



电信运营商知识图谱技术体系研究及应用实践

赵东明

(中国移动通信集团天津有限公司, 天津 300020)

摘要: 知识图谱在电信运营商业务及运维场景的知识搜索、工单处理、智能服务、故障分析等领域发挥着重要作用, 以中国移动业务运营场景为依托, 体系化地研究了知识图谱技术体系, 基于应用场景、服务对象可分为问答检索类知识图谱、工单分析类知识图谱、系统运维类知识图谱和业务运营类知识图谱, 并介绍其技术架构、业务架构及构建方法, 引出知识图谱演进与发展的规划目标。

关键词: 知识图谱; 认知网络; 智能问答; 信息检索; 智能运维

中图分类号: TP391.1

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2022145

Research and application practice of knowledge graph technology system for telecom-operators

ZHAO Dongming

China Mobile Communications Tianjin Co., Ltd., Tianjin 300020, China

Abstract: Knowledge graph plays an important role in knowledge search, work order processing, intelligent service, fault analysis and other fields of China Mobile business and maintenance scenario. The knowledge graph technology system of China Mobile was systematically studied. Based on application scenarios and service objects, it can be divided into question & answer retrieval knowledge graph, work-order analysis knowledge graph, system operation and maintenance knowledge graph and business operation knowledge graph. The technical architecture, business architecture and construction methods were introduced, and the planning goal of the evolution and development of knowledge graph was ultimately led.

Key words: knowledge graph, cognitive network, intelligence question and answer, information retrieval, intelligent operation and maintenance

0 引言

“计算—感知—认知”的人工智能技术发展路

径, 已经成为大多人工智能研究和应用专家的共识。机器具备认知智能, 进而实现推理、归纳、决策甚至创作, 在一定程度上需要一个充满知识

收稿日期: 2021-11-27; 修回日期: 2022-06-27

基金项目: 中国移动战略级重大研发计划项目: 人工智能关键核心技术攻关 (No.R222111101021D5)

Foundation Item: China Mobile Strategic Major Research and Development Plan Project: Key Technology Research in Artificial Intelligence (No.R222111101021D5)



的大脑^[1]。知识图谱(knowledge graph, KG)将互联网的信息表达成更接近人类认知世界的形式^[2],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。知识图谱给互联网语义搜索带来了活力,同时也在智能问答、大数据分析 & 决策中显示出强大威力,已经成为互联网基于知识的智能服务基础^[3-4],已成为推动人工智能从感知能力向认知能力发展的重要途径^[5-6]。知识图谱的应用现在非常广泛:在通用领域,Google、百度等搜索公司利用其提供智能搜索服务^[7],IBM Watson 问答机器人、苹果的 Siri 语音助手和 Wolfram Alpha 都利用图谱进行问题理解、推理和问答^[8];在各垂直领域,行业数据从大规模数据到图谱化知识快速演变,且基于图谱形式的行业知识,对智能客服、智能决策、智能营销等各类智能化服务进行赋能^[9-10]。

随着电信运营商数字化转型的推进,无论是系统架构还是业务的数据都发生着重大的变化^[11],体系趋于复杂化、精细化,业务的数据在成倍增长,这些均对运营、运维工作提出了更高的要求^[12]。因此,需要运用知识图谱技术,发挥其在认知理解、智能分析领域的优势,构建业务、运维智能大脑,以知识图谱的认知关系网络为动态决策提供数据分析引擎,从而支撑相关场景^[13-14],让图谱场景应用结合行业需求,进行

业解决方案的落地,从而提升业务、运维等工作的服务质量和效率^[15]。目前,知识图谱已在电信运营商业务及运维场景的知识搜索、自然语言处理、智能服务助手、故障根因分析等领域发挥着重要作用,成为电信运营商数智化转型和发展的核心驱动力之一。

1 知识图谱的应用价值

知识图谱用节点和关系所组成的图谱,为电信运营商的各个场景直观地建模,运用“图”这种基础性、通用性的“语言”,以网状结构表达业务推荐、存量运营、系统运维、客户服务的各种关系,并且非常直观、自然、直接和高效,不需要中间过程的转换和处理。

目前知识图谱对电信运营商的应用场景和价值如图 1 所示,分类主要集中在:业务推荐及营销、电子渠道触点营销、系统运维、社交网络数据分析、服务决策和支撑、满意度评测和系统故障根因分析等领域。另外在风控管理领域中,知识图谱也可应用于反欺诈、反洗钱、互联网授信风控、保险欺诈、银行欺诈、电商欺诈、项目作假、企业关系分析等场景中。

(1) 复杂业务和运维场景中关系表达能力强
传统数据通常通过表格、字段等方式进行读取,而关系的层级及表达方式多种多样,知识图

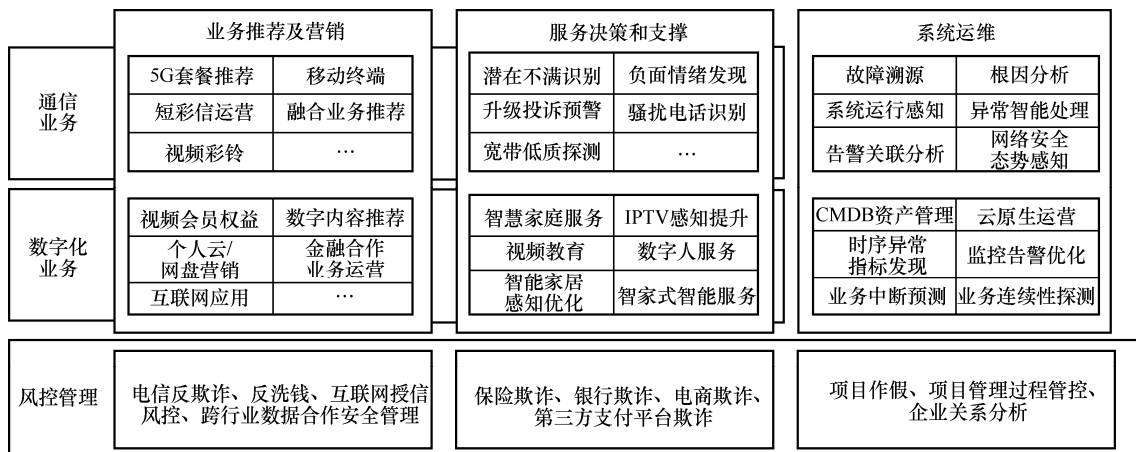


图 1 知识图谱对电信运营商的应用场景和价值

谱基于图论和概率图模型,可以处理复杂多样的关联分析,满足运营商各种角色关系的分析和管理工作需要。

(2) 具有较强的认知分析和逻辑推理能力

基于知识图谱的交互探索式分析,可以模拟人的思考过程去发现、求证、推理,业务人员自己就可以完成全部推荐、分析、表达过程,从而大大提升效率,降本增效又安全可信。

(3) 知识学习从而驱动模型自我优化

知识图谱利用交互式机器学习技术,支持根据推理、纠错、标注等交互动作的学习功能,不断沉淀知识逻辑和模型,提高系统智能性,将知识沉淀在运营商应用领域内部,降低对经验的依赖,也可以基于知识反馈不断优化模型,实现半自动化的知识模型迭代更新。

(4) 跨领域的知识数据合作实现智能策略辅助能力

知识图谱作为 AI 发展的底层技术,能够帮助运营商更好地参与到新的泛行业化的协作网络和联邦主态中,通过跨领域的企业界知识合作,更高效地训练模型辅助自身市场布局、策略优化,从而提升竞争力。

(5) 降低技术成本,促进创新技术发展

知识图谱技术体系能够有效降低业务、运维、服务人员掌握 AI 应用的门槛,扩大技术应用的范围和广度,能够为不同客户提供更加丰富的产品及服务,同时去除数据安全隐患,在提升效率和获得成长的同时,实现自身发展。

2 知识图谱技术体系架构

知识图谱是实现机器认知智能的底层支撑技术,应用场景主要体现在业务推荐、精益运维、精准分析、智慧搜索、智能推荐、智能解释、更自然的人机交互和深层关系推理等各个方面。其框架面向业务类和运维类大体一致,仅在数据源、模型构建等有所区别。

2.1 知识图谱应用分类

针对电信运营商的业务运营、系统运维场景,知识图谱因其构建方式和应用场景的不同,知识图谱技术体系可以分为四大类:问答检索类知识图谱、工单分析类知识图谱、系统运维类知识图谱和业务运营类知识图谱。

知识图谱在共享数据源的情况下,基于统一的知识图谱运营平台,构建“一体四面”的知识图谱运营体系,如图 2 所示,从而实现完整的知识图谱训练和推理应用架构,在图谱构建、营业推荐、存量保有等领域打造一系列标杆应用。

- 一体:知识图谱运营管理系统。以统一 AI 赋能、统一架构管理、统一数据支撑方式,打造知识图谱研发运营一体化的管理系统,入驻统一 AI 平台,嵌入微营销主流程,实现个性化数据选择、模型训练和客群生成。
- 四面:问答检索类知识图谱、工单分析类知识图谱、系统运维类知识图谱和业务运营知识类图谱。针对 4 类知识图谱的构建,可以在机器问答、业务推荐、系统运维等场景打造个性化应用。

(1) 问答检索类知识图谱

问答检索类知识图谱总体遵循自底向上的构建方法,系统通过从业务数据域、基础数据域、应用渠道域等获取海量非结构化数据,进行数据清洗和语义分析,并辅以原子化知识库中的固内容,以无监督/有监督两种模式自动构建面向移动业务服务的知识图谱,对业务和服务渠道提供知识检索及知识推荐能力,赋能知识库、智能应答、工单处理等领域提升效能。问答检索类知识图谱分类对应的实例,包括业务运维知识图谱、行业知识图谱等。

(2) 工单分析类知识图谱

工单分析类知识图谱,基于结构化知识的工单信息检索方法,面向自然语言领域的应答、查

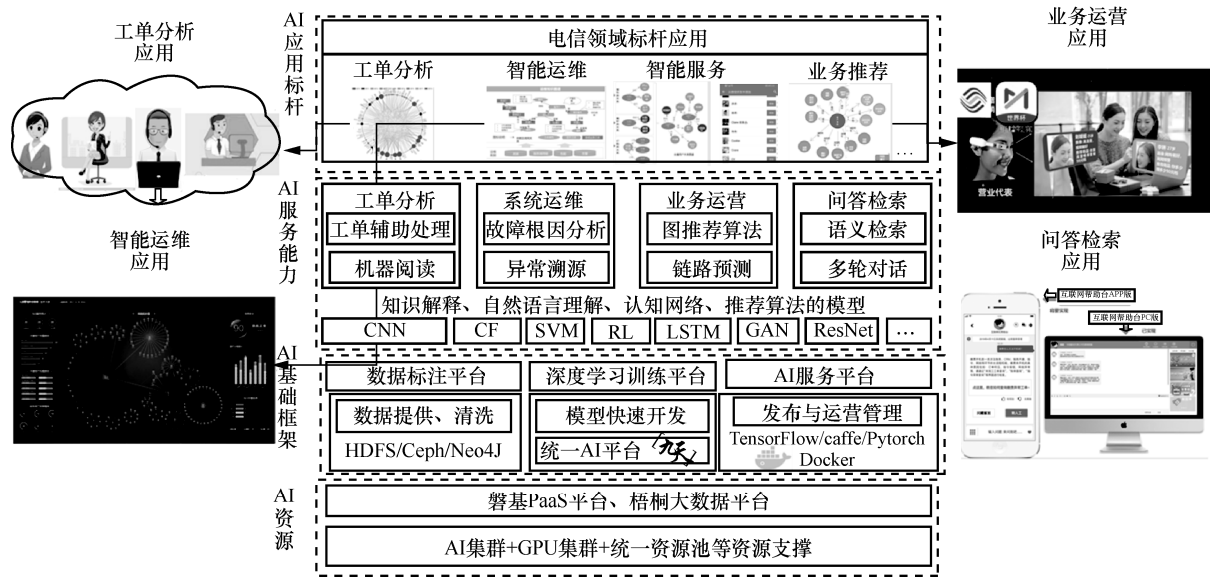


图2 电信运营商“一体四面”的知识图谱技术体系分类

询、检索等问题，该知识图谱结合自然语言理解技术对问题进行理解，并根据问题理解的结果从知识图谱中查找或推理问题对应的答案，给请求方进行反馈。

工单分析类知识图谱可以自动获取解决方案合并后形成处理意见，自动推送到工单处理人待办进行决策辅助。其对应的实例，包括工单分析知识图谱、投诉溯源分析知识图谱等。

(3) 系统运维类知识图谱

传统的监控告警手段已经无法满足高效运维管理的诉求，迫切需要引入领域知识结合 AI 方法构建有效的知识图谱，打造以知识图谱技术驱动的监控和告警智能管理体系，可视化展示告警的关联关系，对告警进行收敛，降低人工投入，提升运维效率。

通过知识图谱与认知推理技术结合，最终以逻辑表达的方式辅助运维分析、系统异常根因挖掘，在提供推理、推荐结果的同时，给出可解释的原因，在面向系统智能运维的推荐、分析、挖掘等场景反馈富语义的认知结论。

系统运维类知识图谱对应的实例，包括监控运维态势感知知识图谱、告警关联关系知识图谱、

故障定位知识图谱、故障处置知识图谱、资源快速溯源 CMDb (configuration management database) 知识图谱和资产优化知识图谱。

(4) 业务运营类知识图谱

业务运营类知识图谱，重点突出知识图谱可解释、可溯源能力，在决策辅助、根因分析、问题溯源、业务推荐、服务预测等领域体现认知能力，尤其在中国移动关键的存量运营领域，可实现基于认知推理的权益推荐、5G 换机推荐和预离网用户识别，在系统运维领域，实现告警分析、故障根因发现等。

业务运营类知识图谱分类对应的实例，包括权益推荐知识图谱、预离网知识图谱、满意度预测知识图谱、骚扰电话识别知识图谱等。

2.2 知识图谱技术架构

知识图谱技术在电信运营商领域的应用，一类是信息检索和工单文本处理，通过文本信息处理、概念提取和关系提取，实现面向信息检索和智能应答的能力输出，其技术特点主要是针对可读文本数据的处理和网络构建。一类是业务运营和系统运维的分析，主要面向系统运维和业务运营知识图谱，重点从隐语义数据（大数据画像、

偏好、标签、主机、应用、数据库、监控等)挖掘逻辑挖掘,针对业务推荐、根因分析、问题溯源等提供认知理解能力。知识图谱技术体系架构如图3所示。

(1) 数据层

知识图谱中的知识来源于结构化、半结构化和非结构化的信息资源,通过各类数据抽取工具从各类业务系统抽取到图谱系统,文本类数据,包括服务工单数据、用户标签数据、订购数据、DPI数据、家庭好友关系以及运维日志等基础数据。运维类数据,包括CMDB、服务调用链、工单、监控、日志等运维系统对接实现多源数据的汇聚及清洗,将数据整理形成层次化、结构化一致数据。标签类数据,可以提供“客户号码—标签—特征—偏好—画像”数据,家庭宽带、无线业务系统均可以提供领域感知数据。

(2) 图谱构建层

图谱构建层提供图谱结构配置、构建模型标注、构建模型训练、三元组生成、三元组审核、三元组发布等页面工具。图谱构建层预置多个智

能分析模型,实现文本分类、实体识别、机器阅读理解、文本聚类、词法/句法分析等自然语言能力遵循“模式设计→数据清洗→实体识别→关系识别→知识融合”的流程将数据层的非结构化数据转化为领域场景需要的图谱实体及关系,生成三元组。

(3) 图谱存储层

图谱存储层使用图数据库将构建层生成的三元组进行图关系存储,不同类别的图谱将分别单独存储。

(4) 图谱计算层

知识计算是基于已构建的知识图谱进行能力输出的过程,是知识图谱能力输出的主要方式。主要包括知识统计与图挖掘、知识推理两大部分内容,知识统计与图挖掘重点在于图查询、图特征统计、关联分析、异常检测、预测推理等能力;知识推理重点研究的是基于图谱的逻辑推理能力,主要包括知识搜索、知识推荐、智能补充、多跳推荐等能力。通过知识中心进行对外能力封装,为上层应用提供能力支撑。

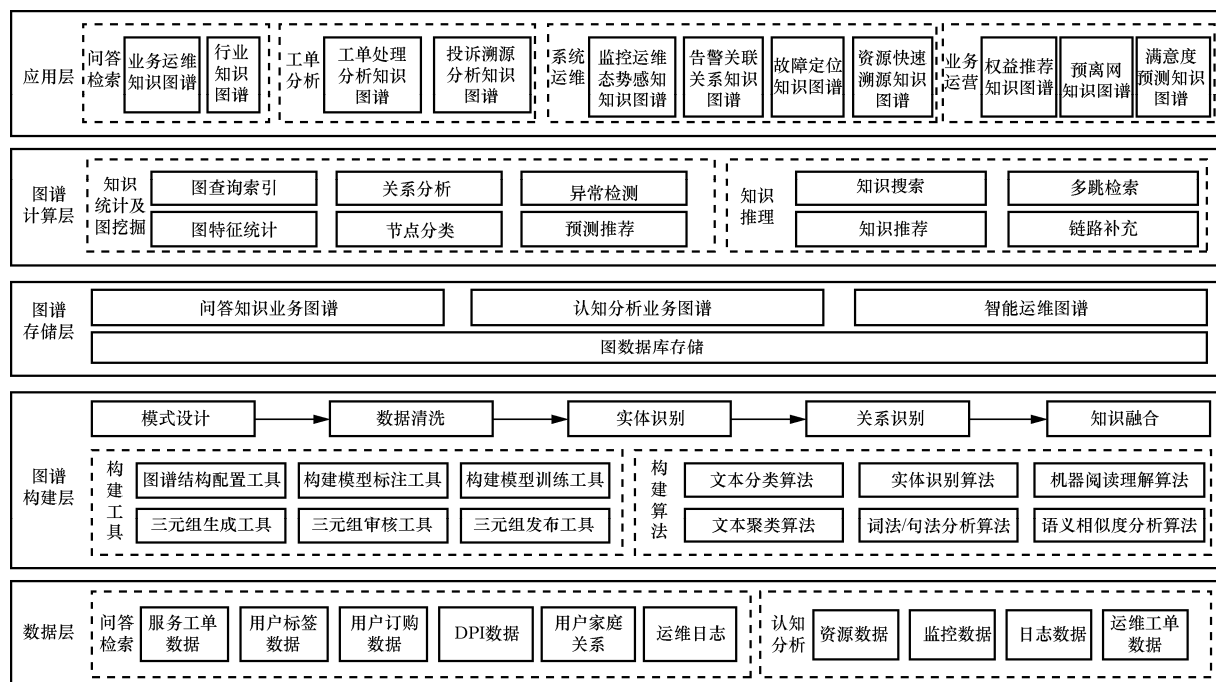


图3 电信运营商知识图谱技术体系架构



(5) 应用层

基于知识中心封装的能力,通过运维场景实现知识图谱的落地,包括基于事件工单、故障单等历史工单的分析,实现问题的关联分析并给出相应的解决方案,提高运维人员处理效率。基于告警的故障根因分析,通过对历史告警的分析结合资源关联关系实现告警收敛及告警根因定位,缩短运维人员告警处理时间,提高告警处理效率。基于资源关联关系的实体关系分析,通过对已有资源关联关系的分析,推导出新的实体关系,从而提高 CMDB 完整性与准确性,为上层运维应用的建立奠定坚实的基础,以及面向业务和服务的应用,包括智能问答机器人、智能知识库、智能工单分析、业务搜索引擎等。

2.3 知识图谱应用实例

电信运营商知识图谱技术体系中,数据层是基础数据的存储、预处理,图谱构建层、图谱存储层、图谱计算层主要针对基础数据构建图谱结构,而应用层则面向具体业务场景进行赋能,根据知识图谱分类的描述,主要分为智能问答、工单分析、系统运维和业务运营。以某大型电信运营商为例,目前具有示范意义的为 14 个应用实例,通过描述知识图谱应用实例的应用场景、服务对象、功能、拓展前景,可以系统性地梳理知识图谱从构建,到应用、服务、技术框架、算法

的整体方案,知识图谱应用及技术架构对应关系如图 4 所示。同时,知识图谱应用实例,是可回溯、可拓展的,当前的知识图谱技术方案,可以支撑问答检索类、工单分析类、业务运营类、系统运维类的场景要求,未来可根据问答、工单、运维、业务的业务发展而不断完善和更新,基于新的需求进行实例扩展,形成具有电信运营特色的一套知识图谱生态循环体系。

电信运营商知识图谱典型应用实例如图 5 所示,主要为以下内容。

(1) 业务运维知识图谱

主要应用于机器人智能应答场景,在智慧化拟人问答领域体现价值。通过对用户问题进行分词、长句理解、语义解析,提取关键语义并与知识图谱模板的匹配,获取图谱实体或属性内容。在图数据库查询并完成消息整合后,以机器人形式与用户进行服务应答,通过知识图谱丰富的逻辑、关系能力提升机器人智能感知。

(2) 行业知识图谱

主要应用场景为行业客户智能服务,对行业数据进行要素提取,并按照业务逻辑进行关系生成及存储,基于用户咨询历史、工单处理记录、企业运营信息等文本信息,提取关键语义、解决方案、问题溯源、核心语义等内容,赋能行业智能应答系统进行自服务、预服务,赋能知识库系

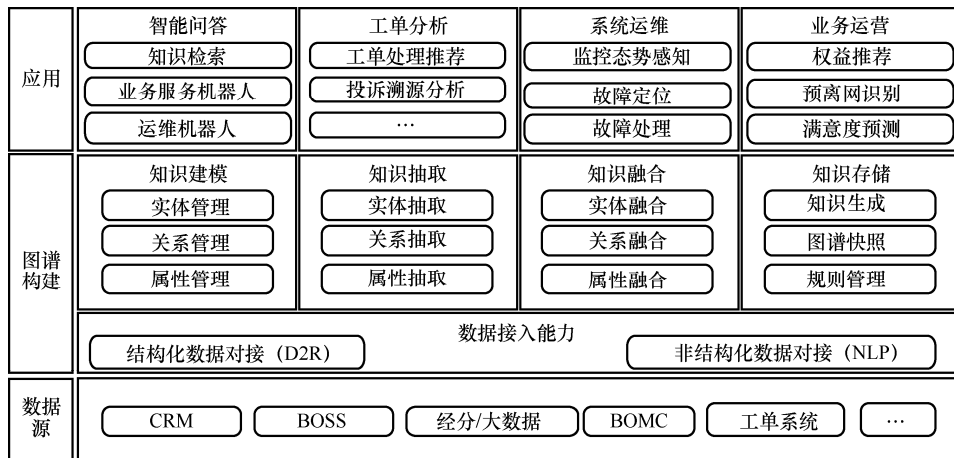


图 4 电信运营商知识图谱应用及技术架构对应关系

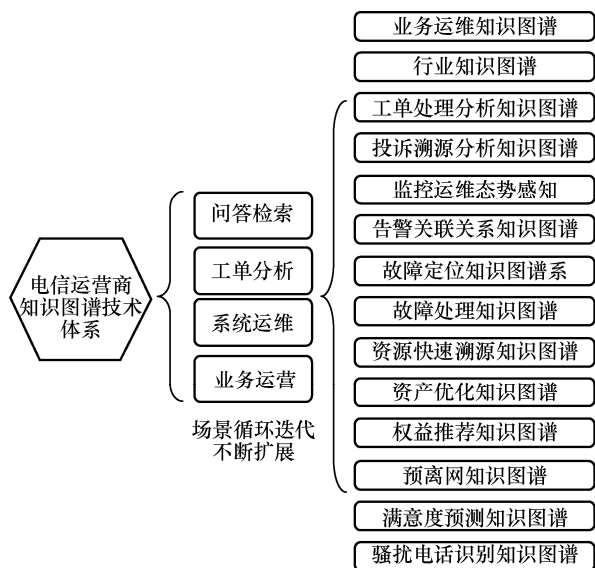


图5 电信运营商知识图谱典型应用实例

统进行图谱式语义搜索，赋能个性化推荐系统进行智慧化服务推荐。

(3) 工单分析知识图谱

主要应用于投诉工单智能分析场景。通过对工单内容进行实体提取和语义理解，并与知识图谱模板的匹配，获取解决方案合并后形成处理意见，自动推送到工单处理人待办进行决策辅助，实现“人—机”协同的投诉工单的自动分析、自动处理、自动回复。

(4) 投诉溯源分析知识图谱

应用于工单分析和溯源。主要基于投诉历史数据，引入语义分析、知识图谱等人工智能技术，构建投诉从产生到解决修复上线全程溯源分析能力，包括全过程溯源追踪、投诉问题根因溯源、投诉处理质量问题溯源分析等，辅助投诉处理及管理人员，提升投诉处理及压降分析过程中的效率和质量。

(5) 监控运维动态感知知识图谱

从海量的日志、调用链、资源库等众多的信息中发现资源之间的图谱关系，构建运维知识图谱，当系统发生故障时，利用运维图谱给运维人员提供系统监控工具，帮助维护人员进行故障影响分析，并引入告警根因 AI 算法能力，实现业务、

应用、平台三层告警根因推导，配合服务调用分析、日志分析、指标异常检测等智能化手段，实现问题快速定位。

(6) 告警关联关系知识图谱

通过对各类数据的挖掘动态构建运维图谱，实现图谱关系实时准确，最终实现故障根因分析结果的推荐，可以辅助运维人员快速定位及解决问题，同时通过运维人员对根因分析结果的不断优化，为运维无人驾驶发展奠定了基础。

(7) 故障定位知识图谱

应用于 SRE (site reliability engineer)/运维人员，通过结合领域相关知识，借助设备固有拓扑结构、应用调用关系、历史告警数据、性能数据、故障数据等梳理相关的实体、属性、关系，利用自然语言处理、因果关系算法、路径搜索算法等 AI 算法从文本告警信息映射为一条完整的根因链路，更好地为 SRE/运维人员提供指引，且具有更强的解释性。

(8) 故障处置知识图谱

应用于故障处理及系统运维人员。通过对历史故障处理方案的梳理及分析，沉淀专家经验、故障分析报告内容、梳理故障与处理预案之间的关系、处理预案之间的关系等，提炼通用化故障处理方案，建立决策中心，为一线运维人员提供故障处理建议及可沉淀的处理作业预案，推动故障的快速处置。

(9) 资源快速溯源知识图谱

通过对资源对象全生命周期的信息以及资源对象之间的关系（包括物理关系、逻辑关系和依赖关系）的梳理及分析，沉淀全局视角下的软硬件知识图谱，实现显性及隐性关系的挖掘，助力一线运维人员，帮助其对庞大体系下的复杂应用和设备进行管理和运营。

(10) 资产优化知识图谱

通过对历史资源使用情况进行分析，沉淀资源对象与适配资源使用情况的关系，实现显性及



隐性关系的挖掘,助力一线运维人员,帮助其在对不同资产资源的使用进行管理和配置,对不同资源给出最优的设备使用状态建议,对低效资产进行回收、对高使用资产进行及时扩容。

(11) 权益推荐知识图谱

应用于电信运营商市场存量运营人员。针对电信运营商、互联网行业重点推荐的权益优惠、产品、营销活动,提供最佳目标客户群,帮助存量运营人员自动化获取每名用户针对每一种权益优惠的推荐概率,并输出到外呼营销、CRM(customer relationship management)弹窗、触点营销、电子渠道营销等,进行针对性业务推荐。

(12) 预离网知识图谱

应用于潜在流失客户智能看护场景,建立以用户离网流失预警为目标的客户行为异动识别看护系统,引入随机森林、KNN(K -nearest neighbor)分类算法和知识图谱,通过对用户每日行为监控、异动行为发现、流失用户识别、运营结果回馈,形成客户高危自识别、模型自优化、结果自呈现、运营自执行的智慧看护系统。

(13) 满意度预测知识图谱

用于挖掘潜在不满意用户,给予精准维系和服务。基于大数据标签进行构建,纳入“原子一概念”生成算法实现语义延伸,在传统大数据分析基础上加入语义理解能力,构建以用户号码节点为中心的服务分析图谱,链接用户服务评测结果(不满意)、大数据标签、原因概念,并实现全量用户潜在不满意的推理,输出用户群进行后续维系。

(14) 骚扰电话识别知识图谱

用于识别电信运营商网内骚扰电话号码,并进行安全管控。将现全网手机用户和大数据属性标签作为输入构建骚扰电话识别知识图谱,并利用紧密度算法构建骚扰电话识别模块,提供骚扰电话识别能力,对判定为骚扰电话的号码进行关停处置。

3 电信运营商知识图谱功能与构建

3.1 知识图谱功能框架

电信运营商知识图谱的功能框架主要体现在整体业务、服务、运维、管理中,主要为文本类应用和分析类应用。对于文本类,知识图谱通过文本信息处理、概念提取和关系提取,实现面向信息检索和智能应答的能力输出,其技术特点主要针对可读文本数据的处理和网络构建。针对分析类,知识图谱重点从隐语义数据(大数据画像、偏好、标签、主机、应用、数据库、监控等)挖掘逻辑挖掘,针对业务推荐、根因分析、问题溯源等提供认知理解能力。

电信运营商知识图谱的功能框架,可以描述各种与知识相关的实体和概念之间的关联关系,从而更快捷准确地进行智能问答和信息检索。知识图谱最常用的语义关系为“实体—关系—实体”与“实体—属性—属性值”,电信客户服务领域存在各种复杂异常、故障、事件,知识图谱是描述其内在联系的最佳载体。知识图谱是将知识实体之间的关系和逻辑进行图形化展示的语义网络,作为中台核心组件,为客户服务域的知识库输出逻辑搜索能力,为智能机器人输出智慧应答能力,为工单处理输出自动回复能力。为 CRM/BOSS(business & operation support system)等营业前端,提供业务/服务/数据/事件之间复杂关系的结构化输出,串联运营与运维,为运营业务、服务、营销、运维等全量场景提供图谱化的实体逻辑支撑,驱动业务服务一体化,服务支撑智能化,海量数据处理自动化。

电信运营商知识图谱功能框架如图6所示,可以用于经营决策、客户服务的分析并具备理解和解释能力,从而在电信领域的认知理解、数据分析、决策支持领域进行能力输出。一方面通过建立从数据到认知图谱中实体、概念、关系的映射,让机器理解数据的本质;另一方面利用认知图谱

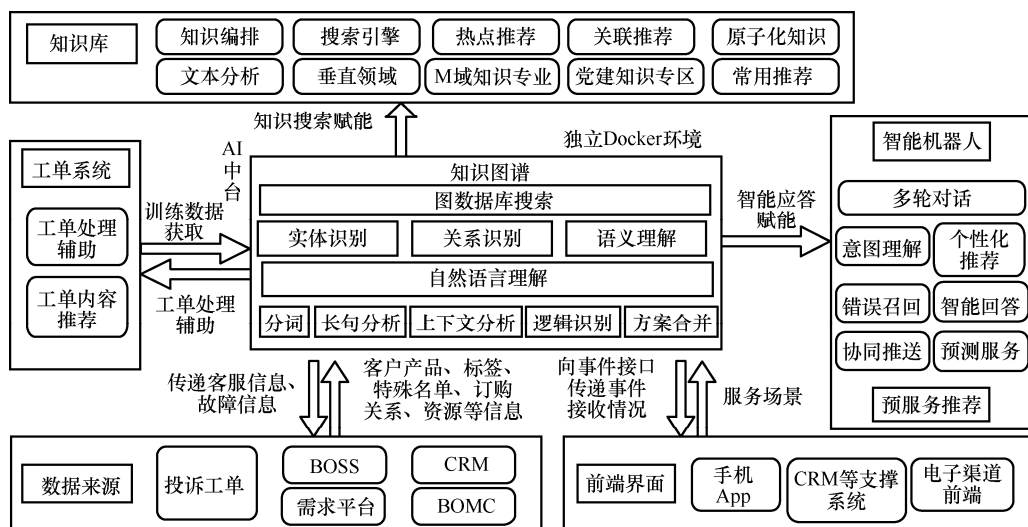


图6 电信运营商知识图谱功能框架

中实体、概念和关系解释现实世界中的事物和现象，让机器解释现象的本质。以市场领域的存量运营为例，基于目前大数据标签库、历史投诉工单、业务管理文档等基础能力和数据，打造面向存量运营“场景”的认知网络，包含面向运营场景的概念层（概念标签）、原子层（原子标签）、分层理论库（基于原子属性分层链接实体）、运营策略库（面向场景的维系、推荐、服务策略）提升运营推荐的精准度，并面向“客户需求”提供可理解、可解释的运营策略。

3.2 知识图谱构建方法

电信运营商知识图谱构建遵循自底向上的逻辑，通过从业务数据域、基础数据域、应用渠道域等获取海量非结构化数据，进行数据清洗和语义分析，并辅以原子化知识库中的固有内容，以无监督/有监督两种模式自动构建面向移动业务服务的知识图谱，针对运营商复杂的投诉和异常场景提出针对性解决方案。

构建知识图谱的步骤分为模式设计、数据清洗、实体识别、关系识别以及知识融合。

(1) 模式设计

模式是知识图谱的概念模型和逻辑基础，借助实体关系的规则定义，约束知识图谱的数据。

模式设计是知识图谱构建的基础，基于现有事件单、投诉单、异常单、客户画像、网络数据，设计一套图谱自动抽取的框架，以运维为例面向故障的模式包括故障内容描述、故障原因、故障解决方案、故障归属、故障详细描述、故障发生地点、客户号码、客户资费、关联需求等。

(2) 数据清洗

基于事件单、异常单、运维总结报告及运营商领域专业词汇进行数据清洗，得到含有相关语义信息的文本内容，用作后续文本语义分析的基础。待清洗数据包括规范化的工单数据、报告数据，也包括非结构化的日志数据、问答数据，数据清洗质量直接关系到模型的训练精度。

(3) 实体识别

实体识别是知识图谱语义分析的关键。采用机器学习技术自动或半自动地从多源数据中提取知识图谱的实体、关系、属性等要素。知识抽取包含实体抽取、关系抽取和属性抽取。

实体识别的工具是 LSTM (long-short term memory) + CRF (conditional random field)，通过 LSTM 网络的处理，得到了输入数据的表示方法，在输出端将 softmax 与 CRF 结合起来，使用 LSTM 解决提取序列特征的问题，使用 CRF 有效利用了



句子级别的标记信息。

(4) 关系识别

关系识别是对非结构化文档处理的重要环节,自然语言中对事件的描述往往会以多个子句的形式存在,且多句之间具有相关性,关系识别需要对投诉工单数据的上下文,进行依存句法相关性分析,从而找出实体间存在哪种关系,以及关系周边的实体分别是什么,例如运维场景中“A导致了B”“C的故障原因是D”“E的变更产生了F现象”,面向运维场景的关系识别是构建RDF(resource description framework)三元组的基础。

(5) 知识融合

知识图谱在进行知识抽取时所使用的数据源是多样化的,存在知识重复、知识间关系不明确等问题。知识融合可消除实体、关系、属性等与事实对象间的歧义,使不同来源的知识能够得到规范化整合。为了让运维知识图谱内容在客服领域具备通用性,本文对抽取出的三元组内容进行二次分析,将根据词性及依存句法抽取出的实体进行聚类,将涉及同类内容的实体进行融合,组成完整语义的回复答案,反馈到知识库及机器人前端。

3.3 知识图谱技术方法及参数

知识图谱构建过程中应用的人工智能技术算法,主要包括以下3种。

- 实体识别:长文本分词算法、层次聚类算法、Bi-LSTM算法、CRF算法。
- 关系识别:词法&依存句法分析,Bert模型(大规模预训练网络)。
- 知识融合:相关性分析算法,相似度计算方法。

知识图谱构建后,以接口方式对应用赋能,具体输入输出数据如下。

(1) 输入数据

在训练过程中,输入主要为投诉处理工单、用户问题记录等信息,需要包括关键信息:故障

内容描述、故障原因、故障解决方案、工单工作、工单处理人、投诉类别、关联问题单、需求单等。在应用过程中,输入为客户的一种问题表述“故障现象”,待获取对应的“故障原因”“问题解决方案”或工单下一步动作、处理人等信息。

知识图谱训练数据以RDF方式存入关系数据库,即Neo4j。

(2) 输出数据

根据输入的问题,在知识图谱关系数据库Neo4j中查询对应的故障原因、解决方案、下一步动作、处理人信息,反馈到请求方。

4 电信运营商知识图谱演进发展

电信运营商知识图谱技术体系,是基于实际需求总结归纳的符合电信特色的知识图谱应用、实例、系统建设运营方法理论,实现对内外部业务的支撑发展。其演进主要分为3个阶段:探索实践阶段、优化提升阶段和中台赋能阶段,以集中化和能力分享两种方式持续推进,前期由技术先进、运营经验丰富的省专公司进行重点知识图谱实例建设,并在业务场景产生实际效果,后期以集中化方式进行全网知识图谱能力聚合、资源集约化、承接全网集中化知识图谱应用需求。

4.1 探索实践阶段

面向智能应答、工单分析、系统运维、业务运营等领域的重点需求场景,研究知识图谱的构建、训练和推理技术方法,并实现基本功能实现和能力输出;形成数据标注、知识图谱体系搭建、知识迭代优化和协同运营等运营流程,结合场景化需要,实现“人—机”协同运营;针对智能应答、工单分析、系统运维、业务运营四大场景,启动一系列知识图谱实例的研究,完成解决方案规划、数据准备和模式设计,并实现重点急需实例的建设,如业务运维、工单分析、故障定位等,完成统一的知识图谱指导白皮书。

4.2 优化提升阶段

面向文本数据、标签数据、运维数据等,完成一系列个性化知识图谱技术方案,如结合机器阅读、实体识别、关系识别、摘要提取、链路预测等机器学习模型的知识图谱技术方法,以解决方案方式完成需求,并将不同模式的知识图谱归纳总结、体系化;形成成熟的业务人员、技术人员协同的知识图谱运营体系,在关键语料库梳理、文本匹配、文本分类标注等方面不断融合协同,形成成熟的迭代运营团队;以知识图谱实例建设成果为基础,形成较完善的知识图谱知识体系、算法体系和方案体系,在业务运营、服务运营、系统运维和问答检索场景中产生效果,并探索知识图谱集中化建设赋能的可行性,指导全行业以自建或者集中化复用方式,广泛开展知识图谱实践工作。

4.3 中台赋能阶段

梳理架构实现中台化管理,统一的运维知识、业务知识、大数据特征知识等,以中台组织方式实现集中化知识管理,并完成基础框架建设,总体资源集约、运行高效,AI与业务形成合力;探索知识图谱集中化赋能的“集团—省”两级协同模式,逐步扩大知识运营范围和赋能范围;建立完备的运营体系,通过内外部运营需求拉动能力运营,赋能业务高质量发展;知识图谱中台能力归纳、收敛、汇聚,形成若干通用知识图谱实例能力,沉淀的共性能力比较丰富,广泛发挥注智赋能作用。同时,支持各省属地化需求进行能力拓展和灵活调度。

5 结束语

本文主要介绍了面向电信运营商的知识图谱技术体系及应用实践方法,通过研究知识图谱架构、分类、技术、服务对象、应用场景、生态循环解决方案等内容,结合现有优秀案例经验,打造了面向外部客户以及内部运维人员的知识图谱

应用实例,包括面向外部客户的权益推荐、预离网分析、满意度预测知识图谱,面向内部业务人员的骚扰电话识别知识图谱,面向内部运维人员的工单分析、投诉溯源分析、监控运维态势感知、告警关联关系、故障定位、故障处置、资源快速溯源CMDB、资产优化知识图谱,以及面向内部和外部用户共用的业务运维、行业知识图谱,并将知识图谱服务于智能问答、日常的运营分析、工单处理流程、系统运维等,体现了知识图谱技术的价值。

未来,在知识图谱完整体系搭建完成后,可服务的电信行业应用形态、服务时采用的具体技术手段均是探索的课题,例如在智能问答、智能客服、智能运维、知识搜索等系统进行交互时,知识数据存放位置、训练数据预处理方法、调用算法和模型方法等。同时,电信运营商知识图谱生态循环如何演进也是需要研究的问题,知识图谱构建后需随着业务、管理的不断更新而迭代优化,系统架构、知识关系也需变化,需要确保知识图谱的生态化演进方式,降低对专家经验的依赖,真正建立知识的时效性、关联性、服务性、共享性的一体化循环迭代方案,提升系统的运营体验,以知识图谱技术赋能运营管理顶层设计,成为数字化变革的新引擎。

参考文献:

- [1] PAN J Z, EDELSTEIN E, BANSKY P, et al. A knowledge graph based approach to social science surveys[J]. Data Intelligence, 2021, 3(4): 477-506.
- [2] HUANG X Q, TANG J Y, TAN Z, et al. Knowledge graph embedding by relational and entity rotation[J]. Knowledge-Based Systems, 2021(229): 107310.
- [3] ZHENG W B, YAN L, GOU C, et al. Pay attention to doctor-patient dialogues: Multi-modal knowledge graph attention image-text embedding for COVID-19 diagnosis[J]. Information Fusion, 2021(75): 168-185.
- [4] TOSI M D L, DOS REIS J C. SciKGraph: a knowledge graph approach to structure a scientific field[J]. Journal of Informetrics, 2021, 15(1): 101109.
- [5] WANG Z K, LI L J, ZENG D, et al. Incorporating prior know-



- ledge from counterfactuals into knowledge graph reasoning[J]. Knowledge-Based Systems, 2021(223): 107035.
- [6] SHA X, SUN Z, ZHANG J. Hierarchical attentive knowledge graph embedding for personalized recommendation[J]. Electronic Commerce Research and Applications, 2021(48): 101071.
- [7] LIN Q K, ZHU Y F, LU H, et al. Improving university faculty evaluations via multi-view knowledge graph[J]. Future Generation Computer Systems, 2021(117): 181-192.
- [8] ABU-SALIH B, AL-TAWIL M, ALJARAH I, et al. Relational learning analysis of social politics using knowledge graph embedding[J]. Data Mining and Knowledge Discovery, 2021, 35(4): 1497-1536.
- [9] BUCHGEHER G, GABAUER D, MARTINEZ-GIL J, et al. Knowledge graphs in manufacturing and production: a systematic literature review[J]. IEEE Access, 2020: 55537-55554.
- [10] XUE Y Y, JIN J H, SONG A B, et al. Relation-based multi-type aware knowledge graph embedding[J]. Neurocomputing, 2021(456): 11-22.
- [11] NIU H R, HE H T, FENG J Z, et al. Knowledge graph completion based on GCN of multi-information fusion and high-dimensional structure analysis weight[J]. Chinese Journal of Electronics, 2022, 31(2): 387-396.
- [12] DANG F R, TANG J T, PANG K Y, et al. Constructing an educational knowledge graph with concepts linked to Wikipedia[J]. Journal of Computer Science and Technology, 2021, 36(5): 1200-1211.
- [13] WANG W G, XU Y L, DU C H, et al. Data set and evaluation of automated construction of financial knowledge graph[J]. Data Intelligence, 2021, 3(3): 418-443.
- [14] CHEN Y. Constructing bug knowledge graph as a service for bug search[J]. International Journal of Software Engineering and Knowledge Engineering, 2021, 31(1): 81-98.
- [15] PU F, ZHANG Z, FENG Y, et al. Learning context-based embeddings for knowledge graph completion[J]. Journal of data and Information Science, 2022, 7(2):23-29.

[作者简介]



赵东明（1984—），男，博士，中国移动通信集团天津有限公司正高级工程师、AI实验室负责人，主要研究方向为知识图谱、情感计算及认知网络。