用方法：

a. 使用 S-BERT 模型进行向量化：S-BERT（Sentence-BERT）是基于 BERT 模型进行预训练的文本表示模型，可以将文本转换为语义丰富的向量表示。

b. 使用 BM25 进行向量化：BM25（Best Matching 25）是一种基于文本检索的向量化方法，常用于计算文本块之间的相似度。

c. 向量数据库的使用：需要设计向量数据库的结构，包括字段定义、索引设计等，并在传统数据库或文件中存储向量数据，确保每个文本块都有对应的向量表示。对于需要查询的文本块，可以使用向量数据库的查询接口进行向量查询。查询接口会返回与查询文本块最相似的文本块向量，用于进行相似度计算和检索操作。

5.4.2 故障辅助决策推荐

故障辅助决策推荐需要利用 5.2 和 5.3 节的预测结果，并基于预测结果确定对应的规程文档的知识范围，并在该知识范围内使用向量检索相关的文本块并返回。具体来说：① 用户输入查询内容后，根据 5.2 和 5.3 节的预测结果，可以确定故障类型及原因，并在规程知识库中确定该类型及原因的范围；② 系统根据预定义的模板拼接用户的查询内容和故障类型及原因，并根据 BM25 和 SBERT 获取查询向量；③ 将查询向量与向量数据库中存储的知识结构向量进行余弦相似度计算；④ 根据相似度计算结果，找到与查询内容最相近的知识。如果相似度超过设定阈值，则生成推荐结果，包括故障处理流程、预防措施等内容；⑤ 将生成的推荐结果返回给用户，用于辅助用户进行故障决策和处理。

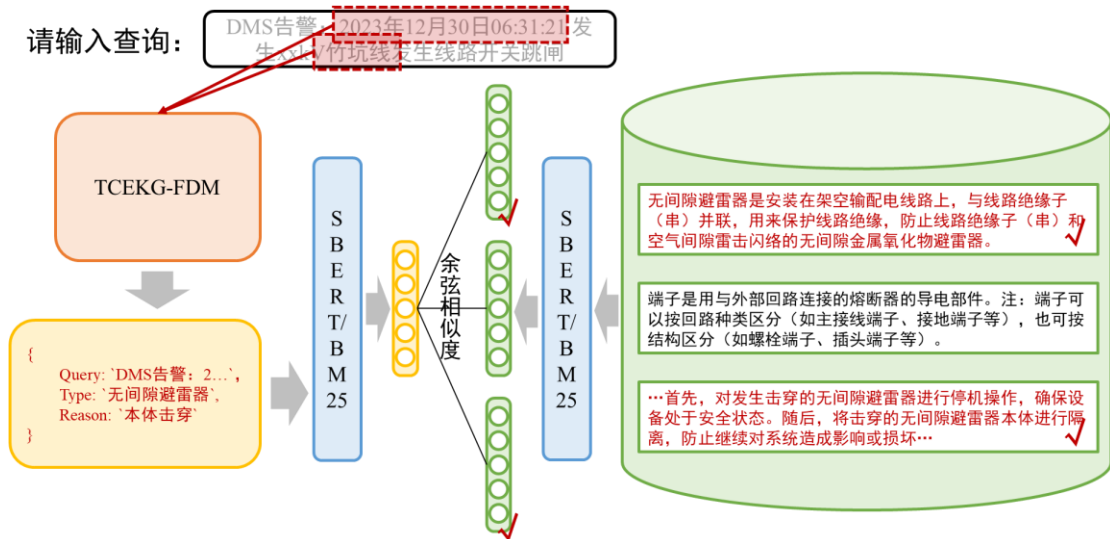


图 5-9 故障辅助决策推荐流程

参考文献

- [1] 漆桂林,高桓,吴天星.知识图谱研究进展[J].情报工程,2017,3(01):4-25.
- [2] 刘峤,李杨,段宏,刘瑶,秦志光.知识图谱构建技术综述[J].计算机研究与发展,2016,53(03):582-600.
- [3] 李涓子,侯磊.知识图谱研究综述[J].山西大学学报(自然科学版),2017,40(03):454-459.
- [4] 李涛,王次臣,李华康.知识图谱的发展与构建[J].南京理工大学学报,2017,41(01):22-34.
- [5] 刘忠宝,党建飞,张志剑.《史记》历史事件自动抽取与事理图谱构建研究[J].图书情报工作,2020,64(11):116-124.
- [6] 李纲,王施运,毛进,李白杨.面向态势感知的国家安全事件图谱构建研究[J].情报学报,2021,40(11):1164-1175.
- [7] 张海涛,李佳玮,刘伟利,刘雅姝.重大突发事件事理图谱构建研究[J].图书情报工作,2021,65(18):133-140.
- [8] 吴克河,王继业,李为,朱亚运.面向能源互联网的新一代电力系统运行模式研究[J].中国电机工程学报,2019,39(04):966-979.
- [9] 乔骥,王新迎,闵睿,白淑华,姚冬,蒲天骄.面向电网调度故障处理的知识图谱框架与关键技术初探[J].中国电机工程学报,2020,40(18):5837-5849.
- [10] 谭刚,陈聿,彭云竹.融合领域特征知识图谱的电网客服问答系统[J].计算机工程与应用,2020,56(03):232-239.
- [11] 刘广一,戴仁昶,路轶,杨小磊,汤亚宸,范宏,谭俊,刘婷婷.电力图计算平台及其在能源互联网中的应用[J].电网技术,2021,45(06):2051-2063.
- [12] 王渊,彭晨辉,王志强,范强,姚一杨,华召云.知识图谱在电网全业务统一数据中心的应用[J].计算机工程与应用,2019,55(15):104-109.
- [13] 李旭斌,田付强,郭亦可.新型电力系统中电力设备健康管理及智能运维关键技术探究[J/OL].电网技术:1-18[2023-02-27].
- [14] 高泽璞,赵云,余伊兰,罗永建,徐紫薇,张莲梅.基于知识图谱的低压配电网拓扑结构辨识方法[J].电力系统保护与控制,2020,48(02):34-43.
- [15] 王骏东,杨军,裴洋舟,詹祥澎,周挺,谢培元.基于知识图谱的配电网故障辅助决策研究[J].电网技术,2021,45(06):2101-2112.
- [16] 叶欣智,尚磊,董旭柱,刘承锡,田野,方华亮.面向配电网故障处置的知识图谱研究与应用[J].电网技术,2022,46(10):3739-3749.
- [17] 姜磊,战文华,张国彦,卢文华,曹磊,张辉,王骏东,张宇威.基于知识图谱的主动配电网电压控制策略[J].科学技术与工程,2021,21(30):12982-12989.
- [18] 徐增林,盛泳潘,贺丽荣,王雅芳.知识图谱技术综述[J].电子科技大学学

报,2016,45(04):589-606.

- [19] 刘知远,孙茂松,林衍凯,谢若冰.知识表示学习研究进展[J].计算机研究与发展,2016,53(02):247-261.
- [20] 胡志磊,靳小龙,陈剑赞,黄冠利.事件图谱的构建、推理与应用[J].大数据,2021,7(03):80-96.
- [21] 王毅,沈喆,姚毅凡,成颖.领域事件图谱构建方法综述[J].数据分析与知识发现,2020,4(10):1-13.
- [22] 蒲天骄,谈元鹏,彭国政,徐会芳,张中浩.电力领域知识图谱的构建与应用[J].电网技术,2021,45(06):2080-2091.
- [23] 闪鑫,陆晓,翟明玉,高宗和,徐春雷,滕贤亮,王波.人工智能应用于电网调控的关键技术分析[J].电力系统自动化,2019,43(01):49-57.
- [24] 李新鹏,徐建航,郭子明,李军良,宁文元,王震学.调度自动化系统知识图谱的构建与应用[J].中国电力,2019,52(02):70-77+157.
- [25] 刘梓权,王慧芳.基于知识图谱技术的电力设备缺陷记录检索方法[J].电力系统自动化,2018,42(14):158-164.
- [26] Ji S, Pan S, Cambria E, et al. A survey on knowledge graphs: Representation, acquisition, and applications[J]. *IEEE transactions on neural networks and learning systems*, 2021, 33(2): 494-514.
- [27] Hogan A, Blomqvist E, Cochez M, et al. Knowledge graphs[J]. *ACM Computing Surveys (CSUR)*, 2021, 54(4): 1-37.
- [28] Fan Y, Wang C, Zhou G, et al. Dkgbuilder: An architecture for building a domain knowledge graph from scratch[C]//*Database Systems for Advanced Applications: 22nd International Conference, DASFAA 2017, Suzhou, China, March 27-30, 2017, Proceedings, Part II* 22. Springer International Publishing, 2017: 663-667.
- [29] Kejriwal M. Domain-specific knowledge graph construction[M]. *New York: Springer International Publishing*, 2019.
- [30] Abu-Salih B. Domain-specific knowledge graphs: A survey[J]. *Journal of Network and Computer Applications*, 2021, 185: 103076.
- [31] Zhang Y, Sheng M, Zhou R, et al. HKGB: an inclusive, extensible, intelligent, semi-auto-constructed knowledge graph framework for healthcare with clinicians' expertise incorporated[J]. *Information Processing & Management*, 2020, 57(6): 102324.
- [32] Shi D, Wang T, Xing H, et al. A learning path recommendation model based on a multidimensional knowledge graph framework for e-learning[J]. *Knowledge-Based Systems*, 2020, 195: 105618.

- [33] Zhu Y, Zhou W, Xu Y, et al. Intelligent learning for knowledge graph towards geological data[J]. *Scientific Programming*, 2017, 2017.
- [34] Glavaš G, Šnajder J. Event graphs for information retrieval and multi-document summarization[J]. *Expert systems with applications*, 2014, 41(15): 6904-6916.
- [35] Rospocher M, Van Erp M, Vossen P, et al. Building event-centric knowledge graphs from news[J]. *Journal of Web Semantics*, 2016, 37: 132-151.
- [36] Li Z, Zhao S, Ding X, et al. EEG: knowledge base for event evolutionary principles and patterns[C]//*Social Media Processing: 6th National Conference, SMP 2017, Beijing, China, September 14-17, 2017, Proceedings*. Springer Singapore, 2017: 40-52.
- [37] Gottschalk S, Demidova E. Eventkg: A multilingual event-centric temporal knowledge graph[C]//*The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15*. Springer International Publishing, 2018: 272-287.
- [38] Ding X, Li Z, Liu T, et al. ELG: an event logic graph[J]. *arXiv preprint arXiv:1907.08015*, 2019.
- [39] Ma Y, Wang Z, Li M, et al. MMEKG: Multi-modal Event Knowledge Graph towards universal representation across modalities[C]. *Association for Computational Linguistics*, 2022.
- [40] Rudnik C, Ehrhart T, Ferret O, et al. Searching news articles using an event knowledge graph leveraged by wikidata[C]//*Companion proceedings of the 2019 world wide web conference*. 2019: 1232-1239.
- [41] Wu J, Zhu X, Zhang C, et al. Event-centric tourism knowledge graph—a case study of Hainan[C]//*Knowledge Science, Engineering and Management: 13th International Conference, KSEM 2020, Hangzhou, China, August 28–30, 2020, Proceedings, Part I 13*. Springer International Publishing, 2020: 3-15.
- [42] Bajracharya R, Wallace R, Aasman J, et al. Entity Event Knowledge Graph for Powerful Health Informatics[C]//*2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)*. IEEE, 2022: 456-460.
- [43] Guan S, Cheng X, Bai L, et al. What is event knowledge graph: a survey[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022.
- [44] Neches R, Fikes R E, Finin T, et al. Enabling technology for knowledge sharing[J]. *AI magazine*, 1991, 12(3): 36-36.
- [45] Grishman R, Sundheim B M. Message understanding conference-6: A brief history[C]//*COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics*. 1996.

- [46] Doddington G R, Mitchell A, Przybocki M A, et al. The automatic content extraction (ace) program-tasks, data, and evaluation[C]//*Lrec*. 2004, 2(1): 837-840.