

一种图智能应用 开发平台及电信运营商应用实践

A Graph Intelligent Application Development Platform and Application Practice of Telecom Operators

贺鸣¹, 郭熹², 秦守浩¹, 张珂珂² (1. 中国联通研究院, 北京 100048; 2. 联通在线信息科技有限公司, 北京 100032)
He Ming¹, Guo Xi², Qin Shouhao¹, Zhang Keke² (1. China Unicom Research Institute, Beijing 100048, China; 2. Unicom Broadband Online Co., Ltd., Beijing 100032, China)

摘要:

图是由实体及其关系构成的非线性数据结构,广泛应用于社交网络、金融风控、生命科学等多个领域。研发了一种全流程的图智能应用开发平台,并基于平台能力构建运营商光缆网资源图谱,研发光缆网资源自动链路拼接应用。该平台为图技术的应用提供了开发工具,降低了企业相关研发工作的技术门槛,可用于包括电信运营商在内的多个领域。

关键词:

图;图计算;图神经网络;知识图谱;光缆网;链路拼接

doi:10.12045/j.issn.1007-3043.2024.10.014

文章编号:1007-3043(2024)10-0073-05

中图分类号:TN919

文献标识码:A

开放科学(资源服务)标识码(OSID):



Abstract:

Graph is a nonlinear data structure based on entities and their relationships, widely existing in various fields such as social networks, financial risk control, and life sciences. It introduces a full-process graph intelligent application development platform. Based on the platform capability, the resource map of telecom operators' optical cable network is constructed and the automatic link splicing application of optical cable network resources is developed. This platform provides development tools for the application of graph technology, reduces the technical threshold of relevant research and development work of enterprises, and can be used in different fields, including telecom operators.

Keywords:

Graph; Graph computing; Graph neural network; Knowledge graph; Optical cable network; Link splicing

引用格式:贺鸣,郭熹,秦守浩,等.一种图智能应用开发平台及电信运营商应用实践[J].邮电设计技术,2024(10):73-77.

0 前言

在当今的数字化转型浪潮中,各行各业积累了海量的数据。随着业务逻辑的日益复杂以及数据来源的多样性,如何有效地对数据进行可视化管理以及挖掘数据之间的深度关联价值成为数字化转型的重要方向。图(Graph)是基于实体及其关系构成的非线性数据结构^[1],现实世界中很多复杂系统可以使用图的相关特性进行描述和分析。图数据建模具有简洁形象的优点,能更清晰地构建复杂的数据模型,有助于

知识检索和价值挖掘。同时,图数据建模也是后续对图模型进行图计算和图训练的基础。由于图相关智能技术目前尚处于快速发展阶段,面对日益增多的图智能应用场景,其学术研究与商业实践之间存在一定的差距。因此,便捷有效且功能丰富的开发工具成为重要需求。

1 图概念及相关技术发展

1.1 图概念

1.1.1 概念

图是一种描述数据之间复杂关系的非线性数据结构。它由节点和边组成,其形式化定义为 $G=(V,E)$,

收稿日期:2024-09-11

其中 V 表示对象或实体的节点集合, E 表示对象或实体之间的边集合^[1]。此外,节点和边还可能具有相应的标签,这些标签用于表示对象、实体或关系的属性信息。根据边是否有方向进行划分,图可分为有向图和无向图两大类;而根据边是否带有权值进行划分,图又可分为有权图和无权图^[2]。在日常生活中,图数据广泛应用于社交网络、内容网络以及交通网络等多个领域。在社交网络中,节点是指每个具体的用户,边是指用户之间的互动关系,例如微信的社交网络可以看作由微信用户(节点)和好友关系(边)构成的图。在铁路交通网络中,节点可以是分散在全国的火车站,边是连接各个火车站之间的铁路轨道。在互联网内容网络中,节点可以是每一个网页,边可以是网页之间的超链接关系。

1.1.2 特点

随着各行各业的数字化转型加速推进,复杂数据之间的关联价值更为突出。关系型数据模型以实体为核心进行构建,缺乏对实体之间关系的直接支持,需要通过关联表间接地维护实体之间的关系。与关系型数据相比,图数据模型以节点和边代表实体和关系,对实体之间的关系提供了原生支持,表达更为直观且结构更为灵活^[3]。因此,在数据种类繁多、需可视化呈现、关联维度和关联价值较高的应用场景中,更适合开发图数据应用。

1.2 图技术及产业应用

1.2.1 技术方向

Gartner预测,到2025年,图技术将被应用于80%的数据和分析创新领域,从而促进企业快速决策^[4]。图技术主要包括图数据库、图计算、图神经网络等一系列基于图的认知技术。

图数据库是面向关联关系图数据的存储数据库,它通过顶点、边和属性来表示和存储数据,同时支持数据的增加、删除、查询和修改等操作。目前,主流的图数据库主要采用属性图(Property Graph)和资源描述框架(Resource Description Framework, RDF)这2种抽象模型表达图数据。属性图使用节点、关系和属性表示图的信息,被广大图数据库厂商所采用。RDF使用三元组表示节点的属性或节点间的关系,它是W3C的一项标准。目前主流的图数据库包括Neo4j、NebulaGraph、ArangoDB和TuGraph等^[5]。

图计算是指所有基于图数据进行的分析计算,其核心在于图计算算法。常见的图计算算法包括遍历

等图分析类算法、最短路径等模式匹配类算法、三角计数等社群发现算法和PageRank等中心性算法。图计算技术解决了传统计算模式下关联查询的效率低、成本高的问题,具有完整的关系刻画能力,以及丰富、高效和敏捷的数据分析能力^[6]。

图神经网络将传统深度学习方法扩展到图数据领域,通过在神经网络模型中添加图操作,对图数据进行学习,提取和发掘图数据中的特征和模式^[7]。大部分深度学习模型是在图片、文本等欧几里德结构数据(Euclidean Structure Data)上进行信息的提取和挖掘,但图数据属于非欧几里德结构数据(Non-Euclidean Structure Data)^[8],传统的深度学习模型无法直接应用于图数据。常见的图神经网络模型包括图卷积网络(Graph Convolution Networks, GCN)、图注意力网络(Graph Attention Networks, GAN)和图生成网络(Graph Generative Networks, GGN)等^[9]。

1.2.2 产业应用

随着图技术的不断发展,它在知识图谱、社交网络、金融风控、网络拓扑分析等领域都得到了广泛应用。知识图谱是图数据的一类下游应用,是一个存储知识实体与实体间关系的结构化网络,适用于各行业和垂直领域,是支撑起各领域智能搜索、智能客服、智能决策、智能营销和直观可视化等智能应用的基础^[10]。在社交网络领域中,可通过图计算算法分析社群结构、用户行为和传播路径等,这些信息可应用于用户画像、个性化推荐、舆情监测等应用场景^[11-12]。在金融风控领域中,通过关联客户身份、行为和社会关系等相关信息,金融机构能够精准识别虚假申请、恶意申请、套现、诈骗等各种犯罪行为,为贷款管理、反欺诈、违规团伙识别等场景提供决策参考,从而降低金融风险^[13-14]。在电网拓扑分析等领域中,针对电网拓扑数据所具有的网络化特征,可采用图数据库和图计算技术来存储和分析电网拓扑数据,进而促进电网营销配一体化应用的研发进程^[15-16]。

1.2.3 电信行业的应用展望

电信行业存在网络资源拓扑(光缆网、无线网等)、人机物交互拓扑(客户关系网、通信行为图谱等)和行业知识图谱(运维知识图谱、企管知识图谱等)等三大类图谱,具有图技术应用的优势。在复杂的自智网络场景中,引入图技术将为网络规建维优营的场景提供智能、高效的决策辅助能力。具体而言,可面向电信运营商的O域(运营域)、B域(业务域)等多维度

数据,开发网络、用户、知识图谱等图数据模型,结合图计算和图神经网络等智能算法,实现网络优化及用户体验优化的智能运营管理。

由于图技术目前尚处于快速发展阶段,为了提高图数据建模和智能分析的效率,运营商可建设一种全流程的图智能应用开发平台,提供端到端的图化建模、图存储、图计算和图训练等功能。该平台支持关系型数据到图数据的转化建模,支持图计算和图训练能力,实现全流程图应用的开发。该智能应用开发平台将图技术能力原子化,可供上层业务调用,实现图应用开发流程标准化,不仅降低了使用门槛,还提高了推广效率。

2 图智能应用开发平台简介

2.1 平台整体框架

如图 1 所示,全流程的图智能应用开发平台整体框架由资源管理模块、图数据管理模块、图分析引擎模块、图训练管理模块、系统管理模块、运维监控模块和模型应用模块构成。该平台为图智能应用的开发提供支撑环境。

资源管理模块包括资源层和资源编排层,提供系统所需的 IT 资源和容器资源。资源层包括 CPU 服

器、智能算力服务器、网络资源等存储、计算和通信资源。资源编排层支持 Docker 容器、Kubernetes 容器集群管理、Harbor 镜像仓库等功能,实现对基础资源的管理和调度。

图数据管理模块包括数据接入、源数据管理和数据处理功能,用于准备图计算和图神经网络模型训练所需的图数据。数据接入功能支持数据上传、数据采集等方式。源数据管理功能支持对源数据的列名重命名,列名和实体、关系、属性的映射配置,支持实体数据、关系数据的增删改查操作以及图数据的存储。数据处理功能支持对图数据的清洗和特征加工等操作。

图分析引擎模块提供包括图神经网络的计算框架和常见算法以及对图数据的常见分析功能。并行计算功能支持数据并行、模型并行和管道并行等方式。计算框架支持 Deep Graph Library (DGL)、PyTorch Geometric (PyG)、Paddle Graph Learning (PGL) 等。图神经网络算法包括图卷积网络 (GCN)、图注意力网络 (GAT) 等。图数据分析功能支持在同构图、异构图、有向图上的常见图计算算法,如路径查询、社区发现等。

图训练管理模块包括模型开发、模型训练和模型部署三大部分。模型开发支持在线开发、自动开发和本地模型上传等方式。模型训练支持数据配置、模型配置和可视化训练等功能。模型部署支持将平台上所管理的模型,按照与推理环境相匹配的方式部署到指定环境中。

模型应用模块支持对已训练完模型的调用,并能以指定的接口形式与其他业务应用集成。

系统管理模块支持用户管理、权限管理、操作审计等功能,保障用户的操作在系统可控范围内。

运维监控模块包括系统监控和系统告警 2 部分。系统监控部分支持 IT 资源监控和进程管理。系统告警部分支持告警规则配置和告警管理功能。

2.2 端到端应用流程

如图 2 所示,图智能应用开发平台的端到端应用流程,主要包括图模型准备、数据准备、图数据建模和图分析应用 4 个关键环节。

2.2.1 图模型准备

创建图模型时,需定义图模型的具体结构信息,包括实体和关系的定义,即图谱的节点和边。如果实体和关系有相对应的属性,则进一步定义实体属性和关系属性。具体定义方式可以选择新定义实体和关

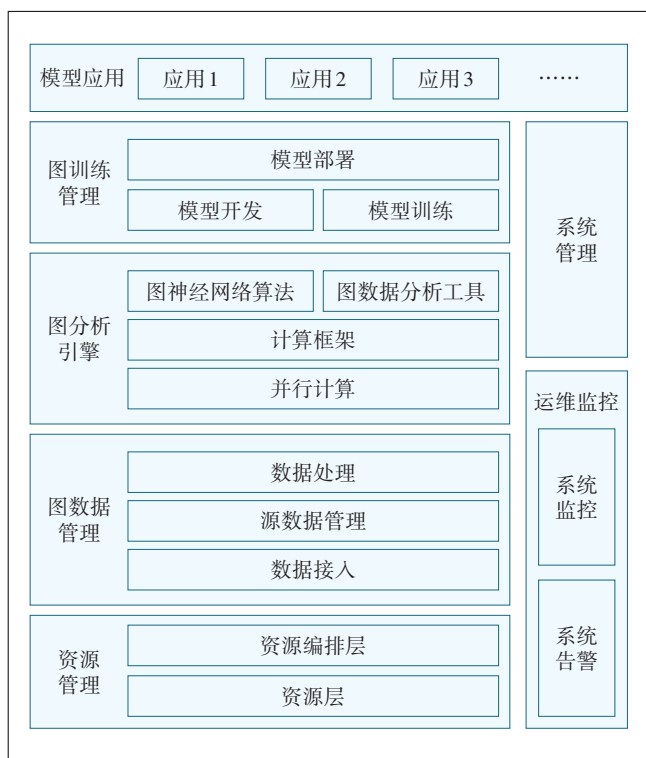


图 1 图智能应用开发平台整体框架

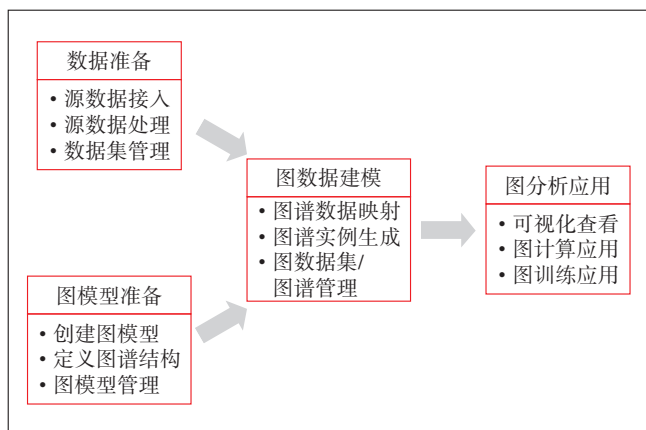


图2 图智能应用开发平台的端到端应用流程

系,或者选择已定义好的图模型进行参考复用。将新定义好具体结构信息的图模型纳入图模型管理,便于对配置的图模型进行存储、查看和复用。

2.2.2 数据准备

在接入原始数据时,可采用本地上传离线数据文件和在线数据采集2种接入方式,采集源包括MySQL、FTP、Kafka、HIVE、Elasticsearch等。针对接入的数据集,可根据实际需求来判断是否需要数据集进行进一步的处理。如果无需处理,该数据集则直接作为源数据集纳入管理;若需处理,则进行源数据处理操作,处理完作为新的源数据集纳入管理。常见简单数据处理操作包括异常字符去除、数值填充、数据去重、字段重命名等。

2.2.3 图数据建模

根据场景需求选择图模型和对应的数据文件,创建图谱实例,并配置图谱实例信息,实现图谱实例的图模型和数据关联融合。将映射完成的图谱实例数据集纳入图数据集管理,便于对创建的图谱实例数据集进行存储和查看。然后,将图谱实例数据集导入图数据库,从而完成图谱实例的生成。生成的图谱实例可纳入图谱管理。

2.2.4 图分析应用

根据业务需求,选择已开发好的图谱或者图数据集,进一步选择图分析操作,这些操作包括图拓扑可视化查看、图计算和图训练。图拓扑可视化操作可基于已开发好的图谱,通过执行相应的图数据库指令实现。择图计算操作可基于已开发好的图谱,通过选择相应的图计算算法模型,如遍历算法、中心性算法和社区发现算法等,对图数据的实体及关联关系进行分析处理。图训练操作可基于已开发好的图数据集,通

过上传或者在线开发具体的图神经网络模型(如图卷积网络、图注意力网络、图生成网络等)对图数据的特征信息进行提取和学习。

3 图智能应用开发平台应用实例

3.1 应用场景

光缆网是网络建设的基础底层部分,为各类业务网提供传输通道,同时也为网络智能化、服务差异化等方向的发展提供有力保证。随着网络规模的不断扩大,如何保障光缆线路既广泛覆盖又高效利用,成为了亟待解决的关键问题。光缆网拓扑结构多样,建设规划方案众多,如何综合考虑多项因素,择取最优方案成为难点。目前,一线作业人员往往仅根据本地需求和资源进行决策,缺少对整体光缆网结构和使用情况的把握,同时各作业人员之间缺乏有效的经验沟通,无法将建设思路和方案进行梳理固化及分享。

通过引入图智能技术,开发光缆网的虚拟模型,增强物理网络所缺少的系统可视化、仿真、优化等能力,可实现对光缆网的链路拼接、路由优化等功能,从而提升光缆网的自智运维效能。

3.2 应用实践

基于图智能应用开发平台的图数据建模能力,我们开发了光缆网资源图谱。以省分实际关系型数据为基础,从中提取了省级光缆网资源数据,经过数据清洗及建模,定义图模型结构,包括机房、机架、光交接箱、光分纤箱、光缆接头、设备室外安置点等实体和相应属性,以及光缆段等实体之间的边和相应属性,从而完成资源数据图谱化。实际光缆网资源图谱节点的数量达700多万,边的数量达400多万,通过图数据建模能力实现光缆网拓扑的可视化展现,点击节点或者边可查看相应的实际属性信息。

基于图智能应用开发平台的图分析应用能力,本文实现了基于光缆网资源图谱的链路拼接。应用流程主要包括4个步骤。

第1步,选取已构建的省级光缆网资源图谱,选择链路拼接的起始节点与结束节点,通过图智能应用开发平台提供的智能路径检索的图计算能力对光缆网资源图谱进行分析。

第2步,实现自动链路拼接,按链路总长度或者跳数排序展示多种拼接方案。

第3步,点击具体方案可查看对应的拼接路由。

第4步,点击具体拼接路由的相应节点或边,可查

看对应的实际属性信息。

现网的链路拼接主要通过人工进行,部分省分公司基于关系型数据开发的本省拼接工具的效率非常低下,且只能支持较少跳数,缺乏统一的应用平台支撑光缆网链路拼接业务。实测效果显示,采用基于图智能应用开发平台的光缆网资源自动链路拼接应用可自动提供拼接方案,支持高跳数拼接,与非图工具相比,其拼接效率提升约百倍。

基于关系型数据拼接工具和基于图智能应用开发平台能力进行链路拼接的应用效果对比情况如表 1 所示。总之,采用关系型数据建模方案,做业务支撑时关联表繁多,多跳拼接和拓扑可视化的能力较差,链路拼接效率低下,且无法直接进行应用功能拓展;基于平台构图能力采用图数据建模方案,支持全图资源拓扑可视化,能更生动形象地展示资源拓扑结构,原生支持多跳拼接且拼接效率大幅提升,同时可基于平台图计算和图训练能力拓展多样的业务支撑应用。

表 1 光缆网资源链路拼接应用效果对比

拼接方式	基于关系型数据拼接工具	基于图智能应用开发平台能力
拼接操作	① 选择拼接区域 ② 选择起始节点与结束节点 ③ 执行链路拼接	① 选择资源图谱 ② 选择起始节点与结束节点 ③ 执行链路拼接
多跳拼接	需要多张表关联查询	原生支持多跳拼接
可视化	无法实现全局资源拓扑可视化	支持全图资源拓扑可视化
拼接效率	低效,6跳以上拼接容易失败,5跳拼接耗时约110s	高效,8跳拼接耗时小于1s
功能扩展	功能单一,无法直接进行功能拓展	基于资源图谱可调用其他图计算能力进行分析,如环路检测、中心性分析等;基于图数据可进行图神经网络模型训练或者推理应用

4 结束语

本文介绍了一种全流程的图智能应用开发平台及应用实践。首先,给出了图概念,并简要分析了图数据特点、图技术和相关产业应用,其次展望了图技术在电信行业中的应用前景。随后,介绍了图智能应用开发平台的实现,该平台支持关系型数据到图数据的转化建模,还具备图计算和图神经网络模型训练能力。面向电信运营商光缆网资源的感知、预测与管控运维场景,基于平台能力进行了光缆网资源图谱的构建及链路拼接应用的实践,以解决复杂网络场景下的资源调配问题,有效提升网络运营的效率 and 智能化水

平。

参考文献:

- [1] 刘梦雅,刘燕兵,于静,等.图数据分析系统计算模型综述[J].计算机应用研究,2017,34(11):3204-3213.
- [2] 黎俊虹.跨平台统一分布式图计算框架研究及系统实现[D].南京:南京大学,2021.
- [3] 沈志宏,赵子豪,王海波.以图为中心的新型大数据技术栈研究[J].数据分析与知识发现,2020,4(7):50-65.
- [4] Gartner. Gartner发布2021年十大数据和分析技术趋势[EB/OL]. [2024-03-24]. https://www.gartner.com/cn/newsroom/press-releases/gartner_2021_.
- [5] 清华大学人工智能研究院,北京智源人工智能研究院,清华-中国工程院知识智能联合研究中心.人工智能之图数据库[EB/OL]. [2024-07-20]. <https://static.aminer.cn/misc/pdf/graphDB.pdf>.
- [6] 清华大学人工智能研究院,北京智源人工智能研究院,清华-中国工程院知识智能联合研究中心.人工智能之图计算[EB/OL]. [2024-02-27]. <https://static.aminer.cn/misc/pdf/GraphComputing.pdf>.
- [7] 赵港,王千阁,姚烽,等.大规模图神经网络系统综述[J].软件学报,2022,33(1):150-170.
- [8] 马帅,刘建伟,左信.图神经网络综述[J].计算机研究与发展,2022,59(1):47-80.
- [9] WU Z H, PAN S R, CHEN F W, et al. A comprehensive survey on graph neural networks[J]. IEEE transactions on neural networks and learning systems, 2021, 32(1):4-24.
- [10] CB Insights. 2022中国图计算技术及应用发展研究报告[EB/OL]. [2024-05-08]. <https://zhuanlan.zhihu.com/p/561321062>.
- [11] 曹杰.基于社交网络的用户人格预测研究[D].合肥:安徽大学,2022.
- [12] 杨佩佩.基于图神经网络的在线社交网络恶意用户检测[D].桂林:桂林电子科技大学,2021.
- [13] 饶逸卓.基于知识引导图神经网络的欺诈检测方法研究[D].北京:军事科学院,2022.
- [14] 董静,韦云健,李丹丹,等.基于图计算技术的车险团伙欺诈识别研究[J].保险理论与实践,2022(7):72-85.
- [15] 韩赫,张沛超,柴博,等.基于图计算的区域热电力系统建模与运行优化方法[J].中国电机工程学报,2022,42(19):7113-7126.
- [16] 周爱华,裘洪彬,高昆仑,等.基于图数据库的电网拓扑分析技术研究[J].电力信息与通信技术,2018,16(8):23-27.

作者简介:

贺鸣,毕业于西安电子科技大学,工程师,硕士,主要从事人工智能平台、网络创新应用等方向的研发工作;郭熹,毕业于华中科技大学,高级工程师,硕士,主要从事分布式计算、神经网络并行训练等方向的研发工作;秦守浩,毕业于北京邮电大学,助理工程师,硕士,主要从事微服务、云原生以及分布式软件的开发工作;张珂珂,毕业于华中科技大学,高级工程师,硕士,主要从事通信计费、能力开放等方向的研究工作。