



DATE ACQUISITION AND ANALYSIS OF B5G6G NETWORK INTELLIGENCE..

摘要

随着日益复杂的移动通信网络和各行业差异化的服务需求，超越5G (B5G) /6G最重要的研究方向和趋势是将人工智能、大数据和计算深度融合到网络中，并开发智能通信网络以实现自动驾驶网络能力和精准定制服务。将数据与智能嵌入网络，以开发具有内生智能以及计算与通信融合特征的新型网络智能技术体系，在架构、数据和人工智能算法方面提出了挑战。数据作为最基础和重要的部分，决定了网络智能性能的上限，而架构和算法只是达到这个上限的手段。因此，我们需要解决如何获取、分析和利用移动大数据的最根本问题，并建立一个全面的数据获取和分析系统，以推动新型B5G/6G网络智能技术体系的发展。本白皮书介绍了一种B5G/6G网络智能的数据获取和分析系统。该系统实现了数据层信息集成、信息层知识提取和表示、智能层知识计算和推理、应用层特征定制。这些功能通过移动网络数据获取和存储、移动通信原理和协议的翻译、网络数据知识图谱构建和基于知识图谱的解析、以及网络智能特征数据集的构建和评估来实现。本文旨在创建一种网络智能数据获取和分析的可行方法，探索网络智能数据技术的一种新型方向和范式。

内容

1 B5G/6G网络智能的愿景与挑战	6
B5G/6G网络智能的数据采集与分析系统.....	9
2.1 无线网络智能开放平台	9
2.2 数据采集.....	10
2.3 基于知识图谱的表示与分析	11
2.4 特征数据集构建.....	12
3 网络数据知识图谱的构建与应用	13
3.1 移动大数据智能特征工程.....	13
3.2 网络数据知识图谱的构建	14
3.2.1 知识图谱的技术架构.....	14
3.2.2 基于通信原理和协议的网路数据知识图谱构建.....	15
3.2.3 知识图谱补全方法	15
3.2.4 知识图谱的边权重	16
3.2.5 知识图谱中节点重要性及节点间影响的效率	16
3.3 网络数据知识图谱的应用	16
4 网络智能特征数据集的构建与评估.....	19
4.1 网络数据KPI的数据模型	19
4.2 针对特定场景影响KPI的特征的初步知识图谱过滤	19
4.3 基于机器学习的特征数据集构建	20
4.4 特征数据集构建过程的评价系统	20
4.5 面向KPI的特征数据集评价系统	21
5 特征数据集构建与评估案例	22
5.1 以上行吞吐量为核心构建本地知识图谱.....	22
5.2 深入分析以上行吞吐量为核心的本地知识图谱.....	23
5.3 构建上行吞吐量特征数据集.....	24
5.4 评估上行吞吐量特征数据集	25
6 展望	26
参考文献	27

前言

随着移动通信网络日益复杂以及各行各业差异化服务需求的多样化，超越5G（B5G）/6G的最重要研究方向和趋势是将人工智能、大数据和计算深度融入网络，并开发智能通信网络，以提供自动驾驶网络能力和精确定制服务。目前，全球在这一方向上的研究仍处于初级阶段。将数据和智能嵌入网络以开发具有内生智能以及计算和通信融合的新型网络智能技术体系，在架构、数据和人工智能算法方面提出了挑战。数据作为最基本和最重要的部分，决定了网络智能性能的上限，而架构和算法只是达到这个上限的手段。在移动通信网络运行过程中，从用户设备（UE）、无线接入网（RAN）、传输承载网络到核心网络，整个网络会生成数百万个数据字段和指标，涉及不同的软件、硬件、功能和协议栈。对于网络智能的发展，建立一套全面有效的网络数据采集和分析系统，以有效采集、分类、分析和利用不同类型的数据，是至关重要的。

作为网络结构、用户设备类型、用户设备行为、数据服务需求以及系统资源配置，在5G/6G中将是高度动态、时间敏感且相互耦合的，移动网络的数据获取和分析将面临许多挑战。主要困难在于数据获取，因为数据分散在不同的设备层。此外，各种数据类型和复杂结构使得数据分析变得困难，高度抽象的数据阻碍了理解。除此之外，具有复杂相关性的耦合数据导致挖掘困难。为了应对这些挑战，需要标准数据属性泛化，高效地澄清数据之间的关系，进一步深入探索关系之间的关联，并清晰地表示关系中所蕴含的数据和关键信息。

网络智能严重依赖数据。建立一个能解决数据获取、开放和利用等基本问题的完整数据获取与分析系统，将推动新型B5G/6G网络智能技术的发展，并成为实现网络智能的关键里程碑。本文试图创建一个可行的网络智能数据获取与分析方法，寻求网络智能数据技术的全新方向和范式。知识图谱作为网络数据知识表示的新方法以及知识管理的新工具，厘清了数据之间复杂的关联，并将数据的类型和属性以及数据间关系的类型和属性进行表征。此外，图的结构和节点信息支持智能推理和智能特征工程实践。知识图谱在解决移动通信网络数据获取与分析方面的上述挑战方面效率显著，它是B5G/6G网络智能数据获取与分析系统的关键。

这份白皮书介绍了一种B5G/6G网络智能数据采集和分析系统。该系统的成果体现在移动网络数据采集和存储、知识表示、知识推理以及特征数据集的构建和评估上。

使用知识图谱等工具对B5G/6G移动通信网络进行知识表示，将移动通信网络复杂的内部关系从“黑匣子”转变为“白匣子”。它基于移动通信原理和协议，以可视化的、分层的、结构化的方式呈现高度动态、时间敏感且易于耦合元素的移动通信网络。在知识表示的基础上，B5G/6G移动通信网络的知识推理有效地处理和整合网络上的海量数据和复杂连接，以实现快速的知识响应和推理。知识推理赋予移动通信网络感知和认知能力。

基于移动通信网络知识图谱中数据字段与服务需求的相关性分析，将数据进一步分类为不同类型，并对数据执行特征提取，以构建B5G/6G移动通信网络的特征数据集。特征数据集代表了显著影响移动通信网络关键性能指标（KPIs）的关键特征。该数据集还包含特征收集的数据。作为一种专家知识与智能算法的组合，该数据集可作为后续智能优化的直接输入。构建此类特征数据集有助于在移动通信网络的智能诊断、定位和优化过程中获取与目标指标密切相关的重要特征。这指导了进一步的智能网络感知、预测、优化，最终实现数字孪生。

组织 紫金山实验室、东南大学、北京邮电大学、中国信息通信研究院移动通信技术有限公司、华为技术有限公司、爱立信（中国）通信有限公司、vivo移动通信有限公司和中国电信集团有限公司研究所

贡献者 小虎友，永明黄，战 hang，石文河，云山伊，俊杰友，文静李，雷峰，惠徐，万飞孙，明宇赵，学强 Yan，乐乐王，任前赵，岩南袁，冰钱，高 Lu，刘杨

1 B5G/6G网络智能的愿景与挑战

在追求超越5G/6G（以下简称B5G/6G）的过程中，移动通信网络将赋能各行各业，因此将面临众多应用场景下多样化需求带来的挑战。B5G/6G的趋势是将人工智能和大数据深度融入通信以实现网络智能化。与侧重手机通信的4G网络相比，5G和B5G需要支持跨行业的全场景全业务，推动各行各业的智能化数字化转型，并提供新的高价值服务和终极服务体验。为满足这些要求，除了整合传统地面网络，6G无线通信网络将整合卫星通信、无人机（UAV）通信等非地面网络，实现空天地海一体化接入。所有频谱将得到充分探索，以进一步提升数据传输速率和连接密度。此外，6G将采用人工智能和大数据技术，引入一系列智能应用，以应对超异构网络使用、各种通信场景、大量天线、高带宽和新的服务需求所产生的海量数据。显然，6G移动网络将深度与计算、大数据和人工智能技术融合，以提供从个人应用、垂直应用到治理应用的全应用服务。预计6G将提供具有“全球覆盖”、“全频谱”、“全应用”[1, 2]等新范式的服务。

未来6G网络预计将满足支持虚拟与物理融合、全息、场景化、个性化和普适通信的服务需求，以及满足异构网络技术融合和空天地海一体化网络的网络需求。然而，目前的基于规则的算法网络运行范式受限于僵化的预设规则，难以动态适应不断变化的用户需求和网络环境。此外，网络运行经验无法有效积累，限制了网络能力的持续改进。也就是说，在当前范式下，网络没有自我进化的能力。升级或改进需要工程师进行耗时研究，这对于规模空前、运行复杂的6G网络是不可接受的。将内源智能引入网络以提升B5G/6G的自动化和智能化，是解决上述问题的关键[1]。特别是，数据对于B5G/6G网络智能至关重要。只有利用移动大数据，才能在服务体验、网络质量、网络效率和网络成本方面优化网络[4]。

主要挑战在于利用移动大数据并将智能嵌入网络，以在数据、架构和人工智能算法方面引入一种新型网络智能技术系统。为应对这些挑战，需要建立一个全面的网络数据采集和分析系统。基于数据采集和分析结果，该系统可以更好地满足差异化的应用需求，实现灵活的网络功能编排、更精细的移动通信网络资源分配以及高效的网络控制。

作为网络结构、终端类型、终端行为、数据服务需求以及系统资源配置，b5g/6g将是高度动态、时间敏感和耦合的，移动大数据的获取和分析将面临许多挑战。主要困难在于数据获取，因为数据分散在不同的设备上，在不同的层级。随之而来的是各种数据类型和复杂结构使数据分析变得困难，高度抽象的数据阻碍了理解。除此之外，具有复杂相关性的耦合数据导致挖掘困难。应对这些挑战需要标准数据属性泛化，有效阐明数据之间的关系，进一步深入探索关系之间的关联，并清晰地表示数据以及关系中固有的关键信息。此外，从终端、接入网络到核心网络，移动通信网络中会生成数万条数据字段和指标，涉及不同的软件、硬件、功能和协议栈。

在移动通信网络智能化发展方面，有效地收集、分类、分析并利用不同类型生成的数据，充分释放网络服务潜力，充分发挥网络技术优势是关键。

传统数据分析方法可分为两类。第一类是基于通信领域积累的理论和经验建模分析。例如，我们有用于异常根本原因分析的专家系统、基于无线网络知识库中多个KPI的异常检测[3]，以及用于无线资源管理优化的优化模型分析。第二类是数据驱动的机器学习模型和算法，它们专注于为特定问题提供一种高效的算法。由于B5G/6G网络预计将比以往更复杂，且数据驱动AI技术存在局限性，这些学习模型和算法无法支持从检测网络症状到实施方案的全过程。这个问题在移动通信中普遍存在，其“黑盒”内部机制缺乏可解释性和可扩展性。因此，单一的知识/模型驱动或数据驱动模式不能满足B5G/6G移动通信大数据分析的需求。移动通信中的数据分析需要转变为由知识/模型和数据共同驱动的智能模式。

作为人工智能在移动通信中的数据分析非常有效，其发展至关重要。早期的人工智能形式，以专家系统为代表，强调因果关系和与符号推理相关的知识结构，代价是违反排中律。后续的人工智能形式，以统计机器学习和深度学习为代表，重视排中律，代价是违反因果关系。人工智能的发展需要这两种类型的人工智能的组合。2018年底，第三代人工智能的理论框架正式发布：(1) 建设可解释和鲁棒的人工智能理论与方法 (2) 发展安全、可靠、可信和可扩展的人工智能技术。人工智能路线图突出了数据与知识融合的人工智能理论与方法。由于知识图谱目前相对成熟，它可以促进第三代人工智能的实施。

基于知识图谱等工具，类脑数据治理系统已被应用于许多领域，形成了一个数据-信息-知识-智能的技术路径。这种闭环类脑系统是两种类型人工智能之间的折中解决方案。同样，在移动通信领域，为了嵌入数据和智能以引入一种新型网络智能技术系统，也需要遵循数据-信息-知识-智能路径。因此，这种技术系统将依赖数据获取与分析。知识图谱是网络智能技术路径的关键，它集成了第一代和第二代人工智能的优势，将在推动B5G/6G网络智能中发挥关键作用。

知识图谱有效地厘清了多源异构移动通信数据之间松散数据结构、分层与扁平化网络以及内生网络性能要素的关联。借此方式，知识图谱有助于实现知识互联互通、高效资源管理以及智能网络维护。移动通信网络的内生要素及其关联可动态呈现并以易于理解的方式准确定位，具有清晰结构。基于知识图谱的数据分析对于移动通信知识的获取、排序及表征至关重要。

基于网络数据知识图谱的相关性分析，初步筛选目标指标的特征。经过持续算法优化和模型训练，构建了针对不同应用场景的特征数据集。整个过程实现了从移动通信知识排序到移动通信智能应用。

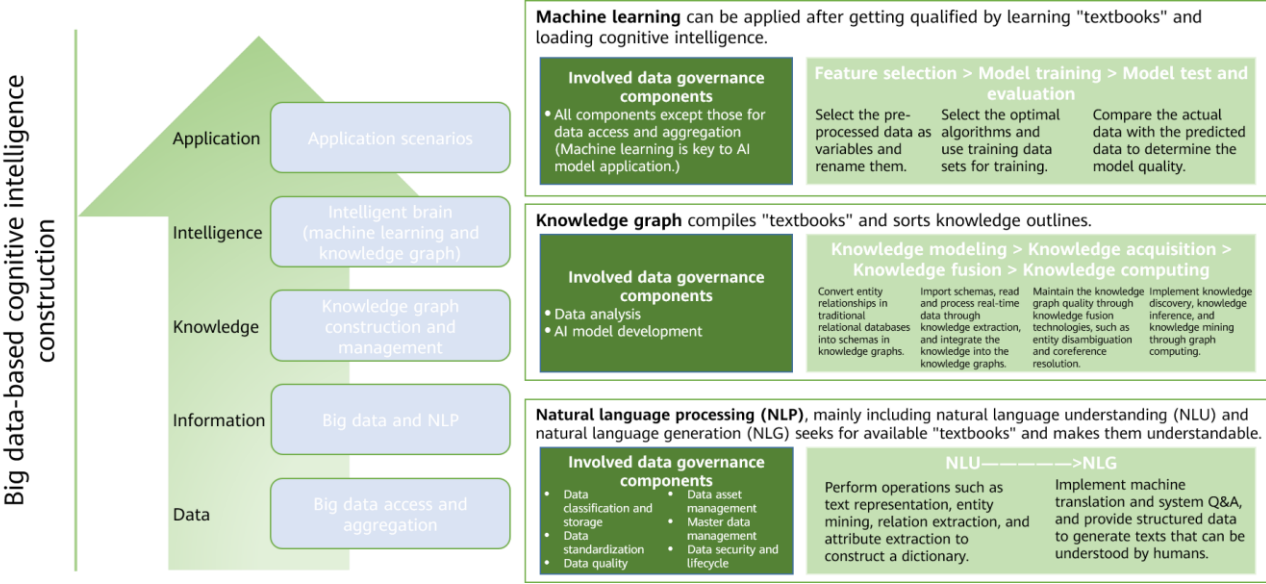


图1-1 基于大数据的认知智能构建在各种行业中常用[13]

B5G/6G网络智力的2数据采集与分析系统

将数据和智能嵌入网络以开发新型技术系统在架构、数据和人工智能算法方面提出了挑战。为应对这些挑战，需要构建一个全面的网络数据获取与分析系统。该系统由四个功能模块组成：无线网络智能开放平台、数据获取、基于知识图谱的相关性分析以及特征数据集构建。这些模块实现了数据层信息集成、信息层知识提取、智能层知识计算和应用层特征定制。

无线网络智能开放平台模块和数采模块聚焦于B5G/6G的逐层数据采集。基于知识图谱的相关分析模块和特征数据集构建模块有助于应对B5G/6G中数据表示、数据分析及特征提取的挑战。这四个模块协同合作，为B5G/6G网络智能奠定数据基础，并探索实现路径。

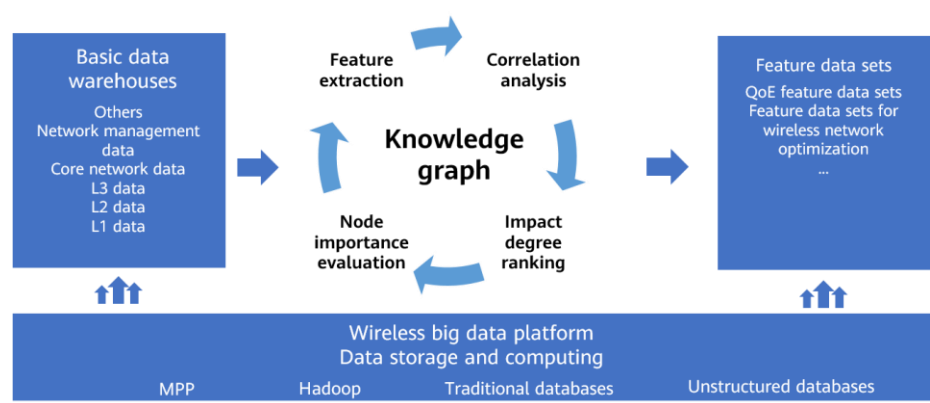


图2-1 B5G/6G网络智能数据采集与分析系统

2.1 无线网络智能开放平台

基于B5G/6G网络和数据仓库的无线网络智能开放平台，深度整合大数据和人工智能。在B5G/6G网络智能数据采集和分析系统的指导下，该平台旨在实现数据开放、数据共享、数据分析和数据应用。

B5G/6G无线网络智能开放平台模块提供可视化一站式解决方案，用于实时移动网络数据的采集、分析和跟踪。它提供专业的网络数据支持，有助于构建优质生态系统，以实现行业内的真实数据开放。该智能开放平台的架构由三层组成：数据采集与预处理层、数据分析层和数据应用层，如图所示。

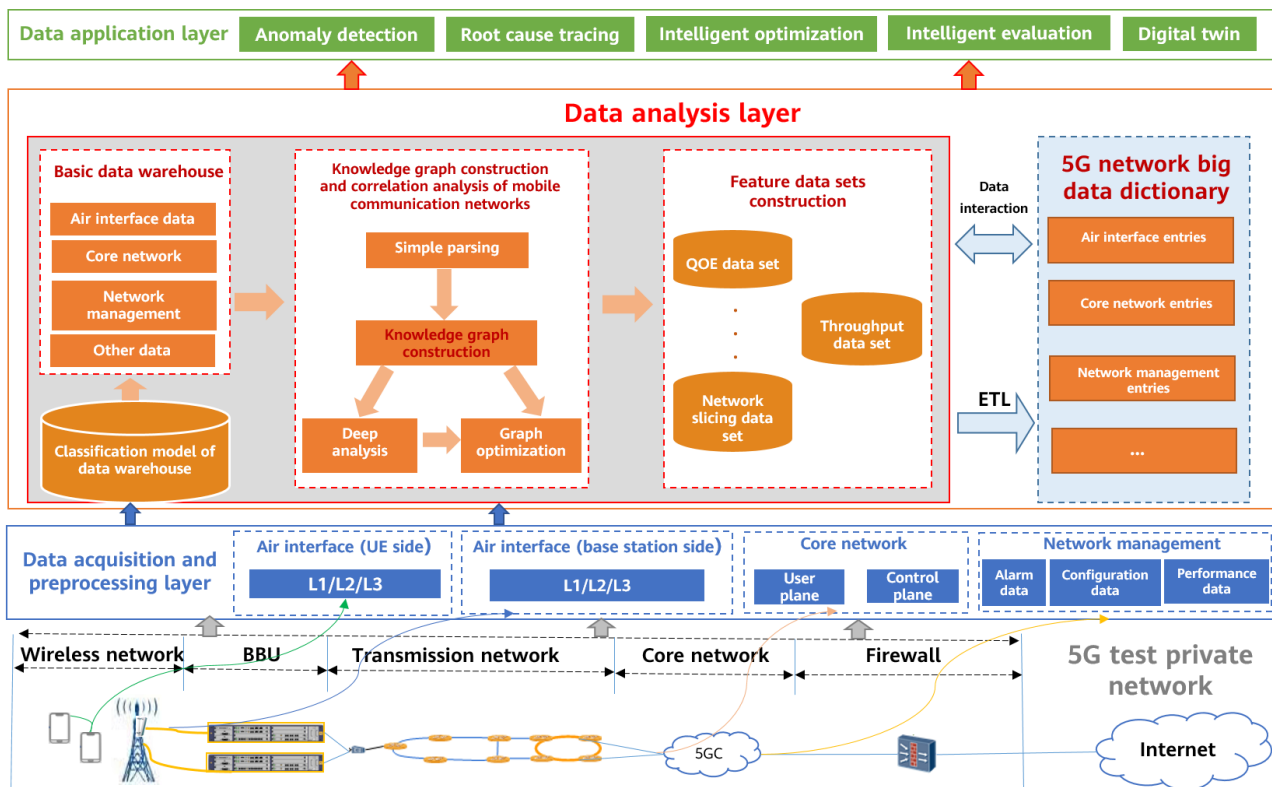


图2-2 B5G/6G无线网络智能开放平台

开放平台通过API访问多维异构B5G/6G数据，形成面向应用、结构相对稳定的时间变化型、高度集成型B5G/6G数据仓库。基于此类数据仓库，建立网络数据采集与分析系统，涵盖网络数据知识图谱构建、基于知识图谱的相关性分析、特征数据集构建以及5G网络大数据词典。5G网络大数据词典是一个完整的数据系统，反映了移动网络运营的特征，将为网络智能化发展奠定坚实基础。输出（由网络数据知识图谱构建、基于知识图谱的相关性分析、特征数据集构建生成）与5G网络大数据词典交互，尤其交互过程中的主要成果将被纳入词典。这些数据分析层上的成果有助于实现数字孪生、异常检测、溯源、智能优化、智能决策和智能评估等应用。

2.2 数据采集

数据采集模块是B5G/6G网络智能数据采集与分析系统的基本组成部分。没有这个模块，所有后续的数据分析和智能数据应用都无法实现。传统数据采集涉及创建表格、过滤数据、收集数据以及将数据导入数据库。它缺乏按需数据处理。传统数据采集方法包括路测、基于硬件的信令数据收集和基于软件的信令数据收集。收集的数据可以分为实时数据和非实时数据。为了使数据采集、数据传输和数据交互更加高效，数据采集功能模块可以根据不同的智能优化应用场景和移动通信网络部署，进一步定制时空数据采集。收集的数据可以存储在基本数据仓库中。

数据采集模块通过各种硬件和软件以及路测收集试点网络的各种类型数据。数据包括无线空中接口数据（用户面和基站面）、核心网数据以及网络管理数据。

无线空中接口数据（用户面和基站面）包括物理层（L1）数据、数据链路层（L2）数据和网络层（L3）数据。物理层包括下行和上行共享信道数据、下行和上行控制信道数据、随机接入信道数据以及探测参考信道数据。数据链路层包括下行调度数据、上行调度数据、下行控制数据以及信道状态信息数据。网络层包括广播数据、寻呼数据和接入数据。通过提取上述数据，基站面建立通信接口，向外部平台发送实时数据，以实现基站面的数据采集。

核心网数据包括控制面数据和用户面数据。控制面数据主要指的是用于控制信令协议类型以控制业务过程建立、维护和释放的数据。用户面数据指的是业务数据，例如语音、分组业务、即时消息、电子邮件和视频数据。

网络管理数据包括性能、告警和配置数据。具体来说，性能数据是指从不同网络元素（NE）收集的性能信息。告警数据是指所有NE为设备故障、网络事件以及网络和服务相关故障生成的各种类型的告警报告。配置数据是指关于NE的基本信息。它映射到实体，主要用于网络拓扑信息管理。

数据预处理涉及对收集到的原始移动网络数据执行操作，例如脏数据处理、基于时间的分区、对非结构化数据进行结构化处理，以及通过ETL将数据导入数据库。

ETL是一个工具，用于从分布式、异构数据源中提取关系数据和平面数据文件到临时中间层进行清洗、转换和集成。最终将数据加载到数据仓库或数据集市，用于在线分析处理和数据挖掘。

2.3 基于知识图谱的表示与分析

知识图谱，一种表示和管理知识的新方式，已在越来越多的垂直应用领域中涌现，并发挥着越来越重要的作用[7-11]。基于移动数据构建的网络数据知识图谱作为移动通信垂直领域的行业知识图谱。知识图谱的构建对移动通信知识的深度、广度和准确性有很高的要求。

基于知识图谱的关联分析模块借鉴网络数据知识图谱进行知识表示、关联分析和深度挖掘，为5G/6G网络智能化提供有效的知识规则和知识计算支持。由于5G/6G网络结构、UE类型、UE行为、数据服务需求和系统资源配置将高度动态、时效性强且耦合，获取分散的移动通信数据、分析多类型和复杂结构的数据以及进行复杂关联的挖掘将面临严峻挑战。知识图谱有助于有效厘清数据字段与通信网络指标之间的关系，并基于这些关系进行深度挖掘。例如，知识图谱可以量化关系的关联程度，并表示数据字段和指标的特征属性。该模块将不同层级的收集到的分散、无序、耦合数据进行有序整理，确保深入的数据-信息-知识处理，为网络智能化提供丰富的输入。

2.4 特征数据集构建

特征数据集构建是闭环网络全局智能通信的一个关键部分，涵盖了数据层信息集成、信息层知识提取、智能层知识计算和应用层特征定制。

例如，对于表示网络性能的重要KPI，如果我们能够使用知识图谱识别所有影响该KPI的节点参数，按照节点的测量重要性以及其对KPI影响的效率进行排序，并使用机器学习算法进一步过滤特征，那么我们就可以构建这个KPI的特征数据集。

以下描述了特征数据集构建模块的工作机制。首先，该模块利用网络数据知识图谱构建后的数据字段和指标的数据分析结果，更新知识图谱中的边权重和图结构，为节点间关联的进一步挖掘奠定基础。然后，通过确定知识图谱中边的权重、节点的权重以及节点间影响的效率，该模块确定影响网络性能和用户体验质量的因素，并执行诊断、定位和分析。本步骤的目的是基于相关因果追溯理论，准确识别影响目标指标的关键参数，并形成关于这些关键参数的关键数据字段的特征数据集参数目录。最后，该模块使用参数目录和机器学习算法来构建关于目标KPI的特征数据集。

3网络数据知识图谱的构建与应用

面向人工智能的移动数据治理首先采用知识规则库，然后使用知识图谱，随后应用机器学习。这三种认知智能技术应用于数据治理的绝大部分组件，并决定了数据处理能力和人工智能应用系统的智能化水平。

移动网络数据具有以下特征：大量不同服务类型的松散数据结构的多样化数据源。网络结构（生成、承载和传输此类移动网络数据的网络结构）、用户终端类型、用户终端行为、数据服务需求以及系统资源配置都高度动态、时效性强且相互耦合。需要一种方法，能够以易于理解的方式动态展示移动通信中的内生要素和内生关联，并具有清晰结构和准确定位。首先，基于移动通信原理和协议，对移动网络数据进行全面梳理。这是面向人工智能的移动数据治理的起点。利用专家知识库和数据挖掘技术，将专家知识和数据转化为人类可理解的信息和知识规则。

然后，采用知识图谱技术构建具有逻辑推理关系的知识挖掘和知识推理框架和模型。在构建网络数据知识图谱后，无线、接入和核心网络可以全面监控5G试点网络提供的服务，甚至可以进行端到端监控。最后，使用机器学习为不同应用场景构建特征数据集。最后一步也是知识图谱的一个重要应用。

3.1 移动大数据的智能特征工程

通常，数据决定了机器学习的上限，而算法只是达到这个上限的手段。这里，数据特指通过特征工程获得的数据。特征工程是一个将原始数据转化为模型的数据训练过程。它的目的是获得更好的训练数据特征，使模型能够达到机器学习的上限。

特征工程有助于提升模型的性能，有时即使在简单模型上也能取得良好效果。特征工程作为机器学习中的一个重要部分，包括特征构造、特征提取和特征选择。特征构造需要垂直领域的实践经验。特征提取和特征选择用于从原始特征中识别最有效的特征。特征选择涉及从特征集中选择具有明显物理或统计意义的特征子集。特征提取有时能帮助发现有意义的特征属性。特征提取和特征选择都能帮助减少特征维度和数据冗余。特征选择需要持续迭代，在此过程中，每个特征对模型构建的重要性可以被表示。

构建网络数据知识图谱，并基于构建的知识图谱对内生因素之间的关联进行深入分析，可以有效地减少特征选择中的重复迭代。这对于智能通信领域的研究人员来说，是优化目标KPIs的一种有效的数据字段选择解决方案。移动大数据的智能特征工程涉及以下方面：基于通信原理和协议构建网络数据知识图谱，边连接强度和边权重计算，知识图谱中节点重要性以及节点之间影响的效率，以及基于知识图谱对在特定场景下影响KPIs的特征进行初始筛选。

基于网络数据知识图谱的智能特征工程能够大幅降低冗余特征的影响，充分发掘知识挖掘和源追溯推理的优势，使特征筛选更加高效。

3.2 网络数据知识图谱构建

自然知识图谱是整合了大量客观人类经验以形成庞大网络的语义网络。构建完整的网络数据知识图谱是一个复杂的系统工程过程。与其他领域基于规则、机器学习、深度学习和强化学习算法的知识抽取相比，网络数据知识图谱的知识抽取主要基于通信原理和协议[5]。因此，需要大量具有通信知识的人力来对图模型进行排序。手动排序的网络数据图模型只是基本图形模型，需要进一步进行实体抽取、关系抽取和属性抽取，以形成网络数据知识图谱的拓扑结构。基于构建的拓扑结构，可以进行数据驱动的图结构更新、深入的关联分析以及利用动态数据感知进行知识计算。

3.2.1 知识图谱技术架构

知识图谱构建涉及数据获取、知识抽取、知识存储和知识融合。

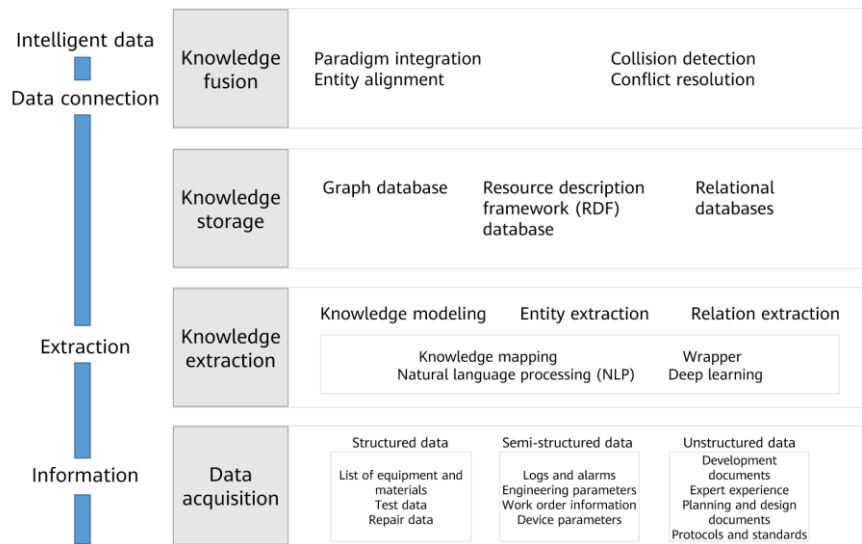


图 3-1 知识图谱技术架构

数据采集：网络数据包括结构化数据以及大量半结构化和非结构化数据，例如网络日志、工程参数、开发文档、设计文档、通信协议和标准。

知识抽取：知识抽取是知识图谱构建的核心，包括知识建模、实体抽取和关系抽取。抽取方法与数据的类型和结构化层次密切相关。常用技术包括知识映射、包装器、自然语言处理(NLP)和深度学习。

知识存储：知识图谱的存储方法取决于应用场景、系统性能和推理能力的要求。可以使用图数据库或资源描述框架数据库进行存储。在某些场景下，也可以使用关系数据库进行存储。

知识融合：由于移动通信网络复杂，知识图谱的构建无法一蹴而就。分别针对无线网络、核心网络和数据传输领域构建了独立的知识图谱。通过融合，最终获得用于全智能决策的综合知识图谱。

3.2.2 基于通信原理和协议的网络数据知识图谱构建

构建网络数据知识图谱旨在以知识图谱的形式可视化数据字段（与移动通信协议、性能以及流程相关）与指标之间的因果关系和关联规则。图中的节点代表数据字段和指标，边连接代表它们之间的因果关系和关联规则。这为探索关联分析方法提供了基础。构建网络数据知识图谱的步骤如下：

1. 确定实体和关系：基于通信领域的专家知识，如通信原理和协议，对网络数据中的相关指标和数据字段进行梳理，并确定可分类的实体类型。此外，确定已定义实体之间的关系。
2. 建立三元组：基于确定的实体和关系，三元组（*head*，*关系*，*尾*）在网络数据中表达实体之间关系的逻辑被建立。*head* 和 *尾* 分别指示三元组中的头部和尾部实体。*关系* 表示实体之间的关系。三元组中的头实体和尾实体在网络数据知识图谱中表现为节点。三元组中实体之间的关系表现为图中节点之间的连接。
3. 构建知识图谱：头实体和尾实体在知识图谱中表现为节点，每个节点代表一个与性能相关的数据字段或指标。关系则描绘为节点间的边。基于通信专家知识库中的知识和规则，具有关系的实体节点进行定向连接，最终形成拓扑网络数据知识图谱。

3.2.3 知识图谱补全方法

知识图谱补全（KGC）是通过预测知识图谱三元组中的缺失部分，使知识图谱更加完整的行为。知识图谱 G 基本上包括一个实体集 $E = \{e_1, e_2, \dots, e_n\}$ （ n 表示实体数量），一个关系集 $R = \{r_1, r_2, \dots, r_m\}$ （ m 表示关系数量），以及相应的三元组 $T = \{(e_i, r_k, e_j) \mid e_i, e_j \in E, r_k \in R\}$ 。由于知识图谱中的实体 E 和关系 R 有限，有些实体和关系可能不属于 G 。因此需要知识图谱补全（KGC）。根据需要的内容，KGC可以分为三个子任务：

预测或补充给定部分三元组 $(?, r, e)$ 的头实体 e 。预测或补充尾实体 e 。
 对于一个给定的部分三元组 $(e, r, ?)$ 。

预测或补充头实体和尾实体之间的关系 r ，对于给定的部分三元组 $(e, ?, e)$ 。

在网络数据知识图谱中，KGC有助于实现特征数据间有效的关系挖掘，并识别用户层KPI与网络层KPI之间的隐含关联。

链接预测是知识图谱构建（KGC）的主要方法。它根据上述三个子任务的要求，涉及头实体链接预测、尾实体链接预测和关系链接预测。链接预测主要基于知识图谱嵌入[12]和图神经网络[5]。

3.2.4 知识图谱的边缘权重

在知识图谱中，一条边表示两个实体节点之间的关联信息，关联强度通常由边的权重表示。边的权重信息有助于知识图谱应用。例如，在移动通信系统中，信道质量指示符（CQI）的下降可以直接导致调制和编码方案（MCS）指数调整。因此，CQI和MCS指数数据字段之间边的权重应该相对较大。这样，当检测到MCS指数调整显著时，系统需要检查调整是否主要是由于CQI下降引起的。同样，多个因素可能导致无线网络空中接口传输速率下降。然而，这些因素的重要性不同，如MIMO层数、调制阶数、编码比特率、物理资源块（PRB）数、资源开销率以及符号数。因此，代表这些因素和代表空中接口传输速率的节点之间边的权重应该不同。这样，系统可以进行更准确的诊断。

对于特定的知识图谱，存在许多计算连接节点的边权重的方法，例如基于两个实体之间的直接相关程度或相关统计信息的计算，以及网络结构中节点之间测量的相似度。需要根据网络数据所表示的具体特征来选择合适的方法。

3.2.5 知识图谱中节点的权重和节点间交互的效率

节点间冲击效率是指节点间最短路径上图边权重的乘积。它表示节点在定义路径上的冲击程度。例如，在移动通信系统中，CQI的降低直接导致MCS指标调整，这可能影响用户吞吐量，进而用户吞吐量的变化直接影响用户感知体验质量。因此，用户感知体验质量指标受到影响链中所有因素的共同影响，而每个因素都有独特的冲击程度。

基于节点间冲击效率，对部分知识图谱中每个节点的重要性进行度量，以表征完整知识图谱中节点的重要性。每个节点的重要性基于节点的输出强度、节点的输入强度以及其他节点对节点输出和接收的贡献所确定。采用这种方式，在构建完整网络数据知识图谱后，可根据数据感知动态调整图谱中所有节点的重要性。在获得所有节点的重要性后，由于节点间的长期相关性，只需监控重要性最高的节点即可监控全网数据质量。

这为后续的网络数据知识图谱应用奠定了基础。

3.3 网络数据知识图谱的应用

应用1：基于内生智能的按需网络优化

未来6G网络将实现大规模物联网（IoT）和万物智能连接，使具有高度多样化需求的应用场景成为可能。随着6G应用范围日益深入和广泛，基于内生智能的准瞬时网络优化至关重要。深入的网络与计算融合将成为趋势。基于生成的网络数据知识图谱，当网络与用户级性能评估指标中的某个指标或某类指标的数据发生剧烈波动并超过预设阈值时，将确定所有连接目标指标的路径。然后，确定并排序沿这些路径上所有节点对目标指标的影响效率。一个或多个具有相对较高

将影响效率作为目标节点。此外，从目标节点的相关节点中识别可调节数据参数，并为决策和执行提供智能优化解决方案。通过这种方式，可以观察到网络和用户级性能评估指标是否得到优化。基于用户需求的内生智能网络优化可初步实现。

应用2：智能无线电网规划

未来的无线网络将涉及空间、空中和地面无线覆盖，以及各种类型的计算能力、数据、控制节点资源以及ToC/ToB服务需求。面对复杂的输入，无线网络规划需要满足多个通常相互冲突或没有直接关系的多目标。例如，在空中对地协同无线网络部署和规划场景中，空中基站满足热点容量需求，地面基站提供广覆盖。这两种类型基站部署不受空间相关性的限制，相关的优化参数也不同。然而，这两种类型的基站受到总可承受部署成本和无线信号干扰的约束。因此，网络规划成为典型的多维多目标优化问题。随着目标数量的增加，常规多目标优化算法在优化过程中产生大量非支配解。结果，最优规划无法实现。

类似大脑的网络数据知识图谱可以指导复杂网络的自动规划。知识图谱用于在特征（例如资源和服务）之间找到最相关的集合以及由待规划指标形成的特征子集。这样，一个复杂的规划问题被划分为几个子问题以进行联合优化。例如，对于空地协同无线网络部署和规划，通过特征分析，将旨在提高吞吐量的空中基站规划作为一个子问题，将旨在提高覆盖水平的地面基站规划作为另一个子问题。为了准确获得网络规划方案，我们需要在每个目标上找到具有出色性能的方案，并利用知识图谱在特定场景（如ToB、ToC和应急服务）中找到子问题（如干扰协调、容量增强和成本优先），对在多个目标上具有全面最优性能的方案进行加权选择。

应用3：网络运维智能、快速故障定位

6G网络将通过卫星、无人机、地面基站和其他接入模式提供空-天-地-海全域三维覆盖。同时，虚拟化和容器技术将使网络极其灵活。

然而，网络将比以往更大、更加复杂。众所周知，一旦节点或网络功能出现故障，相关节点将变得异常。将产生大量警报，并且难以快速定位故障原因。6G网络预计将具有内生智能。基于无处不在的数据、算法和计算能力，数据可以被高效获取和分析，以实现快速网络故障定位和根本原因分析。大量异构站点生成海量数据。基于内生智能和知识图谱，大量历史警报数据被分类为初始警报信息和相关警报信息，以高效提取初始警报和相关警报之间的关联规则。得益于这一点，在高度动态的异构网络上，异常站点或功能可以被快速定位，从而快速定位故障原因。在网络运维方面，除了物理机器的异常，VM或容器异常也需要被妥善定位和处理。相关历史警报可以从运维支撑系统（OSS）或设备日志中获取。然后，基于网络运维知识图谱和其他设备日志（如事件日志以及相关的软硬件日志），进行分析和故障排除，以快速定位故障的根本原因。此外，基于知识图谱的网络数据分析可以在优化网络传输参数方面发挥关键作用。

应用4：无线网络的数字孪生

数字孪生网络是一个由物理网络实体和虚拟孪生组成的数字网络系统。物理网络实体和虚拟孪生可以实时交互并相互映射。构建数字孪生网络需要四个关键要素：数据、建模、交互和映射。通过将数字孪生技术应用于网络并创建物理网络设施的虚拟镜像，构建数字孪生网络平台，该平台的NE、拓扑和数据与物理网络实体一致。数字孪生网络面临的主要挑战是实现高性能孪生并最小化网络开销。应对这一挑战有两种方式。第一种是尽可能重用数字孪生网络和物理网络上的硬件、软件、空中接口频率、UE和其他资源，以保证孪生和隔离特性。第二种是优化孪生建模以减少数字孪生网络的资源开销。数字孪生网络旨在将物理网络的本质从“黑箱”转变为“白箱”，这与基于知识图谱的无线网络知识表示目标一致。知识图谱作为语义网络工具，以及基于知识图谱的深度数据挖掘，有助于构建孪生网络本体模型，并在一定程度上实现虚拟到物理的映射。数字孪生技术已经应用于一些实时网络。然而，由于孪生通常基于大量特定领域专家知识构建，它们缺乏通用性，只能应用于网络环境中的特定物理或逻辑区域。提高通用性和效率的一种方法是结合前两种方式。通过重用数字孪生网络和物理网络上的资源，可以快速获取大量优质无线网络大数据。基于这些数据，可以应用知识图谱和人工智能构建无线网络元模型和元算法。然后，通过资源重用验证模型和算法，并需要根据需要基于数据集进行进一步定制和迭代。最终，可以形成具有强泛化能力的数字孪生网络，并应用于多个无线网络环境。

应用5：跨学科、跨行业、多领域协作的智能研发

现代产品研发通常是一个涉及多学科和专业技术领域的复杂系统工程。随着产品日益复杂，产业分工愈加精细，智能研发需要整个产业链上下游的供应商、合作伙伴和客户之间的协作。针对涵盖机械、通信和软件领域的5G/6G基站研发，迫切需要一种具备跨域协作的智能研发系统，以便基于知识图谱构建专业能力共享与协作系统。

4 网络智能特征数据集的构建与评估

传统的智能网络通信特征工程主要关注应用的几个KPI。在移动通信知识图谱构建完成后，为了智能通信的发展，需要对所有KPI完成特征工程。本章重点在于开发一套标准，用于评估特征数据集构建过程以及面向KPI的特征数据集。

4.1 网络数据 KPI 的数据模型

网络数据KPI可以分为网络质量KPI和网络容量KPI。网络质量KPI包括可访问性、保持性、移动性和完整性KPI。网络容量KPI包括流量（吞吐量）和资源使用KPI。

1. 可访问性关键绩效指标包括RRC连接建立成功率、无线接入成功率和特定服务的无线接入成功率，例如会话式语音服务、会话式实时视频流服务、实时游戏服务和非会话式视频流服务。
2. 可保持性KPIs包括特定服务的无线呼叫掉话率以及无线呼叫掉话率/服务掉话率，例如对话式语音服务、对话式实时视频流服务、实时游戏服务和非对话式视频流服务。
3. 移动性KPI包括 intra-RAT 漂移成功率，inter-RAT 漂移成功率，intra-频率漂移成功率，以及 inter-frequency 漂移成功率。
4. 完整性KPI包括数据包丢弃率和数据包丢失率。
5. 吞吐量kpi包括平均上行/下行ue吞吐量、空中接口上行/下行数据量以及特定业务的上行/下行数据量，例如对话式语音服务、对话式实时视频流服务、实时游戏服务和非对话式视频流服务。
6. 使用kpi包括信道使用率、平均上行/下行RB使用率、寻呼拥塞率、频谱效率和能效。

每种KPI都可能受到一个或多个绩效数据的影响或计算。KPI和绩效数据在知识图谱中表现为实体，而KPI与绩效数据之间的关联则呈现为实体之间的关系。最终，可以构建以KPI为核心的网络数据性能知识图谱。

4.2 针对特定场景影响KPIs的特征的初步知识图谱过滤

针对特定场景下影响KPI的特征进行初步知识图谱过滤，有两个目标：一是减少后续特征工程中的冗余特征，二是增强模型可解释性。在网络数据知识图谱构建完成后，所有移动网络数据字段之间的相关性得到明确和深入解析，因此可以对影响KPI的关键特征进行过滤。这样一来，后续针对目标KPI的特征过滤和拟合中，冗余特征的数目将显著减少。

随着这一点，由于知识图谱工具的因果关系而不是“黑盒”模式下的曲线拟合，影响KPI的过滤特征的可解释性得到了提高。

在特定场景下，选择最受关注的靶标KPIs。基于累积数据，计算所有节点在路径上对靶标KPIs的影响节点重要性和效率。选择在整个图中具有高重要性和巨大影响节点对靶标KPIs的节点。该过程通过知识图谱初步过滤影响KPIs的特征。

4.3 基于机器学习的特征数据集构建

基于对影响特定场景下KPIs的特征进行初步知识图谱过滤，可以使用传统机器学习算法来构建特征数据集。这有助于进一步过滤出一组彼此略微相关、对目标KPIs高度敏感且最能影响目标KPIs并反映其变化的特征。特征数据集的构建是闭环智能通信的关键部分。

根据任务目的，机器学习算法分为三类：分类、回归和聚类算法。

分类算法：使用已知数据集（训练集）来获得相应的模型，该模型可用于对未知数据进行分类。数据集通常是标记集，因此分类是一个具有两个步骤的监督学习任务。第一步是训练数据以获得模型，第二步是使用模型对未知数据进行分类。

回归算法：与分类算法不同，回归算法是监督学习算法，用于预测和建模数值连续随机变量。

聚类算法：聚类是一种无监督学习任务。聚类算法根据数据的内部结构搜索自然组（聚类）的观测样本。

特征数据集的构建主要涉及分类和回归算法。特别是，由于移动通信中的大多数指标具有连续值，因此通常使用回归算法。

4.4 特征数据集构建过程的评价体系

特性和模型是原始数据和预期观测结果之间的中间产品。在机器学习中，模型和特性都需要被选择。这需要在特性和模型之间进行权衡，它们相互影响。好的特性简化建模，所构建的模型能更好地执行所需任务。差的特性可能需要更复杂的模型才能达到相同的性能水平。

有效地评估特征数据集的构建过程和构建的特征数据集是关键。构建过程从三个方面进行评估：完整性、充分性和必要性。

完整性 这意味着所有为确保目标KPI的适配度而必需的特性，都已被添加到特性数据集中。

充足性 指的是特征数据集已经从所有影响目标KPIs的参数中提取出来。在特征工程中，有两个阶段，子集搜索和子集评估，期间计算特征数据集中所有子集的信息增益。得益于知识

图,所有影响目标KPI的节点都被遍历,确保基于知识图谱的特征数据集的充分性。

必要性 意味着没有遗漏关键节点。即使节点对目标KPIs影响不大，它们对整体网络性能也是不可或缺的。

4.5 面向KPI的特征数据集评估系统

作为特征数据集呈现网络通信情报，全面评估所构建的特征数据集至关重要，以便对构建特征数据集的所有前期工作进行恰当评估。面向KPI的特征数据集从以下四个方面进行评估：

适配度 反映了为目标任务KPI构建的特征数据集的适配程度，并有助于精确评估和预测目标任务KPI。

特征提取效率 在确保一定适配度的情况下，应减少冗余特征以提高特征提取效率。

特征敏感度 它表示多维特征对目标KPI的敏感性，并有助于提供关键节点信息。

最小信息标准 应使用最少数的维度来传达最充分、可解释和准确的信息。这有助于确定一个合适的模型。

5 特征数据集构建与评估案例

5.1 以上行吞吐量为核心的本地知识图谱构建

以上行吞吐量为核心，从RAN的473个数据字段中选择了86个节点。基于3GPP规范和通信原理，构建了一个以上行吞吐量为核心的本地知识图谱。

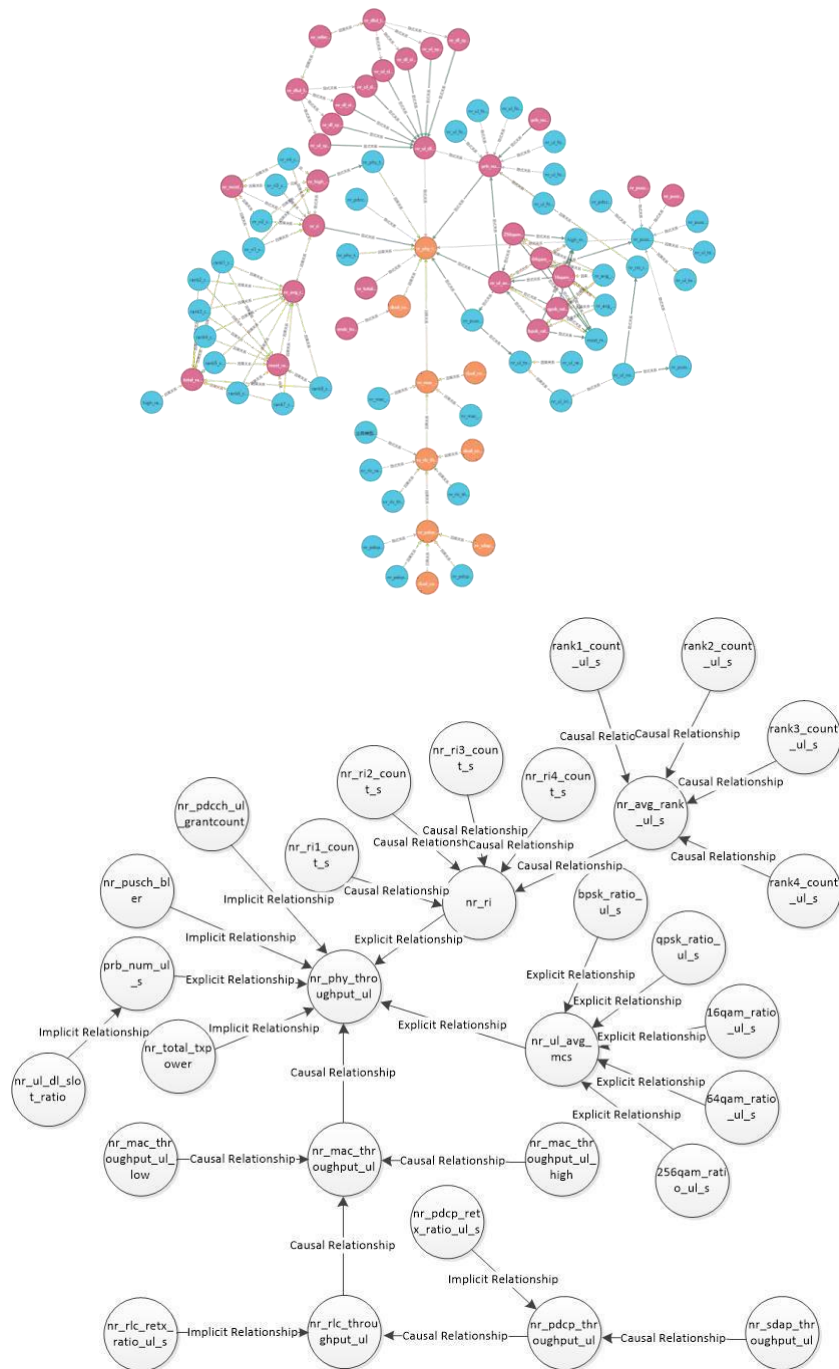


图5-1 完整上行吞吐量知识图谱（上）和放大大局部图（下）

5.2 以上行吞吐量为核心的本地知识图谱的深入分析

首先，分析本地知识图中节点与其核心的上行吞吐量之间的相关性，以获得边的权重。



图 5-2 完全上行吞吐量知识图谱（上部）和放大局部图（下部）中的相关性

然后，根据图结构和边权重，对图中节点的重排序，获得对目标KPI（物理层上行吞吐量）的影响效率。

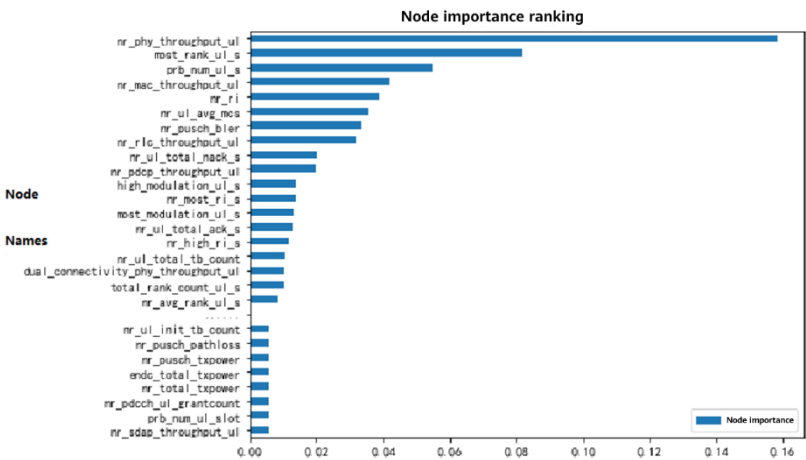


图 5-3 确定知识图谱中节点的权重

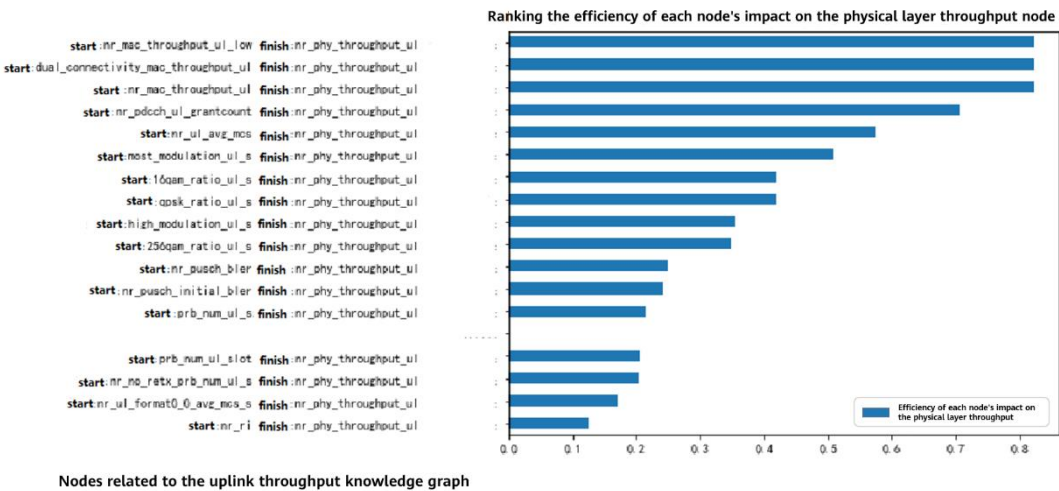


图5-4 排序每个节点对物理层上行吞吐量的影响效率节点

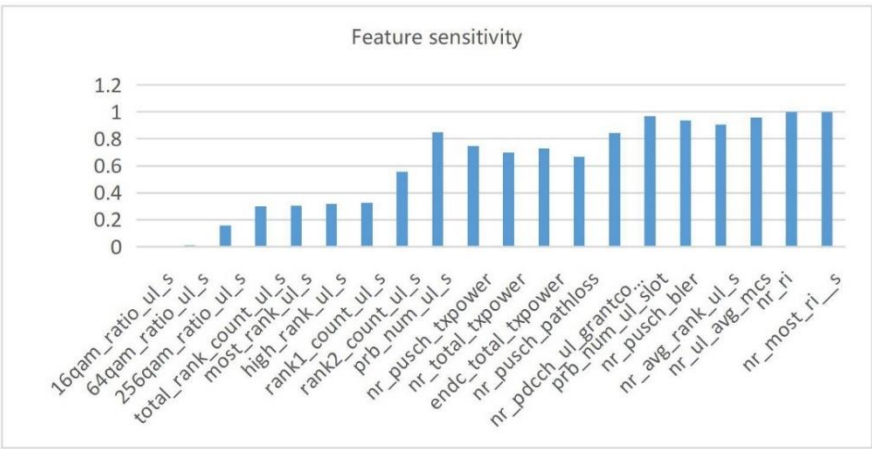
5.3 上行吞吐量特征数据集的构建

基于以核心上行业务吞吐量为关键点的本地知识图谱的深入分析结果，初步筛选出45个对目标KPI影响效率高且在图谱中重要性高的节点。这是构建上行业务吞吐量特征数据集的基础。然后，利用相关的机器学习算法获得上行业务吞吐量特征数据集。该特征数据集包含20个特征（节点）及相关节点数据。

[illegible]

特征数据集可用于直接拟合上行吞吐量。当拟合度达到一定水平时，模型可以部署到网络中以预测和分析上行吞吐量。

在以核心上行业务吞吐量为依托的本地知识图谱的构建与分析过程中完整性，充足性，和必要性已得到保证。



6 展望

当基于知识图谱的网络智能数据获取与分析系统变得更加全面和成熟时，网络数据知识图谱和特征数据集的构建可以逐步标准化。此外，所有上述工作将以适当的方式整合到网络大数据字典中。网络大数据字典将包含开放数据平台获取的移动通信数据信息集、数据字段的排序、分类、含义、归属以及相关性，以及影响KPIs的推荐关键特征。由于数据字段（或特征）之间存在被广泛认可的相关性，AI模型在图像和视频等领域才能创造价值。类似地，移动通信网络的数据字段也需要这样的关系作为参考。随着网络大数据字典的不断丰富，它将指导未来关于网络智能的研究。

参考文献

- [1] 您, X., 王, C. 等. (2021). 迈向6G无线网络: 愿景, 使能技术及新范式转变. 中国科学: 信息科学, 64(1): 1-74. [2] IMT-2030 (6G) 推广组. (2021年10月). 6G愿景与候选技术白皮书. [3] 钱冰 等. (2021). 智能运维, 基于人工智能的应用实践[M]. 机械工业出版社. [4] 黄, Y., 刘, S., 张, C., 您, X., & 吴, H. (2021). 5G/B5G智能网络的真数据测试平台. 智能与融合网络, 2(2), 133-149. [5] 何, S., 熊, S., 欧, Y., 张, J., 王, J., 黄, Y., & 张, Y. (2021). 图神经网络在无线网络中应用综述. 通信学会开放期刊. [6] 佩尔, J. (2018). 因果革命带来的七大火花对机器学习的理论阻碍. arXiv预印本 arXiv:1801.04016. [7] 余, A., 蔡, D. (2021). 基于知识图谱的社区问答生成[J]. 信息科学, 545:132-152. [8] 金, T., 尹, Y., & 金, N. (2021). 基于深度学习的COVID-19知识图谱生成. 可持续发展, 13(4): 2276. [9] 马克西姆, D., 奥利维尔, F., 尼尔斯, P., 弗洛伦斯, V., 克里斯托夫, D., & 威廉, G., 等. (2021). 论坛: 从公共数据库和科学文献构建知识图谱以提取化学物质与疾病之间的关联. 生物信息学 (21): 21. [10] 郝, X., 纪, Z., 李, X., 阴, L., & 杨, R. (2021). 知识图谱的构建与应用. 遥感学报, 13. [11] 杨, M., 陈, K., 孙, S., 韩, Z., & 孟, Q. (2021). 一种基于模式驱动的图排序方法用于知识图的属性提取. 工业信息论进展, PP(99): 1-1. [12] 高伊尔, P., 费拉拉, E. (2018). 图嵌入技术、应用与性能: 综述. 知识系统, 151: 78-94. [13] 一研究. (2021). 中国人工智能基础层产业发展研究报告.

