

2023 年东南科技论坛

“智能算力赋能千行百业” 优秀论文集

福建通信科技

2023 年东南科技论坛
「智能算力赋能千行百业」优秀论文集

◎ 2023 年东南科技论坛

“智能算力赋能千行百业” 优秀论文集

福建省通信学会
福建省互联网协会
福建省信息通信行业协会
福建省邮电规划设计院有限公司

闽内资准字K第111号
内部资料 免费交流

序 言

为打造规模大、层次高、实效好、影响力强的福建科技高端学术品牌，促进科技经济融合和产业创新发展，更好服务福建高质量发展，福建省科学技术协会创办的“东南科技论坛”已成为一年一度全省科技交流合作盛会。由福建省通信学会、福建省互联网协会、中鸿达信息科技有限公司联合承办的“智能算力助力数字经济产业融合发展”是2023年东南科技论坛的重要活动之一，它以开放共享、融合创新、变革转型、引领领域为推进目的，共同探索智能算力在赋能产业发展、促进数实融合方面如何发挥其巨大作用，以推动算力产业高质量发展，为我省数字经济发展注入新动能。

随着新一轮科技革命和产业变革，数字经济迈入高速发展阶段，算力成为激活数据要素潜能、驱动经济社会数字化转型的新引擎，正以新型生产力的姿态站在历史前沿。构建智能算力互联网是我省数字经济发展当下的关键。为确保2023年东南科技论坛——智能算力助力数字经济产业融合发展论坛的成功举办，承办单位联合发文向各相关企事业单位、高等院校、科研机构、设计院的技术专家学者诚征以“智能算力赋能千行百业”为主题的优秀论文。经论文评审委员会专家审核，有43篇论文入选本论文集汇编出版。通过学术论文征集活动，调动了科技工作者的创造、创新精神，使谋划创新、推动创新、落实创新成为自觉行动，鼓励信息通信业科技工作者奋发向上的创新热情。

在此，对参与本次活动的单位及个人表示衷心感谢，我们将不断探索、创新，搭建更多不同类型的学术交流平台，为广大科技工作者服务，为我省信息通信科技的发展做出应有的贡献。

编者

2023年8月18日

目 录

1. 基于算网调度的新型云电脑研究 伍竹歆 苏 蕾 杨慰民 (1)
2. 数字孪生流域大模型的建设思路及场景应用实践 马森标 陈惠祥 (8)
3. 视频监控 AI 产品供给及运营 傅 玮 林 强 江 雨 (23)
4. 智能算力赋能数字政府管控探索：基于数字孪生的视角 朱惠斌 (31)
5. 视频中台与边缘云技术融合应用的研究与实践 李 涵 (36)
6. 智能算力赋能园区数字化转型 范永圣 (41)
7. 基于智算引擎的 AI 应用平台的关键技术与实现 傅 玮 (45)
8. 基于企业 MEC 部署视频算力网络边缘服务的方案研究
..... 詹丰源 骆晓霞 林启笔 张浩宇 林泽剑 游思萍 (52)
9. 基于信息间隙的虚拟能源枢纽工厂响应模型的多目标优化 范林元 林 耿 刘鲤晞 (58)
10. 关系型数据库智能算力实现智能化终端感知画像 潘 嘉 杨慰民 (70)
11. 基于智能算力的智慧园区中智能楼宇管理和节能优化 范永圣 (74)
12. 港口卸船机落料口智能停车定位系统 陈 璟 高伟铭 (78)
13. 算力驱动下的精准医疗发展 李 晨 (89)
14. 基于云计算和 5G 的智能算力赋能仓储物流行业 董帝焱 胡文喜 (94)
15. AI 算力的发展与未来挑战 李 萌 (99)
16. 福建省数据中心建设布局体系研究 张宏锦 (103)
17. 基于算网协同的智慧社区居家养老应用方案研究
..... 林启笔 谢丹卉 张浩宇 骆晓霞 林泽剑 游思萍 詹丰源 (109)
18. 面向 5G 智能网络的实时车辆检测方法 曾泽鹏 (114)
19. 基于互联网 GIS 的无线网络优化可视化分析 董帝焱 黄炳塔 杜丕加 吴晶晶 何晋毅 (119)
20. 智能算力驱动下的数字经济产业融合：机遇与挑战 李 晨 (124)
21. 算云融合赋能千行百业 夏少良 (130)
22. 5G 无线通信中的新型索引调制系统研究 范世文 (135)

23. 单相浸没式液冷 BBU 机柜的优化和模拟分析	张亚洲 沈 斌 (141)
24. 基于 MEC 打造 5G 智慧工厂	陈 锋 李 岩 潘晓宇 (146)
25. 一种基于六维业务特征分析的 PCDN 流量检测工具	陈倩雯 林 伟 杨慰民 (153)
26. 智能算力与医疗产业的融合发展	李 晨 (159)
27. 基于 AI 算法的无人机巡河目标检测应用	黄炳塔 董帝焱 杜丕加 吴晶晶 (165)
28. 5G 全连接工厂关键技术的研究与实践	陈 锋 吴 帆 洪林梦涵 (170)
29. 云网融合算力基础设施应用与发展	陈邦文 (178)
30. 厦门联通算力赋能国贸 AI 视频云	林凯奇 (184)
31. 边缘计算在电化学储能场景中的应用探索	蔡 敏 (187)
32. 智能算力赋能千行百业	张 欢 (196)
33. 基于互联网数据的城市 3D 建筑射线追踪模型研究及应用	董帝焱 许绍松 黄炳塔 杜丕加 吴晶晶 (201)
34. 一种全方位立体监控告警方法的研究与应用	郭龙缘 (210)
35. 一种利用数据行为进行 VoLTE 指标快速动态监测方法	郑银云 杨慰民 (214)
36. 基于机器学习的基站节能方案研究	黄炳塔 董帝焱 杜丕加 吴晶晶 (218)
37. 基站工参平台化管理的前瞻性应用方案	黄炳塔 董帝焱 吴晶晶 杜丕加 (227)
38. 基于 Flink 的网络设备性能实时监控系统的设计与实现	王振峰 陈志安 陈 文 郑银云 杨慰民 (235)
39. 基于 MEC 打造 5G 智慧园区	陈 锋 洪林梦涵 王丽春 (240)
40. 智能算力在互联网与信息通信业中的 融合发展研究	郑文彬 (246)
41. 双谱估计方法计算精度分析	吴文兵 (250)
42. 新时期算力基础设施发展的思考	曹 曦 (254)
43. 基于全过程工程咨询模式的数据中心建设项目探索与应用	高扬明 (259)

基于算网调度的新型云电脑研究

伍竹歆 苏蕾 杨慰民

(中国移动通信集团福建有限公司)

摘要: 随着移动互联网的发展,企业和个人用户对移动办公、远程无感、数据安全提出了更高的要求。依靠高效的互联网传输和先进的虚拟化技术,云电脑可以很好地满足企业的需求。但传统云电脑存在漫游状态接入时延长、个性化定制能力弱、服务器资源无法共享等问题,本文基于泛在融合的算力网络,融合算网智能调度、算网智能编排及边缘计算等关键技术,为企业和个人用户提供一种统一管理、高效安全、稳定流畅的新型云电脑整体解决方案。

关键词: 云电脑;算网调度;算网编排

引言

在信息化浪潮的推动下,云电脑成为下一代信息技术产业的重要组成部分,广泛应用于行政服务大厅、日常办公管理、高校多媒体教室、公共图书馆电脑、医疗桌面云等多种场景,解决了传统计算机构建局域网存在的成本昂贵、维护管理困难、数据安全性低、资源利用率不充分等一系列问题。但是,随着移动互联网的发展,现有云电脑漫游状态接入时延长、服务器资源无法共享、网络安全等问题对于云电脑与题也逐渐显现。

随着全社会数智化转型的不断深入,算力需求的不断增长,算力供给方式的改变以及以智能化、边缘计算为主的新型业务的发展,算力业务和 5G 网络也呈现出深度融合的发展趋势^[1],使得用户能够以更快的速度、更低的成本获得更优质的计算服务,相关行业和企业也将会不断受益于它所带来的效率提升和成本降低,算力网络为数字经济发展提供了重要支撑。

本文旨在基于算力网络的智能调度和编排,为

企业和个人用户提供一种服务质量更高、网络更安全的新型云电脑解决方案。通过搭建算网大脑基础架构、构建 5G MEC 边缘算力资源模型及保证确定性网络时延,针对算力资源和网络资源开展解构和调度的协同编排研究,验证了算网大脑的算力接入统管、算网动态感知、算网协同编排三大功能,为云电脑用户提供更好的感知和使用体验。

1.算网调度的关键功能

算力网络的结合,关键在于实现智能化的算网调度,如图 1 所示,本节主要从泛在算力统管、算网动态感知、算网协同编排三方面进行分析。其中,“泛在算力统管”指通过 5G 边缘云管理系统,对包括第三方算力在内的资源统一输入、度量、管理;“算网动态感知”指通过动态采集边缘算力资源性能并输出可视化能力呈现;“算网协同编排”指根据动态感知的算力资源,匹配前端用户需求,实现对算力资源和网络的最优定位、调度、部署和开通。

作者简介:

伍竹歆: 中国移动通信集团福建有限公司网络部网管中心员工

苏蕾: 中国移动通信集团福建有限公司泉州分公司网络部员工

杨慰民: 中国移动通信集团公司首席专家;中国移动通信集团福建有限公司网络部网管中心副总经理

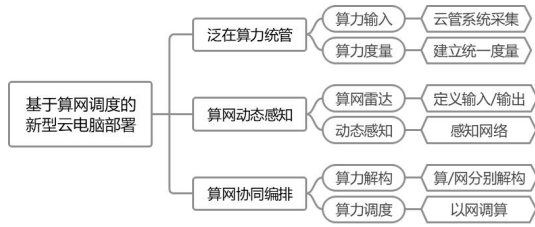


图 1：算网调度的关键功能

1.1 泛在算力统管



序号	资源对象英文名称	资源ID	C名称
1	Manager	管理信息	管理
2	Region	区域	区域
3	Host	物理主机	物理主机
4	HostPart	物理主机-主机端口	物理主机-主机端口
5	VirtualNet	虚拟网络	虚拟网络
6	Project	租户	租户

1.1.2 算力建模

在算力资源输入采集的基础上，对算力节点、算力产品、算力资源池进行建模，具备将各类异构算力(CPU、GPU、NPU、FPGA 等)以及各云平台上不同层次的算力产品(IaaS 的虚机，PaaS 层的 AI、大数据等能力)统一化和标准化的能力，以实现业务逻辑与底层资源的解耦。

算力资源池建模按基本信息配置、接入凭证配置、接入能力设计、接入区域设计、接入配额设计五大步骤进行。

(1) 基本信息配置：对要接入的算力资源池的名称、编码、性质、服务地址等信息进行配置。

(2) 接入凭证配置：对要接入的算力资源池的凭证信息进行配置，认证方式支持 AK_SK、用户名密码两种模式。

(3) 接入能力设计：对要接入的算力资源池的云服务能力进行设计，系统将依据此配置进行云服务能力的按需接入。

(4) 接入区域设计：对要接入的算力资源池的接入区域进行设计，系统将依据此配置进行云服

1.1.1 算力输入

为满足对算力资源准确性和实时性的要求，算网大脑需要统一采集包括 MEC 边缘算力资源数据，主要包括：算力类型、软硬件版本、容量、负载、状态、位置、IP、性能数据、实时告警等。算网大脑与边缘计算管理模块 MEO 的 RESTful API 接口对接，使用 HTTP 协议与 JSON 消息格式，从 MEO 获取当前 MEC 上的 CPU、内存、磁阵等资源信息。^[2]

务能力的按需接入。

(5) 接入配额设计：对要接入的算力资源池的配额限制进行设计，系统将依据此配置进行云服务能力开通的选池操作的调度。

1.1.3 算力度量

通过算力资源池概览，对当前已接入的所有算力资源进行一个清晰的展示，呈现信息包括资源池的名称、编码、云类型、健康评分、计算负载、存储负载、网络负载、算池描述、资源池标签，结合对算力资源信息的周期性定时采集，实现对算力资源池的健康评分以及对计算、存储、网络等各项负载指标的计算、评估。

1.2 算网动态感知

通过算网动态感知功能，可实时查看算力网络纳管的算力资源池总体情况及详情，如对算力节点、类型、状态、容量、负载，各网络域的资源拓扑、端口流量、路径流量、告警状态等实现动态感知，确保闲忙时资源可以被合理调度。

借助算网雷达，选择出对应用户需求的业务解

决方案, 搜索匹配符合条件的算力资源池, 并提供性能最优、成本最优、综合最优等多种方案供客户选订。

1.3 算网协同编排

1.3.1 算力解构

算力解构实现了从云电脑用户意图到网络理解的敏捷转换, 将算力资源和网络能力分别解构独立封装, 针对不同硬件类型的各种算力, 建立统一的度量指标体系, 以便在匹配用户需求时搜索可用资源, 供算力网络调度使用。

1.3.2 算力调度

算力调度有以网调算和以算调网两种方案, 以网调算指根据用户(UE)位置, 确认覆盖的 GNodeB 后, 选择靠近用户的 UPF 提供算力服务; 以算调网指根据算力资源池部署位置、寻找最接近的 UPF 为客户提供算力服务。

对于上述两种方案的选择, 从业务质量上考虑: 需要判断算力是否具备、网络(算力到 UPF 的网络、UPF 到 5GC、UPF 到 UE 的网络)是否具备, 若以 UPF 作为锚点, 可减少算力查勘、算力到 UPF、UPF 到 5GC 的网络查勘, 专注于 UPF 到 UE 之间的网络搭建即可, 整个流程相对会简单; 从业务体验上看: 用户漫游情况下, 以网调算方案能够为客户提供最优网络路径的就近服务, 降低时延提升业务感知。综合以上分析, 本文采用以网调算方案进行测试。

对于算网雷达算力调度策略, 分为四步骤进行。

(1) 根据解构的算力需求, 按照算力基本单元查找符合需求的算力资源池。

(2) 根据需求是否可拆分, 分别计算算力资源池能力容量:

● 对于不可拆分的需求, 直接根据算力需求

来计算所需容量, 然后根据算力资源池的剩余容量进行判断, 过滤出容量符合要求的算力资源池, 并根据调度算法进行排序。

● 对于可拆分的需求, 先将算力资源池的剩余容量按从大到小的顺序进行排序。按照最小可拆分节点数, 先按平均分配的规则, 过滤出容量符合要求的算力资源池, 再根据调度算法进行排序; 如果都不满足, 则增加可拆分节点数, 再次按照平均分配的规则, 过滤出容量符合要求的算力资源池, 如果仍然不满足, 则再次增加可拆分节点数, 直到过滤出符合要求的算力资源池或是达到可拆分节点数上限。

(3) 在确定可用的算力资源池后, 如用户填写了具体的属性参数, 则按照用户填写的参数生成算力基本单元的属性配置; 如用户未填写, 则根据算力域封装的默认规则, 生成算力基本单元的属性配置。

(4) 算力调度的输入和输出:

输入条件: 算力域、算力基本单元、算力基本单元的属性配置、调度算法、用户相关信息、SLA。

输出结果: 算力基本单元、可选资源池(是否分布式、本次分配能力容量)、算力基本单元的属性配置。

1.3.3 智能选路

在算网雷达确定了唯一的组网方案后, 作为输入提交给算网大脑进行选路。算网大脑结合各网络域流量、性能、时延、带宽容量、无线接入、移动性等因素, 根据智能选路策略, 计算域间路径, 选择最优路径和网络切片方案接入云电脑。比如疫情期间针对政府、医院等重保单位, 可针对其高 SLA 要求匹配低时延高可靠网络切片, 全方位保障服务质量; 针对普通企业, 则可以结合性价比选择合适的网络切片服务。

目前已制定五种选路策略:

智能选路策略	筛选和排序值	择优标准
最低时延	算力节点和 UPF 之间最近 N 天的性能指标最差结果	时延最小为最优
最短路径	按跨域总节点数(暂且不考虑光传输层)进行排序。每个域的节点数优先调用相关的域控制器进行路由计算(云专网因为有全量拓扑可由算网大脑自己计算), 如不具备动态计算条件则在入云方案中配置静态值。	最短为最优
最低成本	按跨域 COST 值求和进行排序。跨域 COST 值初期在跨域连接管理的时候会进	成本最小为最优

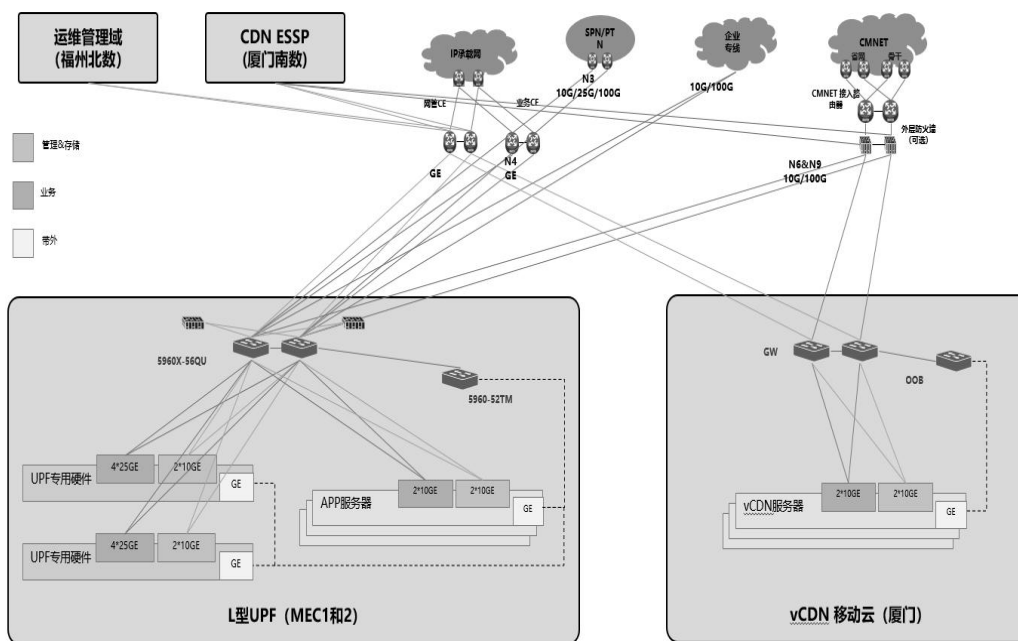
	行配置，后续可根据跨域链路的性能状态等进行动态计算。	
最高可靠	按方案的安全评估（静态）值进行排序。方案的网络整体可用率主要与接入域的连接方式有较大关系，如 PON 接入属于单 CE-单链路-单 PE 模式不具备保护，而 PTN 接入则可用分为双 CE/双上联/双 PE 多种保护。为简化操作，可以不同保护方案封装成不同的入云方案，并给每种方案赋予一个静态值。	可靠性最高为最优
综合推荐	在满足最低要求（可能是时延、带宽、可靠性等多方面）的基础上，对“入云方案和目标云池”组合，按以上几种策略设置不同的权重，进行组合加权评估，按得分高低进行排序推荐。	客户按照业务需求，按照权重选择最优

2.方案搭建与测试

2.1 平台搭建

为验证基于上述算网调度功能的新型云电脑方案部署可行性，我们在移动网络实验室搭建了包括算网编排大脑、边缘云管理系统、5GC、L 型 UPF、

2.6G 5G 基站、SPN、CPE 在内的测试环境，并进行了全局性环境预设：云电脑由 MEC1 和 MEC2 两个算力资源池提供服务，预制四个 5G 切片（MEC1 对应切片 1、2，MEC2 对应切片 3、4），由 5G 专网负责接入，对用户网络时延要求设置低于 50ms。



2.2 场景测试

测试主要包含配置态和运行态两大类，对于配置态，主要针对算力负载均衡、算力异常、算网协同编排三种场景；对于运行态，主要针对切片质量劣化、MEC 质量劣化两种场景。

2.2.1 算力负载均衡调度（配置态）

本场景测试目的是为了验证多台云电脑开通

环节，多个 MEC 资源池算力均衡分配能力。

首先，手动将 MEC1、MEC2 两个资源分配率调整为相近分配率，将网络时延设置为 20ms，模拟用户需求创建 4 台云电脑。

算网大脑对用户需求进行分析并拆解，下发工单给边缘云管理系统，启动算网资源分配；边缘云管理系统根据工单分配算(MEC1+MEC2)和网(切片1+切片4)的资源，部署云电脑虚机；云电脑管

理平台绑定用户实例至新建的 1-4 虚机。如图 2(a) 所示。

测试结果显示在边缘云管理系统可以查看到 MEC1、MEC2 各创建好的两台云电脑，通过云电脑管理平台可以查看到 MEC1、MEC2 绑定好的 4 台云电脑用户，用户可以通过云电脑 APP 访问对应 MEC 算力资源，正常上网和办公。

2.2.2 算力异常按需调度（配置态）

本场景测试目的是为了验证云电脑开通环节，当 MEC 资源池异常时算网大脑对算力的分配能力。

模拟网络环境，将网络时延设置为 40ms，手工将 MEC1 资源分配率调整为 50%，将 MEC2 资源分配率调整为 80%，按用户需求创建 2 台云电脑。

算网雷达根据云电脑用户开通需求，查看 MEC1、MEC2 算力使用情况，监测到 MEC2 资源使用率更高，因此选择在使用率较低的 MEC1 上创建新云电脑的方案，并反馈算网大脑，算网大脑下发工单给边缘云管理系统，在 MEC1 上进行云电脑开通。

测试结果显示通过云电脑管理平台可以查看到 MEC1 上新绑定的 2 台云电脑用户，用户可以通过云电脑 APP 访问 MEC1 算力资源，正常上网和办公。

2.2.3 算网协同编排调度（配置态）

本场景测试目的是为了验证云电脑开通环节，MEC+5G 专网的协同编排和资源调度能力。

手工将 MEC1、MEC2 资源调整为相近分配率，将 UPF1 到 GNodeB 的 5G 切片 1 的时延设置为 40ms，将 UPF2 到 GNodeB 的 5G 切片 2 的时延设置为 210ms，按用户需求创建 2 台云电脑。

算网雷达根据云电脑用户开通需求，查看 MEC1、MEC2 算力状态，监控到 MEC2 网络时延过大，不满足用户需求，因此选择在时延符合要求的 MEC1 上创建新云电脑的方案，并反馈算网大脑，算网大脑下发工单给边缘云管理系统，将 2 台云电脑的开通工单分配至 MEC1 中创建。

测试结果显示通过云电脑管理平台可以查看到 MEC1 新绑定的 2 台云电脑用户，用户可以通过

云电脑 APP 访问 MEC1 算力资源，正常上网和办公。

2.2.4 切片质量劣化网络动态调度（运行态）

本场景是为了验证云电脑正常运行时，出现运行云电脑的网络切片质量劣化，算网感知并做出网络切换调度的能力。

模拟网络环境，云电脑用户 1 已在正常使用切片 1，随后将 5G 切片 1 的时延调整为 210ms，5G 切片 2 的时延保持在 40ms。

算网大脑监测发现 5G 切片 1 的网络时延超过要求的低于 50ms 不达标且出现告警，MEC1 的用户 1 也发现无法正常登入到云桌面。算网大脑通过运维操作系统，查看并提供可切换网络，下发工单进行网络切换，并通知对接 5GC 的 UDM 和 PCF 修改用户关联切片关系，对接 5GC 的 PCF 强制用户 1 从切片 1 切换到切片 2 网络，MEC1 从切片 1 切换到切片 2，手工重启 CPE1。如图 2(b) 所示。

测试结果显示，云电脑用户 1 已通过切片 2 重新登入到 MEC1 的云电脑，恢复正常上网和办公。

2.2.5 MEC 质量劣化算力动态调度（运行态）

本场景是为了验证云电脑正常运行时，出现运行云电脑的 MEC 故障，算网感知并做出 MEC 切换调度的能力。

模拟网络环境，网络时延为 40ms，云电脑用户 3 已在 MEC2 正常使用，随后设置 MEC2 发生故障。

算网大脑监控到 MEC2 资源池告警，分析 MEC2 下的云电脑用户 3 业务存在风险；算网大脑通过运维操作系统，查看并提供可切换的 MEC 资源池，向用户 3 重新分配资源方案，下发工单通知 UDM 修改用户签约数据，在 MEC1 部署虚机 3' 并签约切片 2；边缘云管理系统根据工单在 MEC1 部署虚机 3'，5GC 更改用户 3 所在 CPE2 签约至切片 2，云电脑管理平台绑定用户 3 镜像实例至新创建的虚机 3'。如图 2(c) 所示。

测试结果显示，云电脑用户 3 通过切片 2 重新登入到 MEC1 的云电脑镜像实例，恢复正常上网和办公。

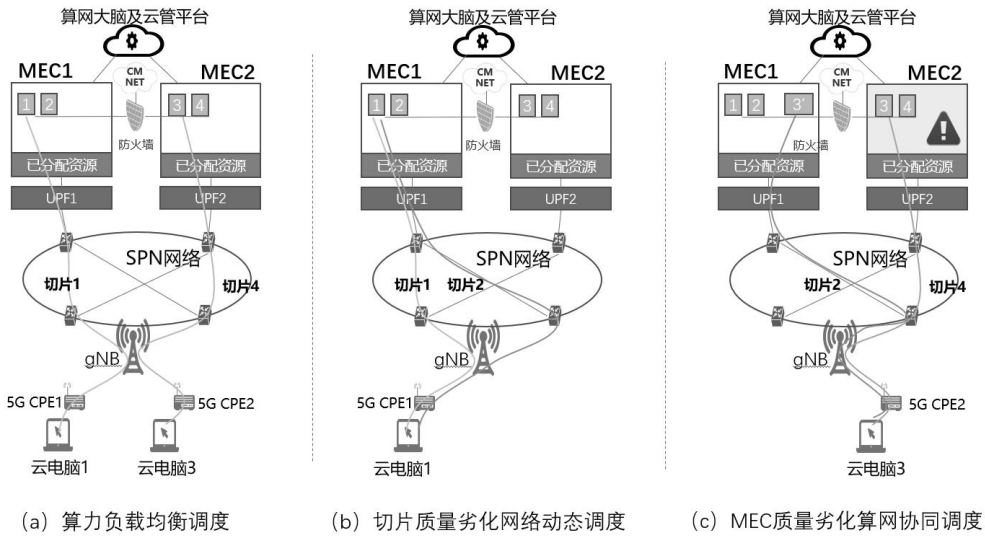


图 2：主要场景测试示意图

2.3 开通全流程测试

在上述五类场景测试正常的基础上，结合现网运营系统、编排系统、能力平台，借助 5G 专网业务开通系统架构，将云电脑应用场景自动映射为底层算力的集合，算网大脑负责底层算力资源的调度和编排，完成了基于算网调度功能的新型云电脑产

品开通全流程测试。如图 3 所示。

定制的云电脑开通任务能够正常呈现在业务订购页面，算力雷达能够根据输入的参数，正确输出备选方案，算力模板能正确解构为算网能力，系统根据基站位置选择最近的 UPF，按时延要求提供算网确定性服务保障。开通结果符合预期。

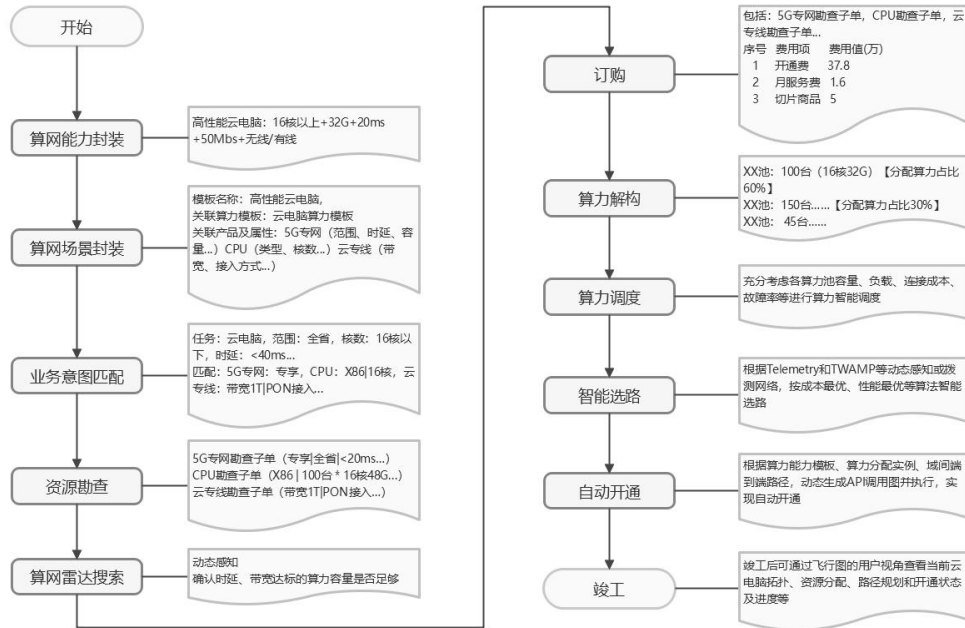


图 3：算网调度场景流程示例

3 结束语

从研究测试结果看，算网大脑能够对 5G MEC 和第三方算力实行智能调度和智能编排，基于算网调度的新型云电脑业务具备商用部署的可行性，能够满足企业和个人对移动办公、高效安全的新型云电脑需求，为电子商务、智能物联网等领域注入了新的活力和创意。目前已在某市开展试商用部署，开通数量超过 500 台，首批客户遍布教育、电商、法院、外呼、影视等行业，通过简易的管理和维护、便利高效的商业体验助力千行百业实现数智化转型。

对于算网调度的新型云电脑，还存在一些未完

善的功能，如异常场景算网调度业务切换无法自动无感实现，需人工进行网络切片调整创建，并重启 CPE；目前仅 5G 专网的接入方式，后期计划拓展有线、WIFI 等多种接入方式，对于这些需求还需要进一步研究完善。预期通过算力网络和云电脑结合的深入研究，能为人工智能产业的发展注入更加强大的动力。

参考文献

- [1]算网大脑白皮书，中国移动，2022.
- [2]中国移动 MEC 边缘云管理系统（MEO）接口规范，中国移动，2021.

数字孪生流域大模型的建设思路及场景应用实践

马森标 陈惠祥

(福建中锐网络股份有限公司)

摘要: 本文提出了数字孪生流域大模型分级架构, 各级互相解耦也可以互相关联, L0 级为基础大模型, 包含预测大模型和科学计算大模型, L1 为行业大模型, 包含气象大模型和水利大模型, L2 为场景大模型, 包含降水预报、洪水预报、内涝预报、水库抗暴雨能力计算、水库水位调度和土石坝渗压验证预警这些场景化大模型, 建议了大模型数据建设思路和基于政务外网的大模型分层算力网络架构, 提供了数字孪生流域大模型的典型应用场景及性能展示。

关键词: 数字孪生流域; 行业大模型; 分层算力; 典型应用场景

中图分类号: TP183 **文献标识码:** A **文章编号:**

引言

2022 年以来, 水利部颁发了数字孪生流域建设的一系列指导文件, 明确了推进数字孪生流域建设的具体时间部署、技术路线规划及相关负责机构的明确任务, 并提出了建设以 AI 智能算力为基础以满足核心业务模型训练和过程推理等需求的数字孪生流域平台。

福建省中小流域分布广泛、数量众多, 全省积雨面积 50 平方公里以上的河流有 763 条, 总河长 24629 公里^[1]; 分布有大中小型水库 3353 多座, 其中小型水库 3078 座, 大中型水库 275 座^[2]。深受洪涝灾害的影响, 福建省水利厅提出了增强水利智慧支撑能力的建设, 人工智能模型逐步得到了应用, 随着 OpenAI chatGPT 大模型的推出, 通用大模型遍地开花, 而随着华为盘古大模型的出现, 大模型在垂直行业的应用也开始崭露头角。

大模型、数据和算力互相促进, 相向发展, 大模型的场景化应用是焦点, 数据是基础, 算力协同是关键, 本文对数字孪生流域的大模型架构、数据建设、分层算力建设及场景化应用进行了综述。

1. 数字孪生流域大模型架构

水利部推进的数字孪生流域建设^[3], 其核心内容是数据、算法和算力的建设, 利用机器学习和深

度学习神经网络架构, 通过对水利行业的理解、模型优选及模型融合技术^[4], 及融合 AI 数据建模和 AI 方程求解的方法^[4], 使用神经网络编码微分方程^[4], 构建数字孪生流域 L0 级通用时间序列预测和科学计算基础大模型^{[5][6]}。

在此基础上, 通过汇聚传感器监测、地理空间、流域水系、跨行业、水利工程等各类多源异构多模态历史结构化数据, 通过调用 L0 级基础大模型, 构建 L1 级气象大模型和水利大模型, 和基于 L0、L1 级大模型构建短临降水、流域洪水、城镇内涝预报、水库抗暴雨能力计算及水库水位调度的不同场景大模型, 并将这些模型加载到边缘算力上进行洪涝灾害的预报预警。

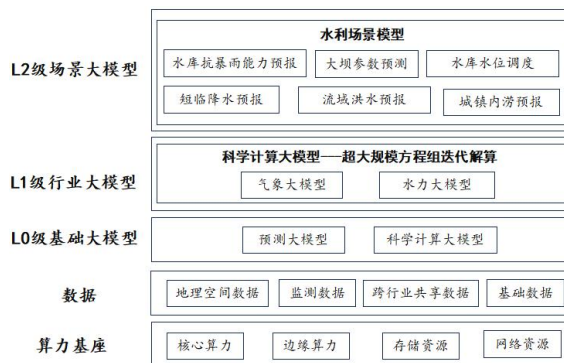


图 1 数字孪生流域大模型核心架构图

针对数字孪生流域大坝等水工建筑物场景，构建以 RBF、XGBOOST、SVM 等机器学习为主的 L0 级基模型空间，通过模型推荐、融合两步优化策略，构建 AI 预测大模型^[4]。

针对数字孪生降水、水库、河流、城镇内涝等场景，构建以 GAN、Transformer、Graphormer、GCN、LSTM、GRU、CNN 为主的 L0 级科学计算基础大模型，采用 AI 数据建模和 AI 方程求解的方法，使用神经网络编码微分方程^[7]，使用 AI 科学计算大模型更快更准的解决气象、水文和水动力计算问题^[4]。

省水利厅可以自己构建数字孪生流域 L0 级基础大模型，也可以跟大模型厂家如华为、阿里和科大讯飞等进行合作，构建有行业属性的数字孪生流域基础大模型，并进行模型的演进和迭代。

2. 数字孪生流域大模型的数据建设

数字流域大模型数据主要包括水位、雨量、渗压、渗流和位移等基础数据，及处理成的多时空尺度数据等，这些数据分布于各建设管理单位物联网数据库中，需对这些数据进行汇聚和质量管控，满

足大模型的高质量和大数据要求。

2.1 流域数据分层汇聚和质量管控

流域数据采用 3 层架构^[8]进行数据汇聚和质量管控，其体系架构如图 2 所示。

1)、在县级对传感器进行统一规划部署，满足降水与洪水之间关联度建模的需要^[9]，对传感数据进行清洗，及将数据处理成多尺度数据，满足多尺度模型训练的需要，并将这些分布于各建设管理单位的结果数据汇集并上传。

2)、在市数据汇聚中心，将县级数据进一步汇聚，并对汇聚上来的数据进行质量监控、审核和数据的应用。

3)、省数据汇聚中心，将市级数据进行汇聚，对数据质量进行考核和数据的应用。

2.2 打造全省统一数据底板

水利厅建设全省统一数据底板，政府和企业紧密合作，利用数据底板共同进行水利大模型的研发和应用，政府可以提供资金支持和政策指导，企业可以提供技术实践和市场应用，整合各方的力量，提高水利大模型的研究水平和应用能力。

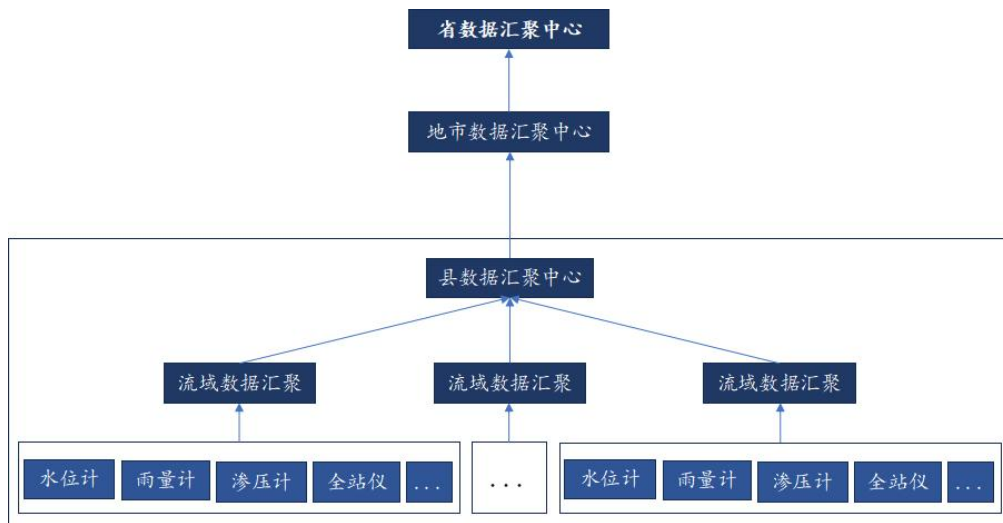


图 2 流域数据分层汇集

3. 数字孪生流域大模型分层算力建设

一般情况下，省级数字水利网络架构依托于电子政务外网，分省、市、县三级进行规划^[10]，在数字水利网络分级实施流域大模型，构建物理分布、

逻辑集中、协同工作的灵活算力，并统筹这些算力以满足数据治理、模型训练的需要，及在汛期或出现突发，本地算力不足的情况下，可以使用到上级单位的算力资源。

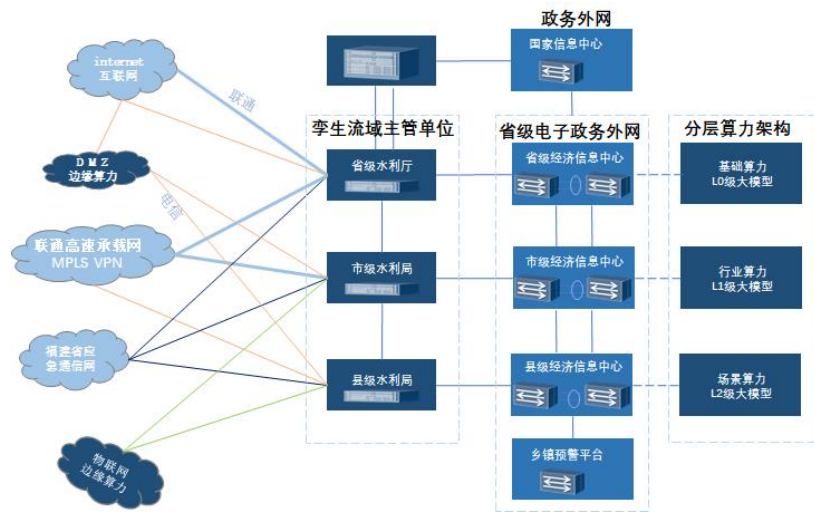


图 3 分层算力网络架构

假设目前有 10% 的流域、10% 的水库进行了数字孪生流域建设，每个流域部署了 200 个流量、水位和降雨量传感器，每个水库部署了 30 个位移、

渗压、渗流、水位和降雨量传感器，并且有了 1 年的数据，监测时间尺度为 1 小时，每个特征 16 字节，则现有数据量估算如表 1 所示：

表 1 全省 1 年数据量估算

	数量(流域、水库)	传感器数量	1 年的数据记录估算	1 年的数据量估算
流域	76	15200	133152000	1.98G
水库	335	9050	79278000	1.18G

根据华为 NLP 盘古- α 模型的经验^[1]，训练参数如表 2 所示，通过线性折算，训练 1G 的数据量训练参数如表 3 所示。

表 2 NLP 盘古- α 训练参数表(600G 数据量)

	模型参数量	数据量	时间(天)	算力(P/天)
盘古- α	2.6B	600G	3	110
盘古- α	13B	600G	7	110

表 3 NLP 盘古- α 训练参数表(折合为 1G 数据量)

	模型参数量	数据量	时间(天)	算力(T/天)
盘古- α	2.6B	1G	3	83
盘古- α	13B	1G	7	83

根据上面的评估，英伟达 RTX 3090 Ti 的 AI 单元算力 320T，可以比较好地适合数字孪生流域大模型的训练需求，可以考虑在省级构建以 RTX 3090 Ti 算力为核心的集群算力，在县市逐步降低算

力的建设，分层设计数字孪生流域算力网络，形成分布式及多样化的算力供给，构建云、边、端算力的分布式部署。

4. 数字孪生流域大模型的典型应用场景

数字孪生流域行业大模型由气象大模型和水利大模型(含水文水动力和大坝)2 个 L1 级的子大模型构成,场景大模型可由 L0 级也可由 L1 级大模型训练而成,由降水预报、洪水预报、内涝预报、水库抗暴雨能力计算、水库水位调度和土石坝渗压验证预警 6 个 L2 级的子大模型构成。

对大模型的训练和部署进行统筹规划,根据算

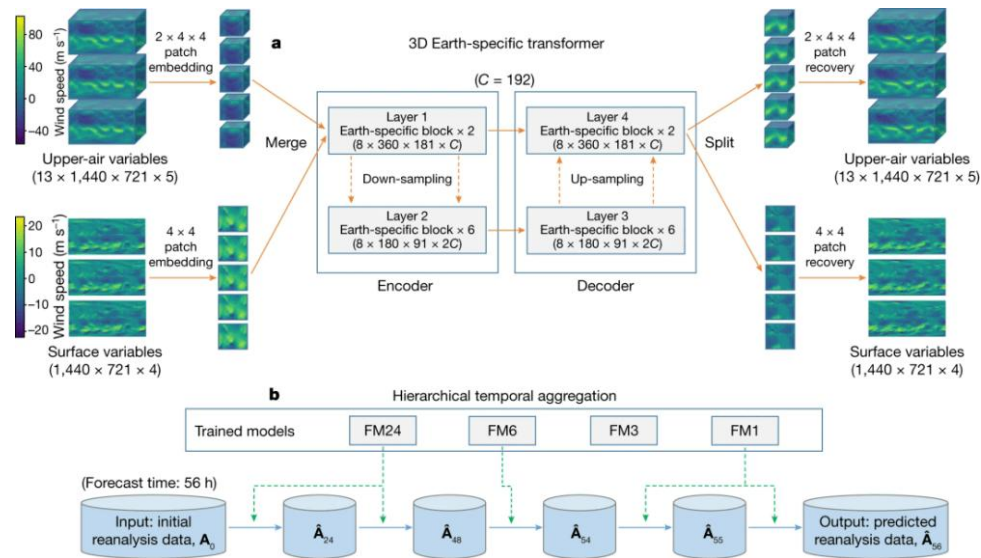


图 4 3DEST 网络训练和推理策略^[11]

华为盘古气象大模型利用 3DEST 三维神经网络进行气象数据处理,利用层次化时域聚合策略减少迭代误差,提高气象预报精度。

4.2 洪涝预报

全球变暖造成越来越多的极端天气,因此,城镇化率的升高为居民生活水平的提高提供了更多可能的同时,由于人口的集中,也给更多的人类带来了集中性的洪涝灾害,通过提高洪涝灾害的预报精度,增加预见期,可以让人们有更多逃生和减少财产损失的机会。

4.2.1 流域洪水预报

目前,一般采用物理模型、数据模型的方法进行流域洪水预报,由于物理模型需要的 DEM (Digital Elevation Model)、下垫面等数据难以获取、水文数据的非平稳性、模型参数的不确定性,数据模型显示出了巨大的优势,可通过流域的降雨量、温度、湿度、蒸散发和水位数据,利用科学计算大

力的分级部署和统一管理,进行算力的集中调度和优化,将部分大模型训练权限下放到地市,结合局地特征微调大模型,提升大模型的性能。

4.1 短临降水预报

通过华为盘古气象大模型,利用局地数据进行短临降水预报^[11]。

模型或者水文大模型,训练洪水预报模型。

通常,针对流域场景,可采用 LSTM(Long Short Time Memory, LSTM)、Transformer 提取流域水位与相关影响因素的时序依赖特征,融合流域的节点物理分布特征,设计图结构学习模块,捕捉流域水位与不同影响因素间的关联关系,利用 GCN(Graph Convolution Neural Network, GCN) 进行表征学习和预测,构建以图神经网络为核心的流域洪水预报深度学习模型。

本文采用罗源流域作为研究对象,该流域主要有护国溪、起步溪和南溪等 3 个主要的河流,在河流上设置了 55 个断面,通过新安江耦合一维水动力模型进行雨量-水位预报,现通过物理模型的预报数据,利用 LSTM-GCN 混合模型对物理模型进行仿真实验,各断面预报仿真结果如图 5-图 14 所示,仿真性能如表 4 所示,表明神经网络对物理模型的仿真性能良好。

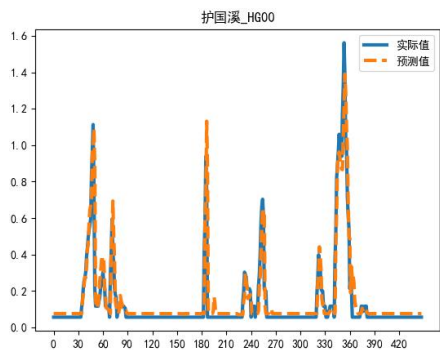


图 5 HG00 断面预报结果可视化(单位:m)

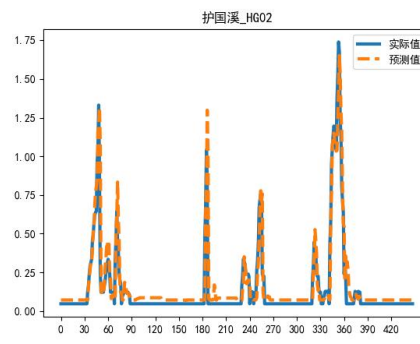


图 6 HG02 断面预报结果可视化(单位:m)

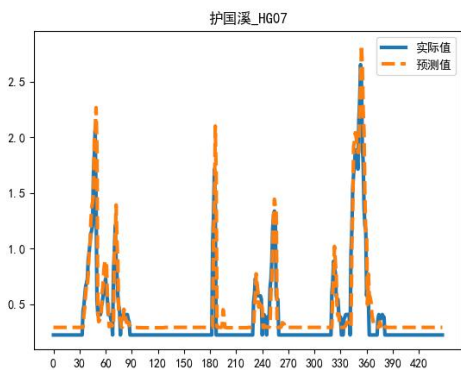


图 7 HG07 断面预报结果可视化(单位:m)

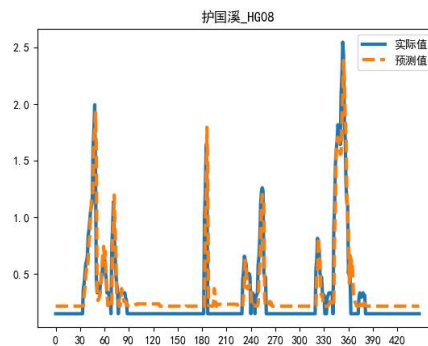


图 8 HG08 断面预报结果可视化(单位:m)

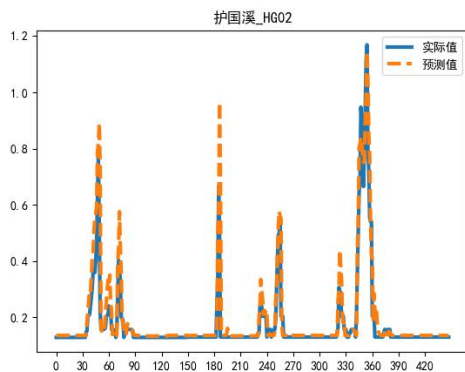


图 9 HG02 断面预报结果可视化(单位:m)

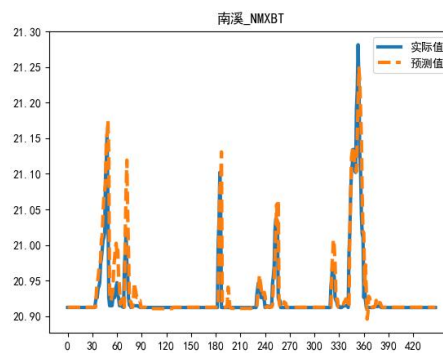


图 10 NXBT 断面预报结果可视化(单位:m)

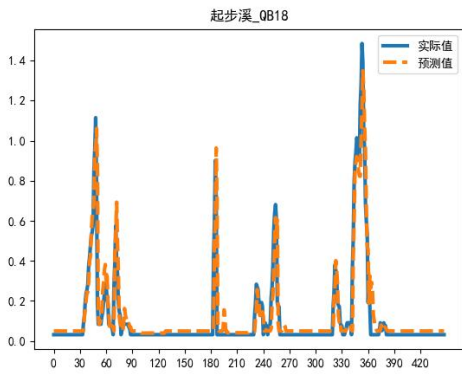


图 11 QB18 断面预报结果可视化(单位:m)

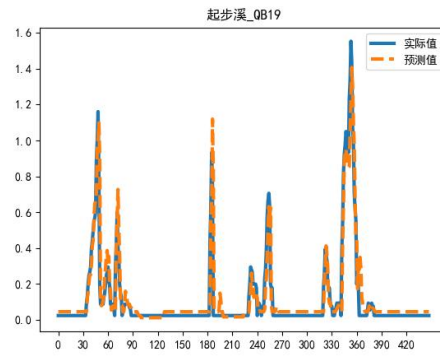


图 12 QB19 断面预报结果可视化(单位:m)

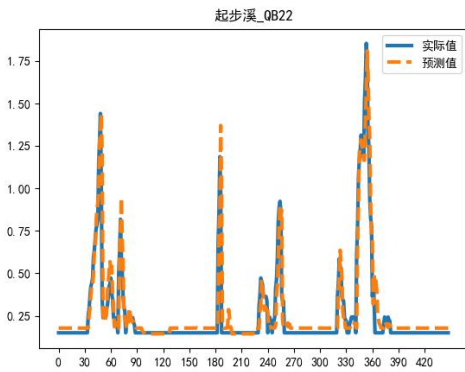


图 13 QB22 断面预报结果可视化(单位:m)

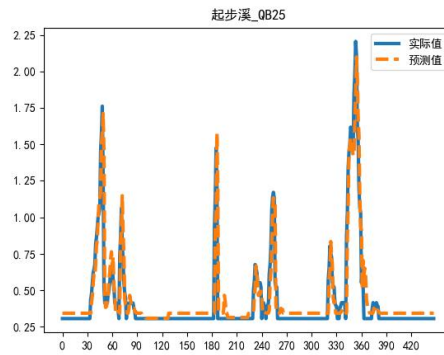


图 14 QB25 断面预报结果可视化(单位:m)

表 4 LSTM-GCN 模型评估性能表

	MSE	MAE	MAPE
指标数值	0.0196	0.0595	0.122

根据《水文情报预报规范(GB/T 22482-2008)》，预报的许可误差为 20%，预报项目精度等级表如表所示。

表 5 预报项目精度等级表

精度等级	甲	乙	丙
合格率/%	$QR \geq 85.0$	$85.0 > QR \geq 70.0$	$70.0 > QR \geq 60.0$

$$QR = \frac{n}{m} \times 100\% \quad \text{公式 1}$$

式中： QR 为合格率，%； n 为合格预报次数； m 为预报总次数。

根据模型的评估指标、断面预报结果可视化图及《水文情报预报规范(GB/T 22482-2008)》，罗源流域的洪水预报等级可以容易达到甲级。

4.2.2 城镇内涝预报

目前，国内外已经出现了大量的内涝模型，如 SWMM，Inforworks CS、SLAMM，STORM，HEC-RAS 等模型，这些模型已被广泛地应用于城市排水系统的规划及灾害的应急管理和预警工作，涉及到了错综复杂部署的管网环境，并且管网的堵塞疏通情况难以探明，提升了城镇内涝模型的建模难度。

相对于流域洪水预报，城镇的下垫面情况比较简单，内涝水深和降雨的关系比较简单明了，可以通过降雨量和水位数据，根据具体的内涝场景，利用机器学习或者深度学习如 ESN^[12](echo state

network)或者 GRU(Gated Recurrent Unit)建立城镇内涝水深预报模型。

本文采用罗源县城杭山村易涝点作为研究对象，通过二次同构模态对雨量进行分解，利用 EMD(Empirical Mode Decomposition)-GRU 模型对该易涝点进行建模预报，模型性能良好，该预报方法申报了相关的知识产权^[3]，杭山村易涝点如图 14 所示，该预报技术方案流程图如图 15 所示，该模型的测试集拟合图如图 16 所示，该模型的性能如表 6 所示，这表明，通过 EMD-GRU 模型，可以很好预测内涝点的积水水深，同时也说明了内涝水

深和降雨的关系比较简单明了。

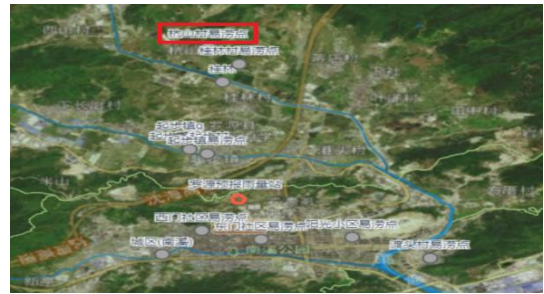


图 15 罗源县易涝点示意图

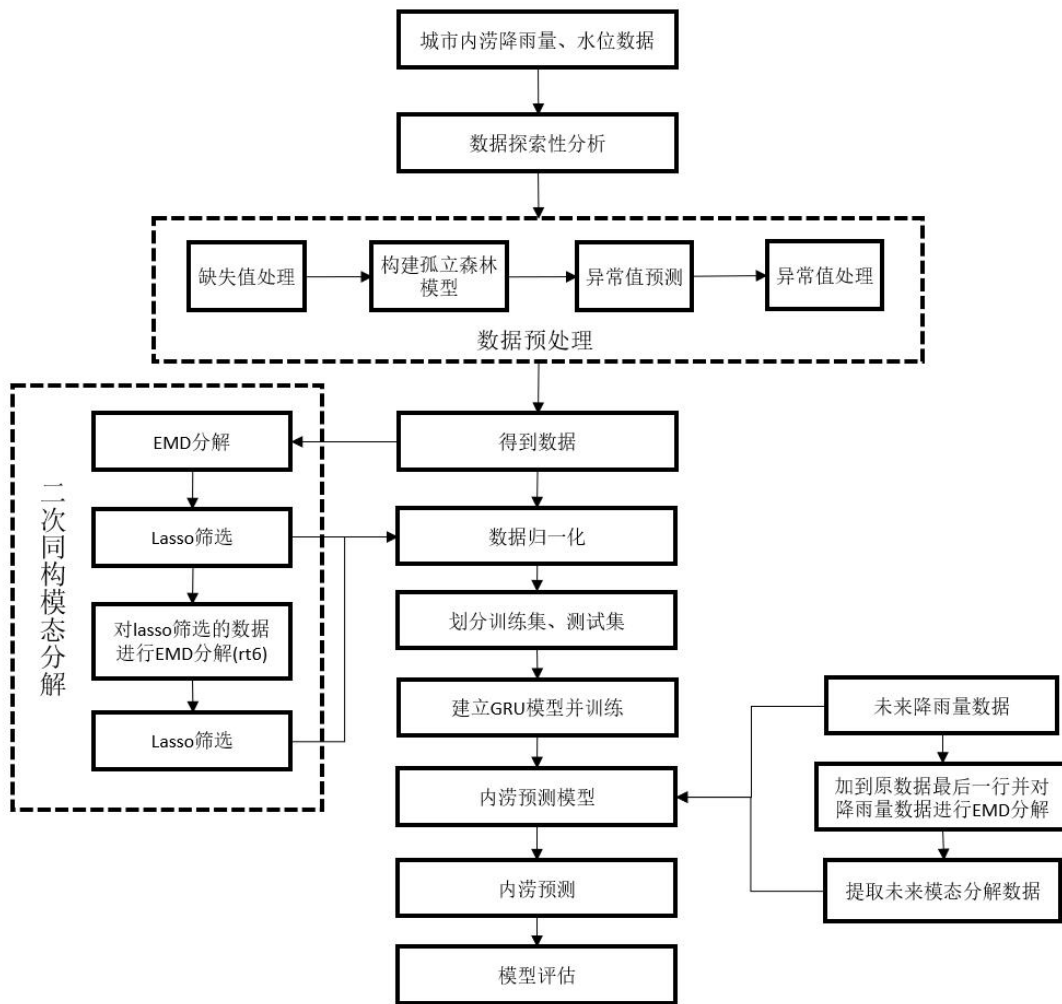


图 16 城镇内涝预报模型流程图

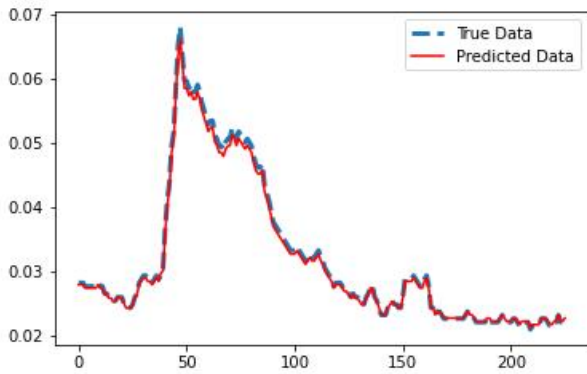


图 17 二次 EMD 特征分解 GRU 模型拟合图
(单位:m)

表 6 二次 EMD 特征分解 GRU 模型评估性能表

	MSE	MAE	MAPE	R ²
指标 数值	0.0000007	0.0004212	1.1075086	0.9974895

根据表 5 内涝点水深预测 GRU 模型评估指标表,参照《水文情报预报规范(GB/T 22482-2008)》,罗源杭山村易涝点的预报等级可以达到甲级。

4.3 水库抗暴雨能力计算

水库的抗暴雨能力,跟当前水库周边的下垫面情况、采用的调度方式和水库剩余库容密切相关。前期的降雨量影响了水库周边的下垫面情况如土壤饱和度,不同的下垫面情况对降水量转化为库容的效率有重大的影响,前期天气干旱,下垫面吸水性强,土壤可以吸收比较多的降雨,汇入库容的降

雨量比较少,水库的抗暴雨能力就比较强;水库的泄洪能力大,水库有比较强的调度能力,水库的抗暴雨能力就比较大;水库水位越低,则剩余库容越大,所能承纳的降雨量越大,抗暴雨能力就越强^[14]。

4.3.1 目前水库抗暴雨能力计算存在的几个问题

目前,水库抗暴雨能力的计算采用数据计算的方式,能以考虑到下垫面的非平稳特征、雨型对入库雨量的影响。

(1) 下垫面的非稳态特性

没有考虑到水库周边下垫面的非稳态特性,土壤的含水饱和特性随着流域的前期降雨是不断变化的,降雨径流系数也是起伏的,在计算抗暴雨能力时,随着选取降雨径流系数的不同,也会出现比较大的波动;

(2) 雨型的多样性

没有考虑到降雨的持续时间和降雨强度,在计算抗暴雨能力时,也会出现比较大的偏差。

(3) 物理模型的误差

各种物理模型都反映了水文学的原理,但人类对流域水文气象规律的认识,存在盲区,而流域环境又复杂多变,各种模型的建立依赖对真实水文现象的模拟概化,所以再好的水文预报模型也难以准确地反映物理流域世界,利用物理模型预报入库流量时总会存在一些差池。

4.3.2 基于深度学习的水库抗暴雨能力计算

利用历史气象数据、水文数据与调度数据,这些数据是各种各样的自然和人为因素作用以后产生的客观结果,利用深度学习,挖掘水库水位和这些数据之间的关系,可以反映出一些在物理模型中难以考虑到的因素,随着数据量的提升,深度学习模型可以比较精准挖掘到它们之间的映射关系。

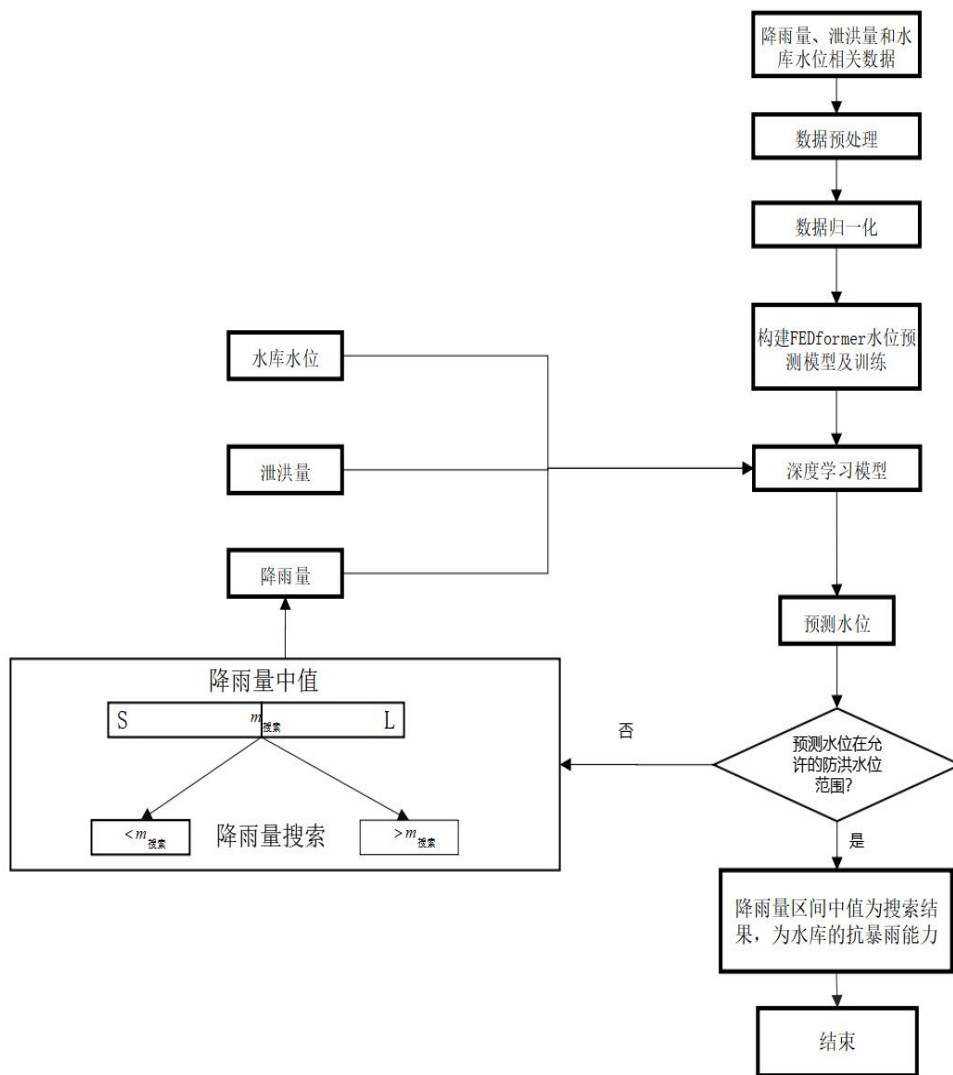


图 18 基于深度学习的水库抗暴雨能力计算流程图

基于精准的深度学习水库水位预测模型，采用二分搜索法进行雨量的搜索，可以计算出水库的抗暴雨能力，如图 17 所示。将水库相关的历史降雨量、泄洪量和水库水位数据，输入 FEDformer 深度学习算法，训练水库水位预测模型，根据设定的水库未来水位和泄洪量，通过二分搜索法，搜索达到水库目标水位的降雨量，这个降雨量，即为水库的抗暴雨能力。

水库抗暴雨能力的计算核心是水库水位

的预测，通过利用基于变分模态分解的 FEDformer 模型进行水库水位预测，得到了良好的预报效果，已经申报了相关的知识产权^[15]，该预报技术方案流程图如图 18 所示，该模型的测试集拟合图如图 19 所示，该模型的性能如表 7 所示，这表明，通过 FEDformer 预报模型，可以很好预测水库的水深，同时也说明了该模型可以很好适应水库周边的下垫面变化情况。

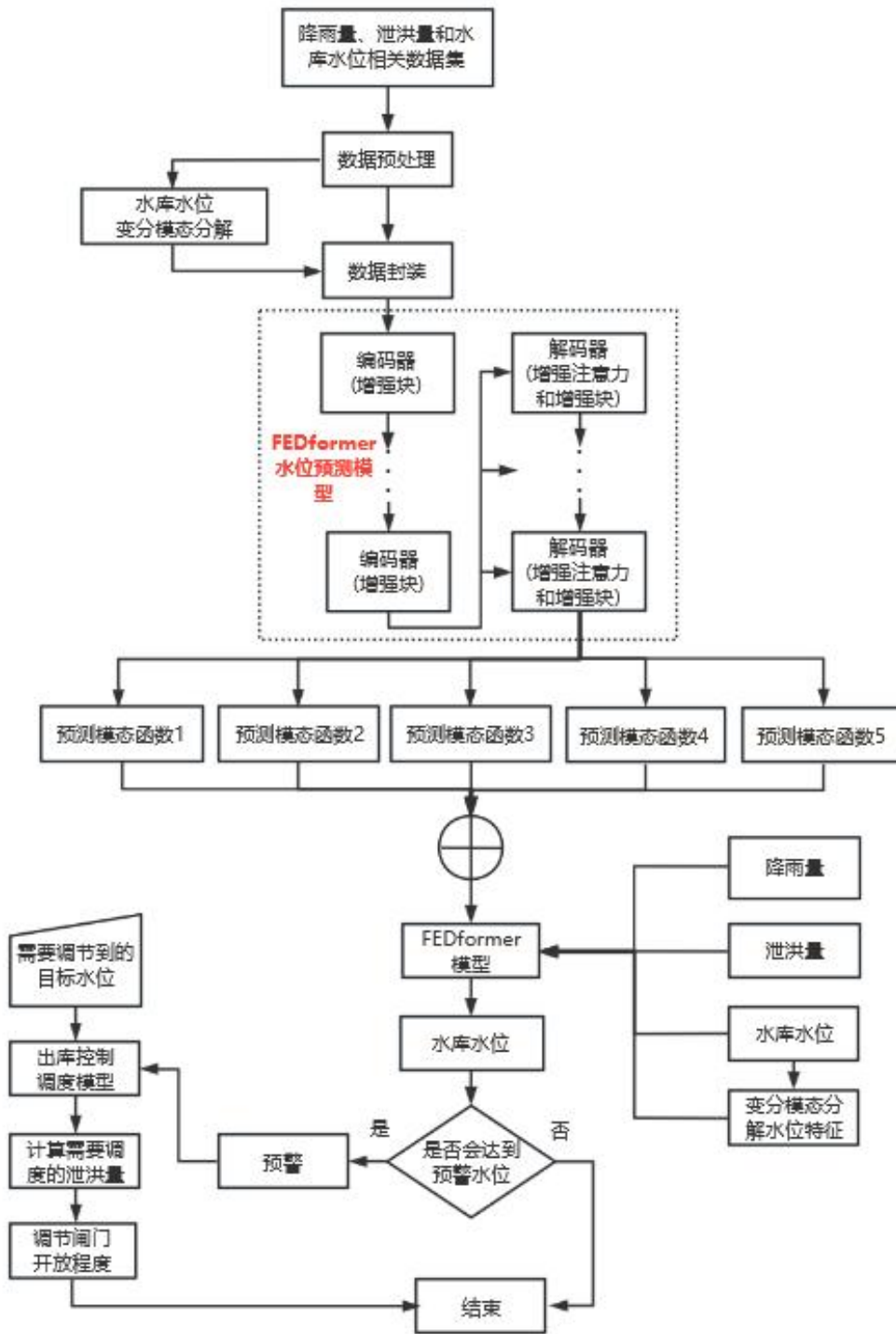


图 19 FEDformer 水库水深预报模型流程图

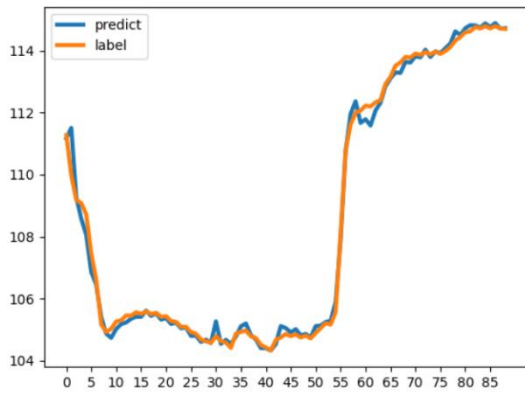


图 20 FEDformer 水库水位预报模型拟合图

表 7 FEDformer 模型评估性能表

评价项目	指标
MAE_1	0.1660
MSE_1	0.0696
MAPE_1	0.1526

根据表 6 水库水深预测 FEDformer 模型评估指标表,参照《水文情报预报规范(GB/T 22482-2008)》,水库水深的预报等级可以达到甲级。

4.4 水库水位调度

水库防洪调度需要考虑水库上下游的防洪要求,还要考虑水库作为防洪工程和兴利工程之间的

利弊平衡,是一个包含多个目标、且具有阶段性特点的一个复杂的决策过程,需要根据洪水情况的变化而变化,因而具有动态性,洪水的不确定性也决定了水库防洪调度工作的不确定性,通过多目标优化算法,生成多个水库调度规则,可以根据实际情况,选择一个合适的水库调度规则。

目前,水库调度技术的调度结果对下游的影响不可预见,存在防洪隐患,可建立关联下游水位的水库水位预测模型,采用 NSGA-II 等多目标优化算法,以水库水位、上游防洪、下游防洪等为多个优化目标,基于水位、下泄流量、下游水位、水量平衡等约束条件,构建水库水位调度模型。

本文通过溪源水库作为研究对象,利用泄洪量、雨量和上下游(在水库下方)水位数据,通过神经网络模型对上下游水位进行预报,取得了不错的效果,已经申报了相关的知识产权^[16],该预报技术方案整体执行过程图如图 20 所示,该模型模块流程图如图 21 所示,该模型的测试集拟合图如图 22-图 23 所示,该模型的性能如表 8 所示,这表明,通过 LSTM-GCN 预报模型,可以根据水库泄洪量很好预报水库上下游(在水库下方)的水位,如果融入水库的水位特征,则可以同时预报水库的水位,这样,在调度水库水位的同时,调度结果对下游水位的影响也是可以预见的。

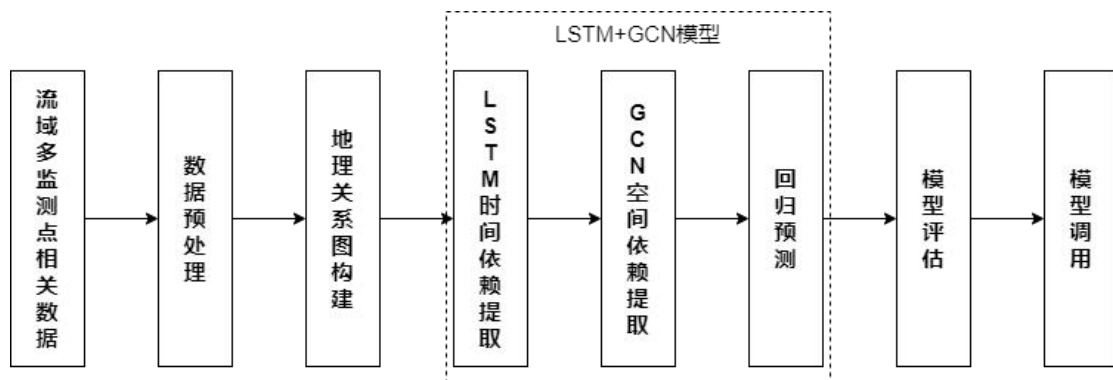


图 21 流域水位预报模型整体构建图

模型输入数据通过 LSTM 编码器、GCN 编码器两个主要模块提取特征,主要步骤通过以下流程实现,图 21 为模型模块流程图。

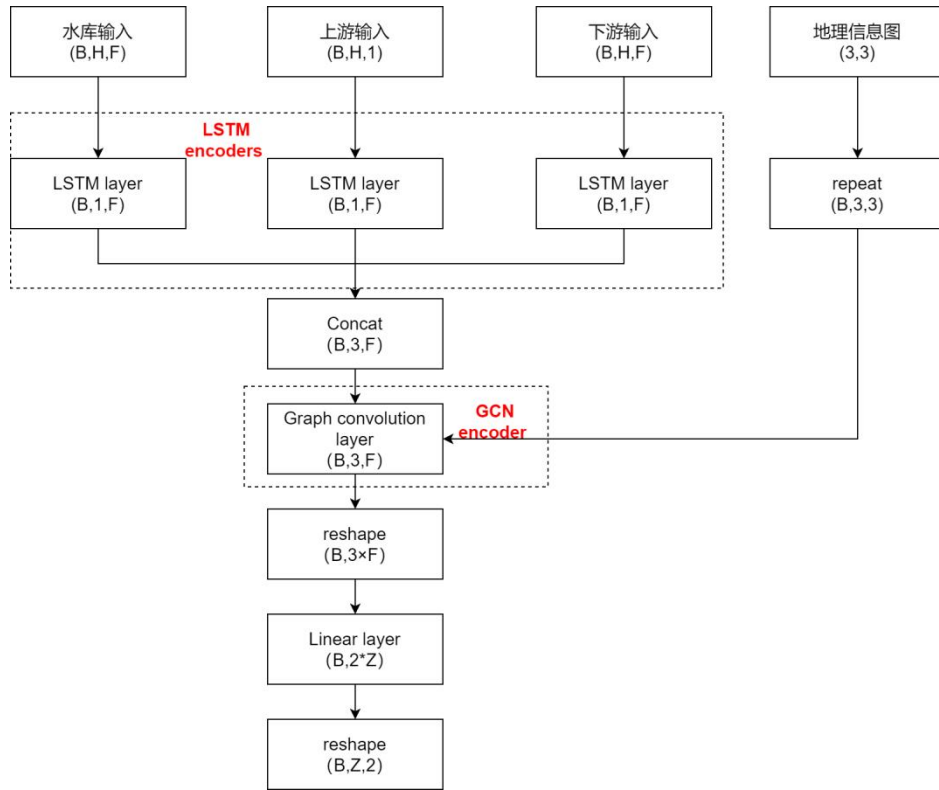


图 22 LSTM+GCN 模型核心流程图

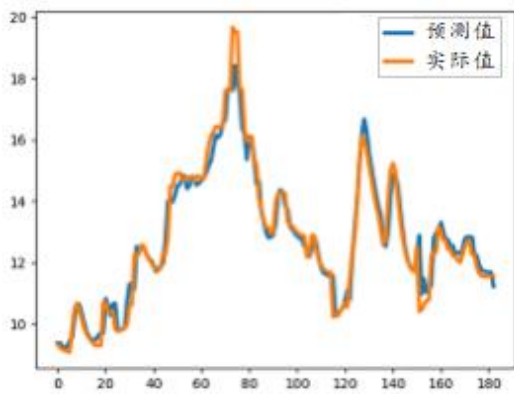


图 23 LSTM-GCN 上游预报结果可视化

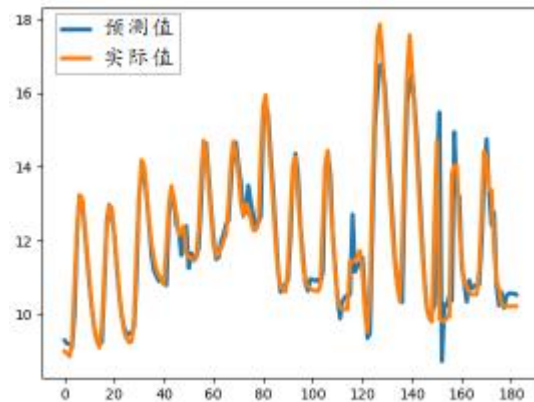


图 24 LSTM-GCN 下游预报结果可视化

表 8 LSTM+GCN 模型评估性能表

	MSE	MAE	MAPE
上游指标数值	0.3633	0.3658	0.4065
下游指标数值	0.9537	0.6249	0.7606

根据表 6 流域上下游预测 LSTM+GCN 模型评估指标表，参照《水文情报预报规范(GB/T

22482-2008)》，流域水位的预报等级可以达到甲级。

4.5 土石坝渗压验证预警

掌握土石坝体在运行期间的渗流压力分布情况和变化规律，是分析有无管涌、流土或接触冲刷等渗透变形或破坏的重要依据，判断防渗、排水、降压设施是否有效的重要手段，掌握土石坝渗压跟水位的变化规律对于保障水库工程的稳定和安全

运行具有重要意义，渗压是影响水库土石坝渗漏和工程稳定性的重要参数之一。

本文通过南平万寿寺水库作为研究对象，通过同一时间的水位数据，利用模型验证同一时间的渗压是否符合模型的预期，决策是否需要预警。利用

水库历史水位和渗压数据训练 RBF 神经网络模型，模型性能良好，已经申报了相关的知识产权^[17]，该预报技术方案整体执行过程图如图 25 所示，该模型的测试集拟合图如图 26-图 28 所示，该模型的性能如表 9 所示。

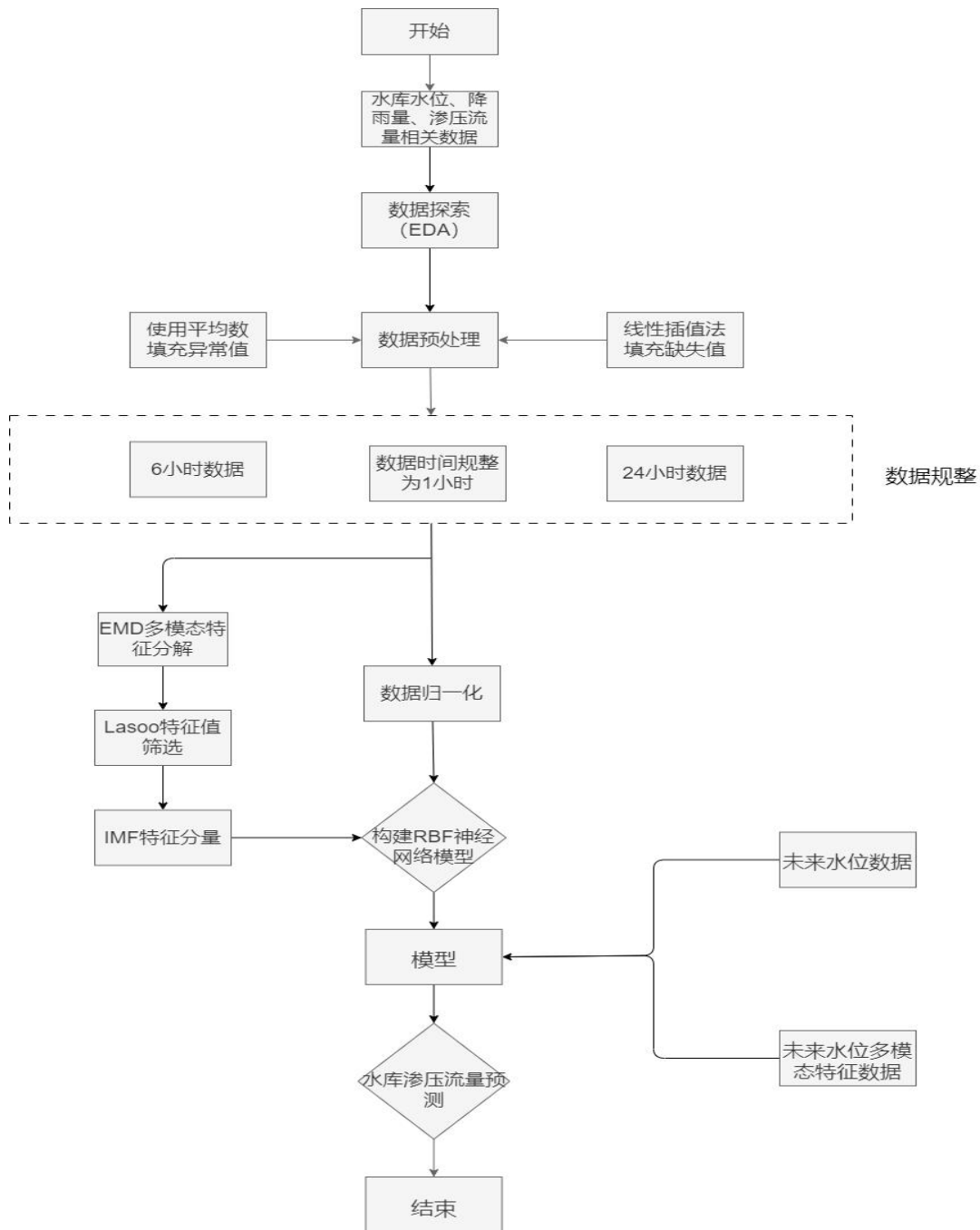


图 25 EMD+RBF 土石坝渗压预测模型流程图

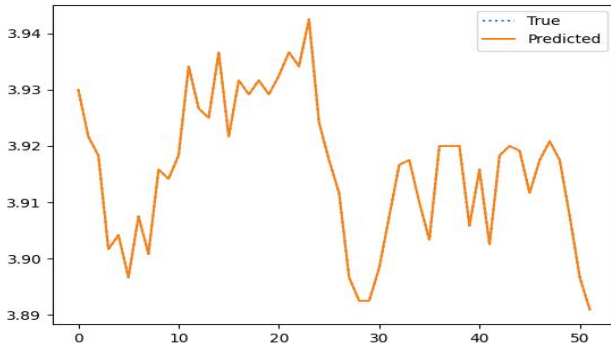


图 26 EMD 数据+RBF 模型拟合图 (1 小时)

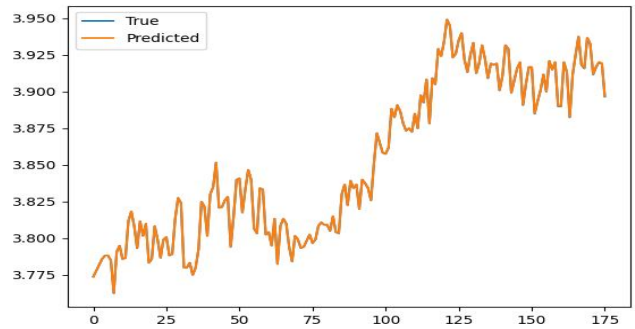


图 27 EMD 数据+RBF 模型拟合图 (6 小时)

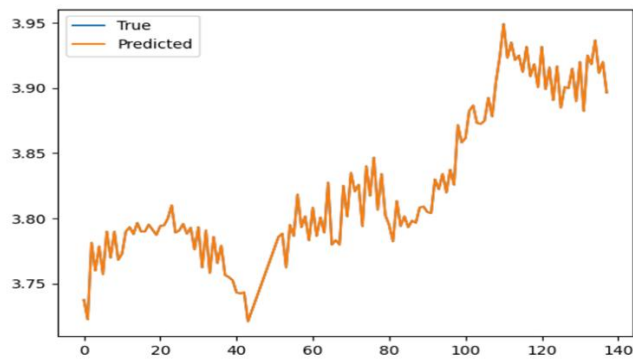


图 28 EMD+RBF 模型拟合图 (24 小时)

表 9 EMD+RBF 模型评估性能表

评价项目	原数据测试集	1 小时数据添加多 模态分解特征	6 小时数据添加多 模态分解特征	24 小时数据添加多 模态分解特征
MAE	0.57032	0.0012	0.01041	0.0018
MSE	1.921	0.0034	0.00126	0.0029
MAPE	14.812	0.0087	0.00267	0.0007
R^2	0.23	0.99	0.54891	0.99

明显地，根据模型拟合图和表 9 的 EMD+RBF 模型评估指标表，模型性能优异。

5. 罗源数字水利项目算力配置

在罗源数字水利项目中配置了 3 台硬件，1 台

边缘算力工作站，承载了流域水位预测模型，1 台数据库服务器，承载了流域的监测数据和预报数据，1 台应用服务器，承载了流域的“四预”平台。

表 10 罗源数字水利项目算力配置表

	型号	数量	用途	配置	承载应用
算力工 作站	Precision 3660	1	边缘算力	1 X i7-13700K CPU,64GB 内存, RTX3090 , 2 X 4TB SATA HDD	新安江模型、一二维水动力模 型、LSTM+_GCN 混合算法
存储服 务器	PowerEdge R750xs	2	流域数据 汇聚	4314 2.4G CPU, 2 X 16GB 内存, 2 X 1.2TB 10k 硬盘	数据存储和数字水利四预平台

存储服务器承载了流域的基础数据、监测数据和预报数据的汇聚和存储,对硬盘的性能有一定的要求,采用了 10k 转速的硬盘。

算力工作站承载流域水位的预报工作,对算力有一定的要求,该算力工作站采用 NVIDIA 的 GeForce RTX 3090 算力卡,该卡的浮点运算性能(FP32)约为 20 TFLOPS,整数运算性能(INT32)约为 80 TOPS,在此算力下,流域整体断面水位预测大概需要 25 秒,流域淹没计算大概需要 5 分钟,可以满足流域洪涝预报的要求。

6. 总结

随着大模型时代的到来,大模型也在向着垂直行业方向延伸,和朝着场景大模型的方向演进,对于数字孪生流域,可以衍生出气象大模型及水利大模型等行业大模型,及降水预报、洪水预报、内涝预报、水库抗暴雨能力计算、水库水位调度和土石坝渗压验证预警这些场景化大模型,随着数字孪生流域的大模型演进,对数据量和数据质量的要求不断提高,对算力的需求急剧增加,同时,数据量的

增加、数据质量和智能算力的提升,将反哺流域四预能力建设,提升流域的预报、预警、预演、预案能力。

参考文献

- [1]福建水资源简报(第八期)
- [2]朱珍香,杨军.福建水库空间分布特征:沿海密度高水量少、内陆密度低水量多.湖泊科学,2018,30(2).
- [3]李国英.建设数字孪生流域推动新阶段水利高质量发展.DOI:10.16616/j.cnki.10-1326/TV.2022.08.02.
- [4]华为.<https://pangu.huaweicloud.com/>.盘古大模型
- [5]Xinli Yu,Zheng Chen,Yuan Ling et al.temporal Data Meets LLM - Explainable Financial Time Series Forecasting. Conference acronym 'XX, June 03 - 05, 2018, Woodstock, NY.
- [6]时娜.华为云“All in ”大模型 盘古 3.0 登场加速百“模”大战.上海证券报,2023-07-10.
- [7]任清华.基于深度学习的偏微分方程求解方法(M).山东大学,2022年5月.

视频监控 AI 产品供给及运营

傅玮 林强 江雨

(中国电信股份有限公司福建分公司)

引言

未来,人工智能技术会渗入到生活,化为无形,融于万物。AI 的智能化、数据化变革的背后既是时代发展的推动,也是政策、用户、技术共同推动的结果,随着 GPU 等硬件设备的研发和人工智能相关算法的成熟,视频 AI 逐渐被大规模应用各个领域, AI+视频应运而生,其中安防行业一直被认为是 AI 技术落地最好的行业。

首先,安防是保障社会正常运行,人民安居乐业的关键,国家十三五规划、十九大报告等均提出加强视频监控智能化的目标,政府在智慧城市、雪亮工程等战略建设又为安防行业发展带来新契机。其次,随着城市化规模扩大,交通、楼宇、公安等场景需求广泛,拓宽了 AI 监控的覆盖面积。最后,大数据、云计算、生物识别、AI 等技术的发展与传统监控沉淀下来的大量图像、视频的融合,实现可视化、网络化、智能化为一体的监控架构将进一步打开和加速 AI 监控产品的扩大。

如果说移动互联网时代,每一部手机皆是智能交互节点,那么产业物联网时代,每一台摄像机就是端上最重要的一个节点。

1. 视频 AI 技术概述

1839 年 1 月法国摄影师达盖尔发明了照相机,自此人类学会了如何保存眼睛看到的图像,但这还不够,我们还想让机器自己去看,并且告诉我它看到了什么,这就是计算机视觉。

计算机视觉比较严谨的定义如下:

“对图像中的客观对象构建明确而有意义的描述” (Ballard & Brown, 1982)

“从一个或多个数字图像中计算三维世界的特性” (Trucco & Verri, 1998)

“基于感知图像做出对客观对象和场景有用

的决策” (Sackman & Shapiro, 2001)

在计算机视觉诞生之前,为创建一个程序来识别特殊的图像,需要耗费数小时的时间,来手动完成这类繁琐的工作。首先,需要核对一个相似图像的数据库。然后,不得不人工分析、测量这类图像,研究人员人工识别存疑的目标,标注相关的数据(如:颜色、测量值及形状)。而现在,计算机视觉采用深度学习这类机器学习方法,自动完成上述的所有流程。

视觉识别是计算机视觉的关键组成部分,神经网络和深度学习等技术的进展极大地推动了视觉识别系统的发展,帮助识别视频流中的事件、对视频进行分类、自动添加字幕以及使用超分辨率等技术增强视频和图像。目前视觉识别应用较为广泛的 AI 技术例如:包括图像分类、对象检测、目标跟踪、语义分割、实例分割等主要技术。

福建电信于 2019 年建设了智能视频云平台,通过自研视频 AI 算法,为行业客户打造智能化应用产品,实现良好商业模式的场景赛道。

2. AI 产品供给

(1) AI 能力建设

过去二十年里,传统安防厂商的工作主要围绕摄像头展开,在硬件设备、终端运维有较好口碑,但在云平台的底层搭建、海量数据处理、高性能的计算能力方面尚存不足。在 AI+产业链中,硬件设备制造、系统集成及运营服务是产业链的核心,渠道推广是产业链的经脉,通过 5G、物联网、大数据与人工智能技术提供智能安防云边端结合的整体解决方案成为视频监控发展的新趋势。

而发展智能视频业务是运营商的必经之路,这不仅体现在国家层面将 AI 纳入新基建的政策支持,且从需求侧和供给侧,智能产业生态正在不断丰富,聚焦价值领域,获得越来越多元的应用场景和更大

规模的受众。现在，不论是算法厂家、设备厂家还是系统集成商，都不具备全行业应用的开发能力。而运营商恰恰有面向全行业应用解决方案的能力：从 AI 技术到客户价值实现，需要一整套产品与解决方案，向客户最终交付的是应用，尤其是不同行业、不同场景、不同用途的业务应用都不尽相同，需要针对性的集成与开发。

福建电信智能视频云平台 2022 年以 AI+ 为基础，结合大数据对平台核心能力进行升级。至今，省内已建设了 28 台 GPU 算力资源池，选用英伟达 T4 和华为 Atlas300 智能推理芯片。通过容器管理平台在算力节点上按需创建容器 docker，将算法镜像包下发到容器上使用。按业务需求，省内逐步完善自研 AI 的品类，包括区域入侵（人、车、船）、明厨亮灶、垃圾屋监控、城市治理等 15 类算法。海量的 AI 事件经过视图大数据技术的加工后，生产出基于人脸、人体、车辆特征的时空轨迹等技战法应用，为行业客户定制智能化应用产品。因此，AI 产品化工作的核心是算法，本质是大数据。同时，集团也逐渐重视大数据 AI 的建设工作，并规划了

集团的 4 级算力架构。

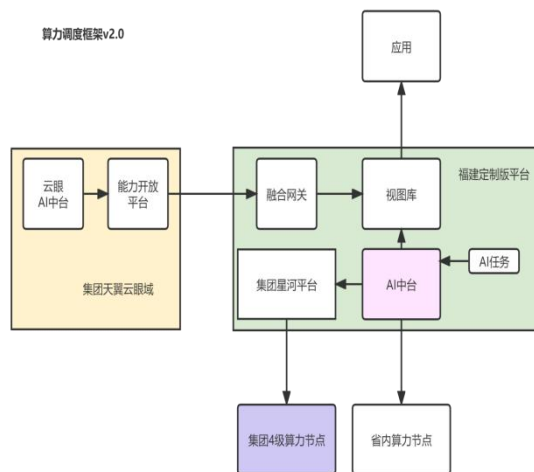


图 1 福建电信算力调度框架

(2) AI 中台架构

AI 产品的供给核心在 AI 中台，省内视频分析业务在控制单元的调度下，实现业务管理、算法开通、算法任务监测等功能，如下图：

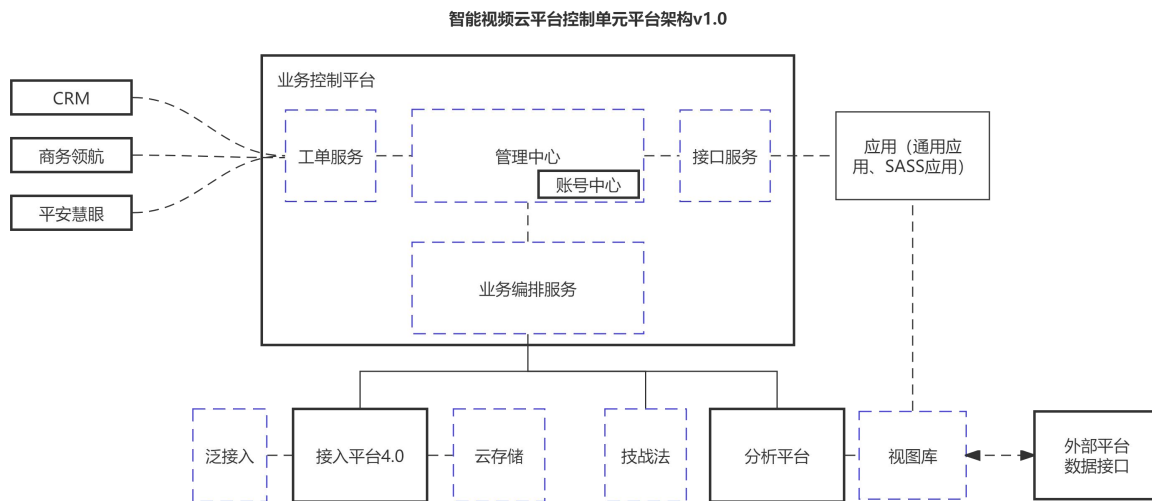


图 2 视频云平台架构

其中，AI 中台围绕算法引入标准化、算力管理集约化、AI 赋能便捷化三大核心功能，打造 AI 生态圈，为行业客户提供安全可信的智能视频数字化

使能服务。提供运营型平台+项目型平台两种产品架构，支持轻量化部署，满足高可用特性，保障业务稳定性。其功能架构如下：

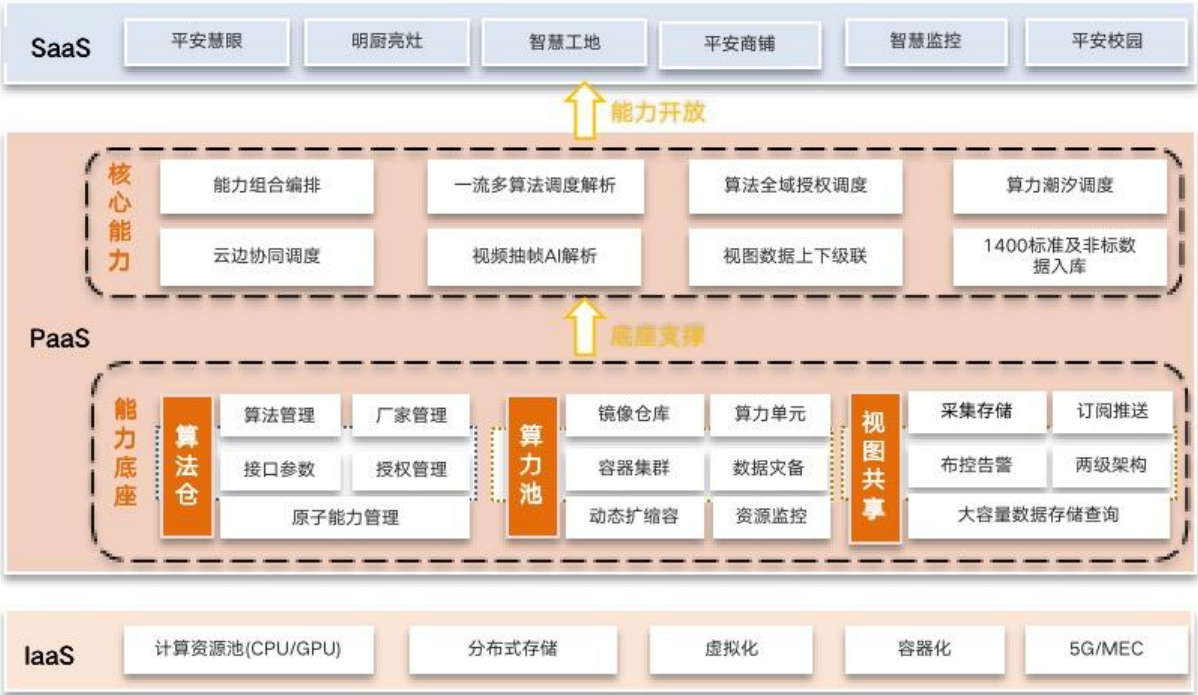


图 3 AI 中台功能架构

AI 中台具备跨域授权、潮汐调度、算力池化、多算法编排等核心能力，可实现：与行业主流算法厂商开展实质性合作，实现算法跨域授权、全网调度功能；具备多厂家多算法动态授权能力，支持 License 申请、下发、回收等功能；实现同一算力资源在不同时段支撑不同算法的分析需求；采用算法任务+场景策略因子+多算法能力组合方式，充分利用计算资源，实现 AI 解析智能化、算力集约化，降低视频业务的分析成本；适配多种 GPU 型号，采用容器化技术，结合基础配置的资源模版，快速生成 AI 算力单元，提供 AI 分析能力；根据业务需求，具备高可用及集群负载能力，保障业务的快速扩容及稳定性；多算法厂商多 AI 能力的自定义编排功能，实现同一视频源不同 AI 目标检测及解析，输出丰富的结构化数据。

(3) AI 产品打造

传统算法是：输入数据和规则，产生结果；
 人工智能是：输入数据和结果，产生规则。
 面向智慧农业、智能制造、智能交通、智慧物

流、数字金融、数字商贸、数字社会、数字政府、其他数字化效率提升业等领域的机关、企事业单位，在发展过程中存在痛点、难点、堵点问题，都需要优秀数字技术、产品和解决方案。因此福建电信 AI 产品的思路，不是单纯输出结果，而是要基于客户的行业属性，运用这个结果打造应用。

具体做法是：将 AI 原子能力注册到控制单元的能力服务，然后通过控制单元的业务编排服务，实现小 B 产品按需定制。比如为独居老人做一个产品，老人步履蹒跚，最怕跌倒，在业务编排中 AI 能力的倒地识别加入原子能力；老人体弱多病，可能猝死，在业务编排中 AI 能力的久居未出算法加入原子能力；老人反应迟钝，遇事不决，在业务编排中将音频能力的一键告警能力加入原子能力。在通用应用底座上，开通客户的账号，并将一键告警、倒地识别、久居未出三个能力勾选出来，即完成了应用产品的定制界面。

福建电信已完成明厨亮灶、城中村、城管三大应用标品的打造与推广，具体如下：



图 4 厦门智慧城管



图 5 厦门智慧城中村



图 6 明厨亮灶福建版

3.AI 中台在业内先进性

三大标品均已获得客户的高度认同并带动了视联网定制版业务的发展。AI 中台充分利用企业现有的能力，具备**系统稳定可靠、经济原则、标准性、共建共享原则、易操作原则**等先进的特性，当前端有新的需求出现时，后端可以用自动化的工具迅速迭代实现功能。

在短短一年中，AI 中台已实现 100+算法的部署和使用，有力的支撑了省内智能视频业务的发展。在技术上也紧随业内头部企业的步伐，实现以下几个创新能力：

(1) 告警证据链：支持将同一摄像头下的所有告警信息进行事件归档，通过目标特征和告警坐标等多维度进行研判，生成多条告警证据链，避免同一告警事件多次重复告警，提高数据价值，并为事件追踪回溯提供有力支持。

(2) 高效抽帧：通过对视频流的高效抽帧技术（CPU+GPU 抽帧方式的结合），调用自研的人像模型，结合人脸特征判别规则，将检测框标在图片中，再拼合图片成视频流，通过流媒体服务器输出。

(3) 算力分级调度：支持根据任务的不同重要等级，对算力资源进行不同通道划分，针对一类任务，建立算力高架专道，保证任务高效运行；针对二类任务，建立算力共享资源池，充分利用资源的同时保障业务正常运行；针对三类任务，数据允许 T+1 告警。

福建电信 AI 中台运营至今，已在 100 多个项目，加载 5000 多路算法，其中 TOP5 的 AI 算法分别是：人脸识别、车辆识别、区域入侵、明厨亮灶、垃圾分类。在厦门同安一网统管项目中，福建电信 AI 中台总评分排名第二，仅次于百度。

4.AI 运营能力

得益于视频爆炸时代下，指数级增长的海量视频数据处理需求，以及重点场景对于技术精度的需求，更精准的处理需求，都推动了视频 AI+业务的发展。业务需求往往只关注“算法有没有”，而运营的主要责任在“算法好不好”。在 AI 需求汹涌而来的同时，也反映出 AI+的运营体系的短板，如 AI 业务的建章立制还不够完善，AI 运维队伍还未建立等。因此，视联网团队解决方案和测试交付两条支撑线，向上承接集团，向下训战队伍，同时加强技术自研，提升核心竞争力。AI 运营核心能力如下：

(1) 云边协同-能力编排：减少资源消耗、打造极致成本

云边协同：边端采用低廉的国产 CPU 对视频流进行抽帧及目标检测，云端采用高端 GPU 对边缘上传的图片流进行识别解析，降低边到云的带宽消耗、提升云侧的解析效率；

能力编排：拉取一路视频流，通过视频抽帧能力+多图片能力的编排，实现一个设备赋能多种 AI 能力，降低流媒体资源消耗、利用图片 AI 代替视频 AI，降低算法软件成本。

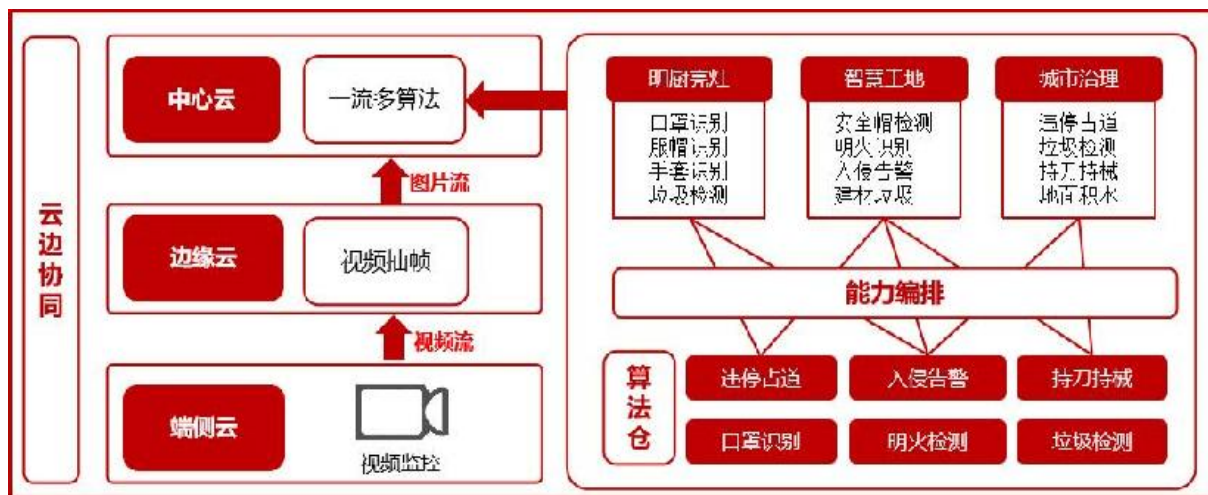


图 7 能力编排

(2) 算力池化-跨域授权：算力集约管理，授权全网调度

算力集约化：通过算力容器化技术，实现对算力资源的集约化管理，将原本独立的服务器进行资源池化，适配多种 GPU 型号，结合资源配置模版，快速生成 AI 算力单元，提供 AI 分析能力；

算力动态扩容：在资源池的基础上，实现算力集群负载，根据业务需求，具备高可用及集群负载能力，保障业务的快速扩容及稳定性；

跨域授权：与行业主流算法厂商开展实质性合作，基于 2+31+X 部署架构，实现算法跨域授权、全网调度功能；



图 8 算力池化

(3) AI 生产中心：一站式 AI 生产管理业务闭环

从场景化角度对算法进行全方面的测评，为项目客户筛选可靠、性价比高的算法；集成标准数据

集管理、数据标注、数据回流等功能，为算法测评及模型训练提供标准数据来源；针对项目标注数据进行自动化算法模型训练，优化特定场景下的算法准确率。

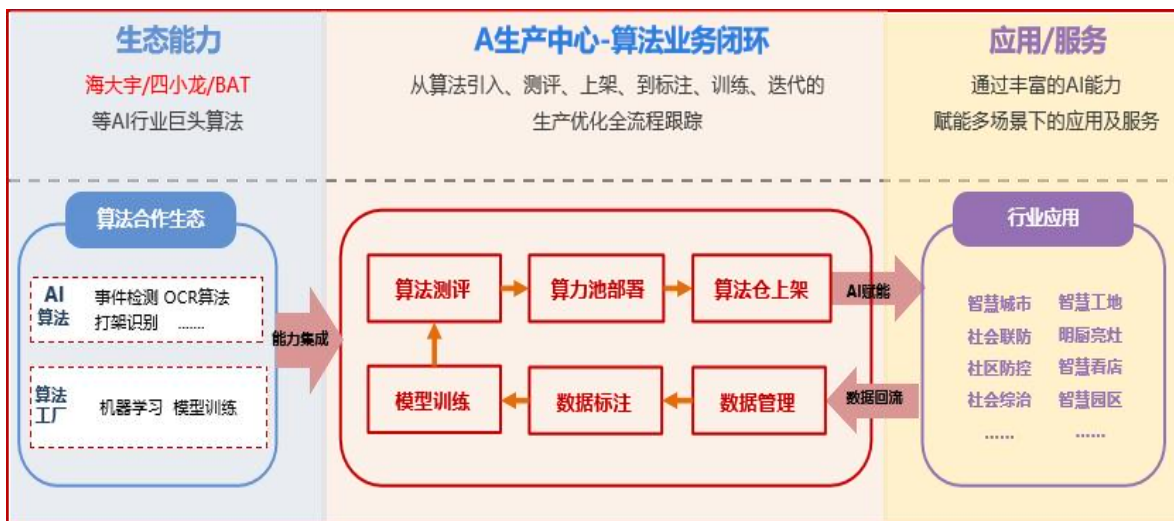


图 9 AI 生产中心

(4) 完善算法优化机制，满足解析实战要求当前行业算法“出厂精度”在落地部署后，受

现场环境、标注定义、视频流质量等客观因素，无法达到实验室精度，所以，AI 算法基于真实场景数

据做增量训练，提升算法精度以满足实战要求。福建电信建立了一套算法精度测算的工作流程。

首先确定了算法成熟度标准，在项目中可交付

为 L4，准确率 95%以上，召回率 90%以上。

其次，建立了算法测算方法：

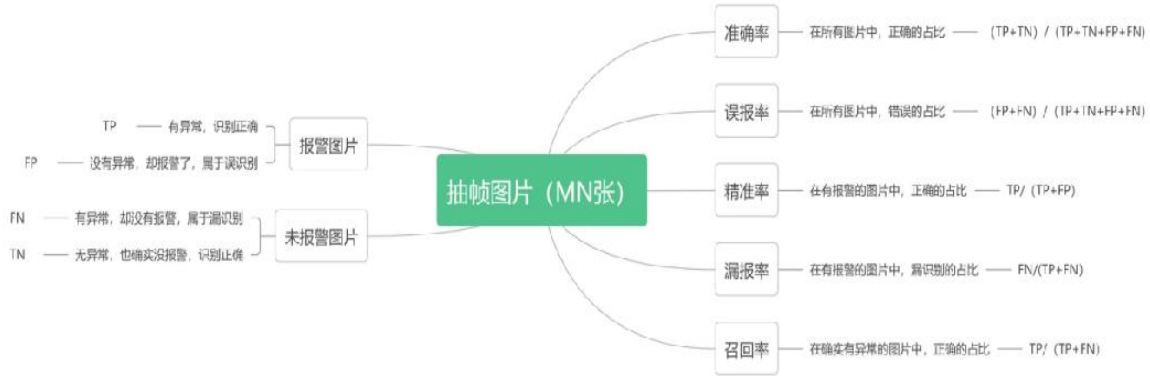


图 10 算法测算方法

最后，每月都安排项目中算法的测算，并定期发布。

算法名	测试团队	算子	场景	选点位置	置信度	测试日期	TP	FP	FN	TN	精准率 = TP/(TP+FP)	召回率 = TP/(TP+FN)	准确率 = (TP+TN)/(TP+TN+FP+FN)
明厨亮灶 (销售品增强)	视频部	地面垃圾识别 (集团AI)	室内 (纸张纸团、包装盒 (如饮料盒、鞋盒等)、纸板纸箱、垃圾袋)	省公司食堂后厨	0.65	20230113	217	4	284	496	98.19%	43.31%	71.23%
			湖里万达商厦的味道监控-夜晚	湖里万达商厦的味道门口	0.65	20230125	120	5	124	1	99%	96%	98%
			湖里万达商厦汕头牛肉粿条监控-白天	湖里万达商厦汕头牛肉粿条门口	0.65	20230125	121	4	125	0	100%	97%	98%
			浙美村8号垃圾分类投放点监控-夜晚	小区垃圾投放处	0.65	20230125	109	141	245	5	96%	44%	71%
			湖里万达商厦的味道监控-白天	湖里万达商厦的味道门口	0.65	20230125	120	5	125	0	100%	96%	98%
			湖里万达商厦鲜美多营养餐监控-白天	湖里万达商厦鲜美多营养餐门口	0.65	20230125	0	125	123	2	0%	0%	49%
			湖里万达写字楼海希房产-白天	湖里万达写字楼海希房产门口	0.65	20230125	0	125	123	2	0%	0%	49%
	视频部	口罩识别 (集团AI)	口罩 (一次性医用口罩、N95口罩、红色口罩、粉色口罩、黑色口罩)	省公司食堂后厨	0.65	20221111	8	0	15	0	100.00%	34.78%	34.78%
	视频部	未戴口罩	省公司食堂后厨	0.65	20221111	33	0	0	0	100.00%	100.00%	100.00%	
	视频部	厨师帽识别 (集团AI)	厨师帽 (白色、一次性白色/蓝色)	省公司食堂后厨	0.65	20221111	26	0	0	7	100.00%	100.00%	100.00%
	视频部	未戴厨师帽	省公司食堂后厨	0.65	20221111	8	0	0	59	100.00%	100.00%	100.00%	
	视频部	厨师服识别 (集团AI)	厨师服 (白色)	省公司食堂后厨	0.65								
	视频部	手套识别 (集团AI)	手套 (胶皮手套、医用手套)	省公司食堂后厨	0.65	20230325	80	0	0	80	100.00%	100.00%	100.00%
	视频部	垃圾桶未盖检测 (集团AI)	室内厨房-标准绿色垃圾桶	省公司食堂后厨	0.65	20230325	81	0	39	120	100.00%	67.50%	83.75%
			室内厨房-标准小垃圾桶绿色	省公司食堂后厨	0.65	20230325	84	0	36	120	100.00%	70.00%	85.00%
			室内厨房-铁桶、油漆桶	厦门六中食堂	0.65	20230325	0	0	140	140	0.00%	0.00%	50.00%
			室外白天垃圾点-垃圾站垃圾桶	小区垃圾投放处	0.65	20230325	85	1	65	149	98.84%	56.67%	78.00%
			室外白天垃圾点-小区楼下垃圾桶	小区垃圾投放处	0.65	20230325	16	0	114	130	100.00%	12.31%	56.15%
	视频部	抽纸识别 (集团AI)	室内厨房	厨房	0.65	20220824	173	12	20	3	93.51%	89.64%	84.62%

图 11 算法测算结果

4. 算法交付任重道远

AI 算法生产出来后，并不能满足所有场景的使用，准确率、召回率都比较低，这和 AI 训练所用的样本数量不足、工程化体系不完善等有很大关

系，具体如下：

(1) 训练点位零散接入，无规模化训练，无法覆盖多场景

目前算法团队能用于训练的场景一般是项目中提供的少量路数点位训练出来的算法，在当下点

位可以做到非常高的准确率，但后续推广接入的点位仍然是几路几路地接进来，无法泛化批量训练，因此在新接入的点位未必都能够出成效，新的项目都可能存在高错误率、漏识别率的问题。

(2) 缺乏算法效果自评估工具，以致问题难发现

在有些场景中，识别到人没戴口罩，要提醒客户，比如公共交通或影剧院，但有的场景，要客户摘下口罩，比如通过人脸识别进行考勤，所以，戴或者不戴口罩，都可能是正确的结果，我们无法通过短时间是否有 AI 事件来评估算法任务是否正常，但可以通过一段时间的跟踪来发现视频流是否稳定、算法任务是否正常等问题。此外，我们已知摄像头安装对算法效果有很大的影响，但我们没有手段，通过 2D 的视频图片反推摄像头安装的高度、俯视角、指北角、亮度等参数与所开算法之间的匹配问题。如果只能依赖于装维人员个人的经验与理解，很容易造成 AI 装维需要多次上门调测。

(3) 算法上架后应用感知差

算法产品上架后需要应用加载配置以及应用逻辑处理，均需要应用运营支撑团队人工处理，比如区域入侵需要绘制区域识别框，客流统计需要配置人数阈值才能生成告警、告警的时长需要配置、应用要结合客户需求过滤算法置信度、目标分析框显示等；如若没有应用运营团队的高效配合与支持，必然给客户对算法的使用感知造成影响。

(4) 客户期望值没有管理，以致交付验收难
客户期望的识别要准确，识别不能漏，对应到

算法指标分别是准确率和召回率。而准确率与召回率是一对矛盾的指标，追求准确率可能要牺牲召回率，反之亦然。在有些场景，准确率更加重要，如人脸比对，公安技术的要求是每百亿次比对只能一次出错。有的场景，不能漏更重要，如区域入侵，在客户认为极其重要的区域，无论白天晚上，晴天雨天，客户都不希望忽略了闯入的每一个人。因此，要管理好客户的预期值，算法自研的交付人员在项目初期，需求沟通时，就要介入项目，确定指标标准。

(5) 项目运营环节无闭环，反馈机制不健全

在项目中产生的误告警缺乏自动化反馈流程，仍需客户或运营人员搜集反馈，因此算法在项目上的精准优化无法快速启动负样本的自动化收集。而运营环节与算法研发团队之间没有闭环，无法将算法应用的场景分门别类后将项目信息同步给研发团队，并按需回流告警图片，这就造成研发团队无法第一时间掌握使用情况，并做出快速的优化迭代响应。

5. 总结与展望

自 2022 年 OpenAI 的 ChatGPT 横空出世，大模型多模态的 AI 技术火爆出圈，集团大数据 AI 中心快速跟进，打造大数据 PAAS 底座，推出 AI 赋能的星河平台和面向政法公安的天元平台，整个战略布局雏形已现。省内 AI 发展战略也将更深度的融入到集团的视联网发展规划中，在控制单元上完成能力与业务分拆，快速定制出千人千面的 AI 应用，为主航道提供轻交付的 AI 产品能力。

智能算力赋能数字政府管控探索： 基于数字孪生的视角

朱惠斌

(中国移动通信集团广东有限公司深圳分公司)

摘要：数字政府管控因涉及数据部门广、数据类型多、数据总量大，面对智能算力的供给和分配等难点，已成为信息通信行业的研究热点。在“探索建设数字孪生城市”成为国家“十四五”规划的契机下，数字孪生技术成为智能算力赋能数字政府管控的重要策略。从发展背景和应用前景 2 个方面分析数字孪生技术的发展现状，从建筑信息系统模型、地理信息系统模型、三维场景系统模型、物联网系统模型和人工智能系统模型 5 个方面探索数字孪生技术的实践模型与主要平台，从实践制约和发展展望 2 个方面分析数字孪生技术的未来发展潜力。研究结果表明在智能算力赋能下，数字孪生技术应用于数字政府管控具备可系统仿真模拟、智能算力融合多层面数据要素的优点，需从平台开放标准化、多接口接入统一和设施的共建共享 3 个方面进行革新和优化以降低总体技术应用成本，为智能算力赋能数字政府管控提供重要的基础性支撑。

关键词：智能算力；数字孪生；数字政府

中图分类号：TN91；TU984 **文章识别码：**A **文章编号：**

引言

算力是数字政府管控效率的重要支撑要素，其具备处理数字政府海量数据规模和类型的重要作用。国家“十四五”规划提出“探索建设数字孪生城市”的发展目标，将数字孪生技术作为数字政府规划、建设和运营的全生命周期发展技术，为智能算力赋能数字政府管控提供重要基础，为未来城市高质量发展指明道路。数字孪生技术属于系统模拟仿真技术的优化升级，其通过对现实情况进行数字化建模，以供数字政府的数据分析、趋势推演、生产调度等实际发展需求，是当前社会经济数字化转型的重要途径，为提升城市基础算力、融合多时空、多维度数据、形成城市虚拟管控空间具有重要的基础性作用。数字孪生技术兼顾本地化部署和云端部署能力，可根据实际发展需求进行大网和专网间的灵活调度调整，具备业务部署灵活性，其技术在部署数字政府过程中具备显著的先进性。

数字孪生技术在经济层面的重要作用在于推动数字经济发展和传统产业数字化，为数字城市总体效率提升提供支撑作用。国外数字孪生技术的主要应用前景为城市规划、建设和运营，其改变了原有城市规划、城市建设和城市运营相互关联度有待加强的特点，实现城市全生命周期管理。在城市全生命周期管理的基础上，数字孪生技术逐步向工业等领域渗透，为生产效率提升带来立竿见影的效果。我国目前已从国家战略等层面充分重视数字孪生技术的应用，并予以政策和配套支持，为未来数字孪生技术更好地服务我国高质量城市发展提供支撑。通过分析数字孪生技术的发展背景与应用前景，探索数字孪生技术的实践模型与主要平台，从数字孪生技术面临的实践制约入手，探讨数字孪生技术的发展展望，有助于充分实现国内数字政府建设与国外实践经验的接轨，响应国家战略，从数字经济和传统经济数字化两个层面入手，从技术层面切实推动城市新型经济发展增量和城市生产生活效率

提升,对数字政府管控具有紧迫性和必要性。

1. 智能算力赋能数字政府管控的发展背景与应用前景

1.1 发展背景

智能算力赋能数字政府管控的发展背景主要包括两个方面:首先,部分城市发展模式探索的实际经济和时间成本过大,导致系统模拟仿真等技术逐步出现。但系统模拟仿真技术采取特定条件假设,其实际模拟结果只能部分拟合现实发展情况。因此,数字孪生技术通过系统构建与现实完全一致的模型进行仿真模拟,是对原有系统模拟仿真技术的重要发展创新,大幅度降低实际应用成本。其次,数字孪生技术有助于融合多目标、多专业、多类型的发展目标,实施融合,改变原有系统模拟仿真技术针对一个局部类型的局限性,提升总体城市系统效率。在降低成本和扩大范围两个因素的共同影响下,数字孪生技术在海外数字政府管理领域得到迅速发展,且已逐步拓展至其他领域。

1.2 应用前景

目前,数字孪生技术已初步接入的数字政府典型应用类型包括智慧交通(含机场、高速公路、高铁等)、智慧园区,智慧基础设施(电力、供水等)、智慧文旅等。这部分典型应用得以实践的原因是多样化的。首先,智慧交通、智慧园区、智慧基础设施和智慧文旅均有明确的运营主体,在多年发展进程中已逐步形成数字化发展基础。以智慧交通为例,机场、高速公路、高铁目前已具备科学的调度系统,可实现实时监控、调度等职能,以保障具体应用的实施。数字孪生技术是在一定的数字化发展基础上,根据新时期的需求进行优化和提升,其实施难度低、实施成本低,且对相应运营主体而言是提升运营效率和收入利润的途径,运营主体具备充分的主观意愿以推动数字孪生技术在相应领域的实施。

数字孪生技术正在尝试进入的数字政府典型应用类型则包括智慧工厂、智慧物流仓储、智慧港口等。这部分典型应用相对运营稳定性较强,虽然也具备一定的数字化基础,但主要集中于生产范畴,缺乏从监控、调度等领域的全链条数字化基础,需要耗费大量成本对现有网络架构、调度系统和设备

内容进行大幅度优化更新。若实施数字孪生技术,典型应用类型的生产效率将得到大幅度提升,属于传统行业数字化的进程,但其前期投入的硬件、软件层面成本也极大,需作为资本开支摊入产业总体折旧成本中。相应运营主体则需考虑其带来的效率提升是否能覆盖其建设成本,在长周期内实现总体效率提升,运营主体的风险偏好和创新偏好则显著影响数字孪生技术的进一步发展和使用。

数字孪生技术未来潜在数字政府发展领域包括智慧乡村、智慧城市等。因智慧乡村和智慧城市与前两种典型应用类型不同,其属于系统工程,需要平台构筑和基层工程的同步配合。系统工程的实施包括规划、建设等多个层面,单一层面实施效果也会影响总体实施效果。且系统工程构筑的筹资是阶段性的,如果分阶段实施过程中的总体规划建设方针进行调整,则影响系统工程的后续实践效率。这类典型应用类型是属于政府主导的,兼具一定的社会福利和基础设施属性,是未来数字孪生技术提升社会生产生活效率的重要目标。

2. 智能算力赋能数字政府管控的实践模型与主要平台

2.1 建筑信息系统模型

建筑信息系统模型(Building Information Model)是建筑智能化发展的重要模型支撑,其实现对原有建筑设计通过二维模型对平面、立面和剖面进行分类设计的特点,将建筑作为三维立体空间进行建模,有效减少原有建筑设计容易出现的管线位置冲突、不同二维图纸相互不对应的特点,提高建筑总体可视化效果的提升。国内部分城市正对新建建筑推动建筑信息系统模型的使用和存档,确保新建建筑信息的数字化留存,为未来城市数字化管理提供助力。

因建筑是城市的重要载体,其具备高度、承载人数等参数基础,具备结构、给排水、暖通、电气等专业基础,其具备大量设计层面信息。建筑承载的人员、生产生活活动是动态调整的,与城市同样存在设计、建设、运营等全流程,其信息保存和处理是城市数字孪生系统的重要基础性要素,对数字政府管控具有重要的作用。目前,建筑信息系统模型主要为静态模型,即以设计指导施工,施工过程中的设计变更等未必会体现在建筑信息模型中,后

期的建筑运行过程中产生的变化则更难体现在建筑信息系统模型中。且不同使用主体对建筑信息系统模型的数据类型和颗粒度定义不同,难以实现不同建筑间的对接。因此,建筑信息系统模型需统一建筑信息类型和颗粒度,将其应用和数据更新拓展至建筑全生命周期,且确定全周期的统筹模型运营主体,以确定建筑信息的真实性、准确性和实时性,为数字政府管控奠定数据基础。

2.2 地理信息系统模型

地理信息系统模型 (Geographic Information System) 是城市智能化管理的重要模型支撑,其使用普遍结合地理信息系统、遥感数据等进行统筹数字政府管控。地理信息系统承载区域、城市等地域的各类基础属性,涵盖内容极多。但地理信息系统模型作为静态模型,缺乏动态数据更新基础,主要应用于现状分析和历史演变研究。且地理信息系统模型主要来源于政府统计数据 and 遥感数据等内容,缺乏与建筑信息系统模型的直接对接;其重视物质空间数据,缺乏与其他类型模型对接使用等特点均影响其应用于实践中。因此,对地理信息系统模型进行优化和调整,有助于提升地理信息系统模型的实际作用,成为数字政府管控的重要组成部分。

2.3 三维场景系统模型

三维场景系统模型是在建筑信息系统模型和地理信息系统模型基础上,结合现场多视角照片、多专业图纸和无人机高空拍摄的模式,形成的三维仿真场景。三维场景系统模型是重要的可视化场景平台,可根据场景进行各类操作(含视角放大缩小、平移旋转等)以查看现场实际情况和细节。三维场景系统模型软件类型较为繁杂,具备多样化的渲染引擎和建模方式。三维场景系统模型是数字孪生技术中最需标准化和规范化的模型,将建筑信息模型和地理信息模型进行可视化转化,并增加模型系统可操作空间。三维场景系统模型具备的数据属性映射是根据数字孪生技术需达到目标而确定的,可调整相应显示分辨率和操作颗粒度,以适配各种需求类型下的实际情况。虽然建筑信息模型可间接反映三维情况,但三维场景系统模型增加场地等数据要素,提升数字孪生技术的界面友好性,为数字政府管控提供可视化基础。

2.3 物联网络系统模型

物联网络系统模型 (Internet of Things) 指实现“万物互联”的发展愿景,其核心是通过多种传感器和网络基础设施以实现信息的快速处理、传输和存储。根据实际发展需求,需对传感器进行分级分类,在保障城市发展目标的同时控制总体资本开支和设备成本。传统物联网络系统包括摄像头等设备,但目前已逐步转型为视、声、味等多种类型,并逐步复杂化。因此,物联网络系统模型间的兼容性和数据接口统一性是制约模型能否得到实施的重要因素。因物联网络系统的供应商和服务商众多,各专网间难以直接并存,且供应商和服务商也不愿意实现底层架构开放,以达到共建共享和便捷接入的目标。因此,政府主导统筹,各类供应商和服务商相配合的模式,是实现物联网络系统模型的重要路径,也是实现数字政府管控的重要路径。

2.4 人工智能系统模型

人工智能系统模型 (Artificial Intelligence) 包括资源管理、算法模式和安全预警等层面。资源管理指数字孪生技术包含的数据类型和数量极大,需通过人工智能系统实现资源高效统筹,增加频繁使用数据的调用效率,降低稀少使用数据的占用空间;算法模型则以提高效率为基础,即增加数据处理的速度和效率;安全预警则是对系统可能的潜在危险进行趋势外推和提前预判,并进行处理。人工智能系统是一项系统性工程,其面对的重要挑战在于模型目标的多元化,即通过目标的重要程度确定算法中的分项权重。人工智能系统模型是数字孪生技术能否得到充分应用的重要组成部分,直接决定数字孪生技术演算结果的准确性和导向性,对数字政府管控效率具有重要的影响作用。

3. 智能算力赋能数字政府管控的实践制约与发展展望

3.1 实践制约

(1) 成本约束

互联网和移动互联网规划建设过程中,科学进步带来信息通信效率大幅度提升。其中,新型信息通信技术 (5G、全光网等无线和有线传输技术) 革新对全社会信息交互效率提高的支撑作用极其显

著。因信息通信系统包括数据处理、存储、传输等多个层面，若其中单一层面出现显著短板，会影响总体系统效率。新型信息通信技术研发需耗费大量成本，但其具备显著的特点是开发完成后，可根据具体需求修改参数，并广泛应用于各地域，属于前期研发大规模资本开支，后期可大规模应用的技术场景，其研发成本在后期可扩展地域的应用过程中通过大规模应用产生的利用摊薄研发单价，因此技术革新具备可持续性。但技术革新的应用具备基础设施层面和扩展应用层面。其中，基础设施层面可通过扩大用户数量和提高用户收入单价，以简单地提升总体营业收入。扩展应用层面则需通过具体应用需求，进行再次技术研发并实施，不仅带来新增的研发成本，且还需评估新增营业收入能否覆盖新增研发成本、技术研发场景是否具备可拓展性和广阔的市场前景，影响数字政府实施的技术基础。

因数字孪生技术的应用单位、应用需求呈现多样化的特点，其决定了数字孪生技术无法通过传统的前期大规模研发资本开始、后期大规模低成本应用的模式回收利润，其应用过程中则需平衡研发投入和研发成本间的关系。因数字孪生技术采用的主要为主流模型，各服务商均可调度，不具备显著的技术壁垒，单一服务商在开发固定类型数字孪生技术应用后需与其他服务商进行竞争。因此，数字孪生技术服务提供商普遍采取根据需求特定开发的模式，即特定需求产生的营业收入能覆盖研发成本，服务提供商才会进行开发。因此，数字孪生技术多应用于特定化场景，特定市场推广规模有限，在服务成本大规模降低前，难以实现广泛应用。因此，采取私人定制化方案并不利于数字孪生技术应用，推出特定套餐模式以适当覆盖数字政府同质化业务需求达到服务供给标准化则有利于数字孪生技术发展。

（2）数据基础

数字孪生技术受到的另一个重要实践制约为数据基础。因数据具备动态性，需根据特定数据类型的时间周期进行调整。因此，基础数据来源的真实性和安全性是影响数字孪生技术发展的基础性要素。基础数据来源主要包括传感器采集、人工采集和历史数据等模式。其中，传感器采集的基础是根据需求部署大规模的传感器，其导致大量的前期

投资成本。且传感器的统计口径、核算校准等情况直接影响各地域、各主管部门间的基础数据能否实现比较，影响数据的准确性。人工采集的基础是大量人力成本的投入，且需保障人工采集的真实性，既耗费大量人力成本，其准确性也难以得到充分保障。历史数据则是根据原数据基础进行直接录入，统计成本较低，但难以保障其数据准确性。

除传感器采集可实现一定周期的数据自动采集和上报外，人工采集的周期普遍较长，难以保障数据基础的及时和可靠。且历史数据可能面对同类数据统计口径调整，导致采用历史数据作为基础的趋势研究科学性有待提升。即时数字孪生技术能获得足够的成本以开展数据搜集工作，目前数据的基础与采集模式也难以保障技术正常开展与应用。因此，从数据架构、数据类型和数据特点上进行深入研究，降低数字孪生技术需要的必要性数据数量，是数字孪生技术在数字政府管控领域得到广泛应用的重要保障。

3.2 发展展望

（1）开放平台标准化

数字政府管控面对多层级、多类型、多需求的特点，目前各主导单位建立属于自身的标准化开放平台，但相互间却难以对接。为数字孪生技术进一步指导数字政府发展，需开发整套开放平台作为数字政府的核心连接组织，各主导单位逐步介入开放平台，实现平台功能的拓展。开放平台需标准化，实现对各类数据的兼容并蓄。且开放平台需明确各类数据的标准格式和标准来源，作为未来数字政府数据大幅增加的提前准备。

（2）多接口并行统一

数字政府管理需求中，存在于多个主导单位间的连接，多接口并行带来数据量和数据优先度的差异。根据数字政府需求确定数据优先度，并编制相应编号层级体系，有助于实现并行多接口间的数据统一。数字政府存在的多个主导单位间，需兼顾政府扁平化管理模式，对现有政府审批模式进行进一步优化，确保数字政府为市民带来更好的便利。

（3）设施的共建共享

数字政府管理中，各类运营主体会建设自身的数字基础设施，作为相应运营主体在城市发展进程

中的重要优势和未来盈利的主要来源。在现实运行过程中,让运营主体放弃自身的比较竞争优势并不容易,因此基础设施层面共建共享推动难度极大。从数字政府管理角度,为实现数字孪生技术效果,各设施间共建共享将有助于降低数字政府管理总体成本,为设施规划、建设和维护成本带来重要的经济支撑。

4.小结与启示

城市是一套复杂系统,对城市进行研究的学科包括城市规划学、城市经济学、城市地理学、城市社会学等多种内容,其均从一个侧面揭示城市系统的内容。但目前对城市的研究主要包括现状研究、历史演变研究,难以对城市目前情况进行分析和对城市未来趋势进行研判。在科学进步的情况下,国外学者提出数字孪生技术,对城市系统进行系统仿真模拟,降低实践成本,且已在城市管理领域有所尝试。数字孪生技术的重要创新点在于数字城市全生命周期的管理,实现政府管理的现状分析、趋势研判和安全预警等重要功能,是智能算力赋能数字政府管控的重要策略。

数字孪生技术包含多个系统模型,在未来数字政府进一步发展提供多种可能性。但数字孪生技术当前面对的主要问题在于底层系统和接口的标准化,即在同一标准下实现系统开源和应用拓展。现有建筑信息模型、地理信息模型、三维场景系统模型、物联网系统模型和人工智能系统模型均为数字孪生技术应用提供良好的系统基础,但将相关系

统统筹形成重要数字政府管控平台,为未来各领域逐步接入和应用于数字孪生技术提供重要的助力。对数字孪生技术的背景与发展、模型组成部分、实践制约与发展展望进行研究,将为智能算力赋能数字政府管控提供新型技术策略,对数字经济和传统产业数字化转型提供重要的基础支撑,对实践国家“探索建设数字孪生城市”战略具有紧迫性和必要性。

参考文献:

- [1]朱惠斌.国外网络空间治理思路与展望[J].通信与信息技术,2022(S1):63-66.
- [2]胡梦珠.我国城市韧性研究的文献计量分析[J].山东行政学院学报,2022(05):118-128.
- [3]朱惠斌.国内外数字孪生技术研究进展与实践展望[J].信息通信技术,2022,16(05):75-80.
- [4]汤资岚,杨旭.数字技术赋能乡村协商民主:价值、困境与进路[J].辽宁行政学院学报,2022(05):35-40.
- [5]朱惠斌.城市智慧交通策略探索[J].江南论坛,2022(07):43-46.
- [6]朱惠斌.汇聚机房算力效率模型与提升策略探索[J].电信快报,2022(07):10-13+17.
- [7]朱惠斌.国有自然资源数字化管理思路探索[J].国有资产管理,2022(07):38-41.

视频中台与边缘云技术融合应用的研究与实践

李 涵

(福建省邮电规划设计院有限公司)

摘要: 随着网络信息化的不断发展,行业对于视频类业务的需求在体量高速增长的同时愈发多元化,视频中台的规划与建设将成为这一发展阶段下的重要工作;与此同时,我国边缘云计算的相关技术正在日趋成熟,亟待推广与应用。本文对此提出一种基于边缘云架构的视频中台解决方案,重点介绍了视频中台与边缘云技术在融合应用中的主要考量因素及其价值。

关键词: 视频中台,边缘云,智能化

引言

近年来,我国持续推进网络基础设施建设,以第五代移动通信技术(5G)、数据中心(IDC)、人工智能(AI)、物联网等为代表的信息数字化能力不断提升。移动通信技术的发展不仅带来网速的提升,同时推动流量单位成本下降,“提速降费”让视频服务类应用程序(App)的用户量和活跃度不断增加。高新视频将成为主流媒体业务形态,相关客户的要求也随之严格,在低延时、高效率等关键技术能力上提出了更高的要求。

目前我国一二线城市已基本进入海量视图大数据分析处理的应用和建设阶段。这一阶段下单一的视频基础体系架构逐步向视频云体系架构转移,传统的云体系架构由于其集约化的典型特征,无法很好地支撑现有及未来的多样化应用和适配新形势下业务发展需求,例如海量视频流媒体需要高带宽保证实时调取、分布式多节点智能解析处理等。面对行业的风险和挑战,如何探索一套新的视频业务体系是适应技术发展趋势的当务之急。

1. 视频中台的建设模式探讨

1.1 中台化下的视频云平台

随着 5G 时代的到来,视频、图像业务应用不

断深入拓展,视频云平台的运行需要海量计算和存储资源来支撑图形分析、流媒体存储、机器视觉以及视频资源共享与开放,对计算、存储、网络等资源的利用效率提出了更高的标准,这将在一定程度上要求视频云平台具有中台级的统一管理调度能力。

视频中台模式是视频云平台发展到一定规模阶段后的外在体现,在这个阶段,平台将在确保功能兼容性和稳定性的前提下,突破单一业务片区的局限,实现模块化的、可复制的多点部署,形成资源融合、云网融合、视频能力融合的新型平台架构,从而满足大量新增业务对平台功能、接入及存储等能力的需求。

1.2 中台建设思路

传统的视频业务架构中,前端业务与后端功能模块通常是点对点适配。随着新业务复杂度的提高,在传统的“烟囱式”架构下易形成数据孤岛,业务难以实现快速开发和部署。

为了改善这种状况,视频中台将对市面主流业务场景视频类功能需求进行沉淀,从而实现功能模块与资源的有效复用,提升业务响应和规模化创新能力,降低建设成本并提高服务质量的的功能。

视频中台的总体架构可在云计算的基础设施

作者简介:

李涵(1991-),男,工程师,从事业务网通信方面的设计、研究工作。

即服务(IaaS)/平台即服务(PaaS)/软件即服务(SaaS)标准结构基础上,根据行业对各项业务功能需求度的实际差异,在市场调研后按照基础功能、可选功能、定制化功能三类功能点对其进行补充。视频中台能力架构样例如图 1 所示。

(1) 基础功能:全省较为共性的需求,建议集中规划、统一部署。

(2) 可选功能:个性化场景能力,本地侧根据实际业务情况按需开通。

(3) 定制化功能:对于尚未统筹规划,并在业务发展过程中出现的能力需求,采用大数据时代的信息通信技术服务(DICT)方式快速迭代开发满足。

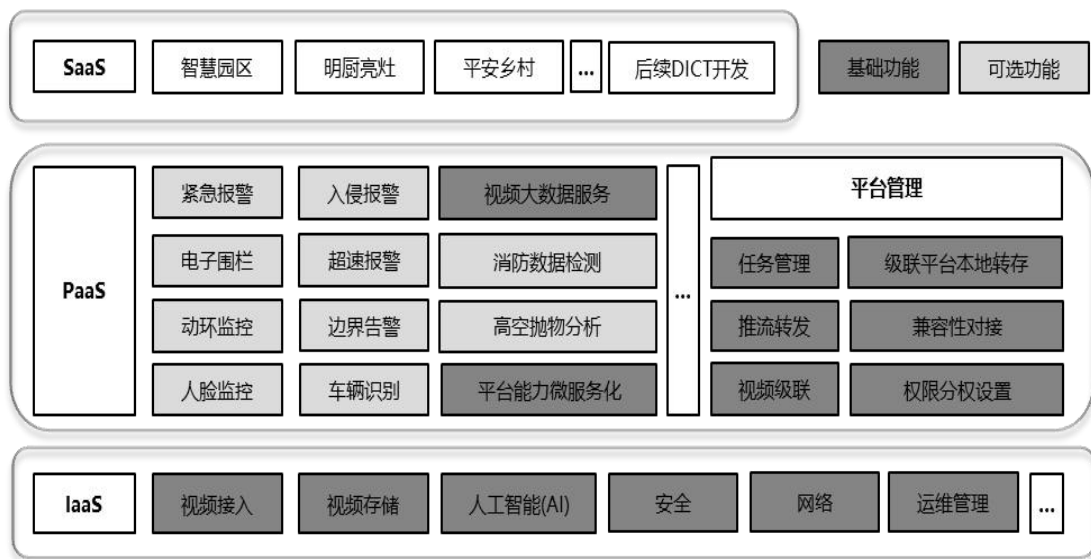


图 1 视频中台能力架构样例图

以某视频中台为例,其涉及以下能力建设:

(1) 视频接入

终端设备通过国标协议、厂商协议或私有协议接入平台,平台通过信令及视频流、网络交互,保证视频预览稳定性,支持所有国标设备(GB28181)接入,支持接入、直播、录像、回放、云台控制、负载均衡等主要能力。

(2) 推流转发

管理端通过与设备信令交互,前端设备推流至平台流媒体服务器,再由流媒体服务器利用流式传输技术把视频流进行转发、分发至存储处理服务器或第三方视频播放器,实现视频流资源合理使用和视频流复用。

(3) 视频存储

存储处理程序把流媒体服务器分发的实时视频流转换成录像文件 mp4 格式并写入本地磁盘中,客户端可通过录像查询进行调阅,支持对接分布式存储、简单存储服务(S3)协议存储、本地存储。

(4) 视频级联

支持级联下级平台,通过国标级联的方式,接入下级平台,同步下级平台数据,以同步下级平台的设备资源,实现下级平台的设备管理等功能。支持级联上级平台,通过国标级联的方式,接入上级平台,视频数据推送到上级平台。

(5) 级联平台本地转存

将下级平台的级联对接上来的视频数据进行本平台转存,在本平台存储。

(5) AI 能力

为满足客户对于 AI 能力的使用需求,开发或引入 AI 引擎,进行 AI 能力孵化及推演,相关能力包括图像识别、语音识别、文字识别(OCR)、事件识别等。边缘节点建设图像处理(GPU)服务器作为 AI 硬件能力,为平台赋能 AI 处理能力。

(6) 兼容性对接能力

平台具备灵活定制开发对接地市个性化平台、并预备与其它大型视频中台通过国标级联或者接口对接的方式进行数据接入的能力。

(7) 权限分权设置

管理员可自行创建子账号并配置账号权限，管理员可以自行进行平台权限的分配和管理，自定义账号的查看权限。

(8) 视频解码展示

根据项目实际情况，后台需适配不同型号的解码器或网络视频录像机（NVR）等设备，对接视频流数据，进行视频流编解码，并将视频解码上墙进行大屏展示。前端界面需新增大屏展示的功能，包括但不限于设备管理、大屏区域管理等。

(9) 对讲功能

前端摄像头具有拾音器的情况下可实现对讲功能，客户端可与摄像头进行语音对讲。

(10) 多屏查看

客户端多显示屏查看，支持在不同显示器上展示不同的监控画面。

(11) 上级联视频调阅

上级联平台视频调阅，支持上级联平台调阅下级联平台的录像存储数据。

(12) 手机端 NVR 前端录像调阅

手机端 APP 支持多客户并发访问调阅、查看、回放等功能并新增调阅前端 NVR 的录像。

(13) 代理登录功能

有些客户不能直接访问外网，需要通过代理功能才能登入个人电脑（PC）端，平台需要增加代理模块方便用户访问 PC 端。

2. 基于边缘云技术的视频中台实现方案

如前文所述，传统集约化的数据中心式云资源池已逐渐无法满足当下视频媒体业务的发展以及视频中台建设的要求，本文将以边缘云作为切入点，探索一种基于边缘云技术的视频中台实现方案。

2.1 边缘云模式的优势

市场层面

2020 年 3 月 4 日，中共中央政治局常委会提出加快包括数据中心在内的“新基建”进度，加大建设投资力度，强化要素保障与安全保障体系。数据中心作为数字时代各项服务的基础设施，将得到更大力度的发展支持。随着公有云竞争壁垒日益提高，边缘云将成为云市场最佳切入点。边缘云为企业提供更稳定的业务发布，交付可承诺的维护响应速度（SLA）、更快的访问速度、更灵活的业务架

构，加速企业转型。从目前规模来看，国内边缘云（专有云）市场整体上还处于比较初级的阶段，发展潜力较大。

技术层面

边缘计算节点能够及时有效地处理边缘端的数据请求，提升边缘端的信息处理效率。目前边缘云需求主要涉及政府、工业、医疗、交通、能源等行业，通过以下技术特点形成其独有的对客户吸引力：

(1) 保证敏感数据的安全存储需求，实现数据不出地市，满足政企机构上云对业务稳定、数据安全和灵活性的双重关注；

(2) 通过属地化部署，实现对本地资源池的低时延访问、以及本市网际互连协议（IP）地址本地 IDC 互联。利于企业内部信息流程与云端无缝衔接、数据双向迁移，帮助企业产生新的业务价值；

(3) 计算和存储资源的边缘化，极大的降低了海量视频、图片数据在传输、存储、应用等方面对数据中心的严重依赖；

(4) 可有效减轻边缘端到数据中心端的网络带宽和应用请求带来的资源占用，降低对高成本跨市专线的投资；

(5) 有效满足党政军及行业重点客户，如 XX 分局/派出所对本端的人脸比对、布控告警等实时业务的数据处理、分析需求。

2.2 应用案例

以运营商建设模式为例，视频中台的构建应全面统筹、统一规划，聚焦平台 IaaS 层资源搭建、PaaS 层能力迭代共享——通过 IaaS 层提供统一的硬件环境、通过 PaaS 层提供丰富的功能模块、通过 SaaS 层对功能模块的调用组合，实现业务快速部署。对省内统一提供资源开放、云网融合、视频能力融合的服务，承载地区个性化政企视频监控业务发展。

结合边缘云技术特点，采用“中心—边缘”的总体布局，中心节点作为算力中枢，对全省的视频接入和视频存储进行统一管理及资源调度、形成能力库并向前端提供中台能力、同时向其它内部系统和外部平台提供能力接口；边缘节点下沉部署至各地市，为本地视频流的接入和存储提供硬件能力，并与客户平台实现个性化对接。总体架构如图 2 所示。

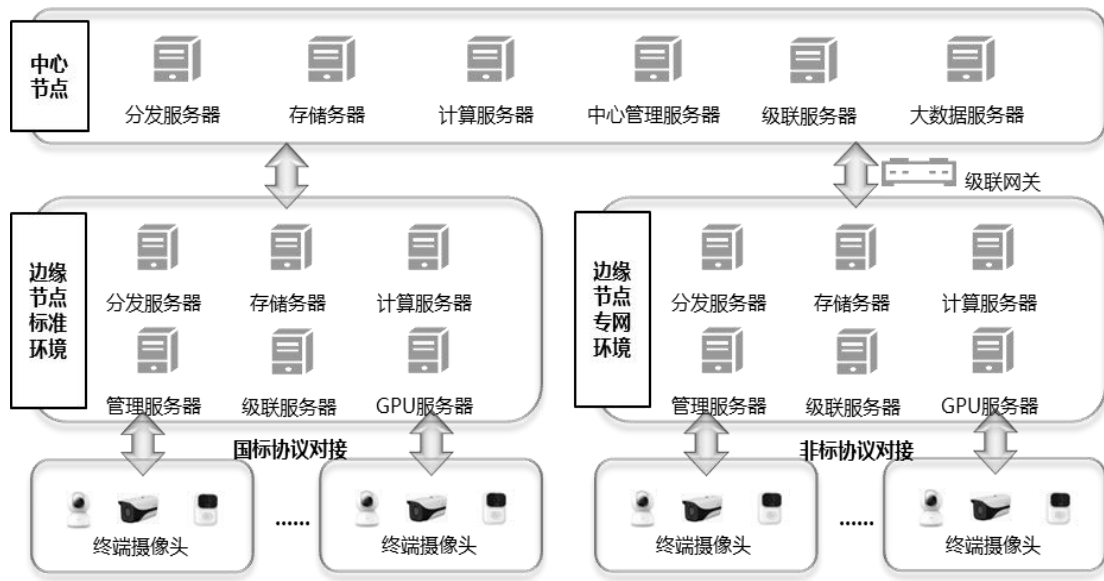


图 2 视频中台组织架构样例图

以某省运营商的视频中台建设项目为例，考虑到项目业务重要性及对输入输出（I/O）能力的较高要求，采用“中心—边缘”的两级布局架构，在省会数据中心建设中心节点，构建计算、存储、管理、网络、安全、视频中台等总体能力；并在各地市部署接入、存储节点，通过统一的边缘云技术架

构实现对全省资源的动态分配和异地调度，降低延时，提高处理效率，避免单点故障。中心侧通过公网实现对全省公网接入业务的统一调度管理、边缘测可根据本地业务拓展需要，个性化对接本地政企平台，数据级联到省专网平台，满足专网接入的业务需求。项目网络拓扑如图 3 所示。

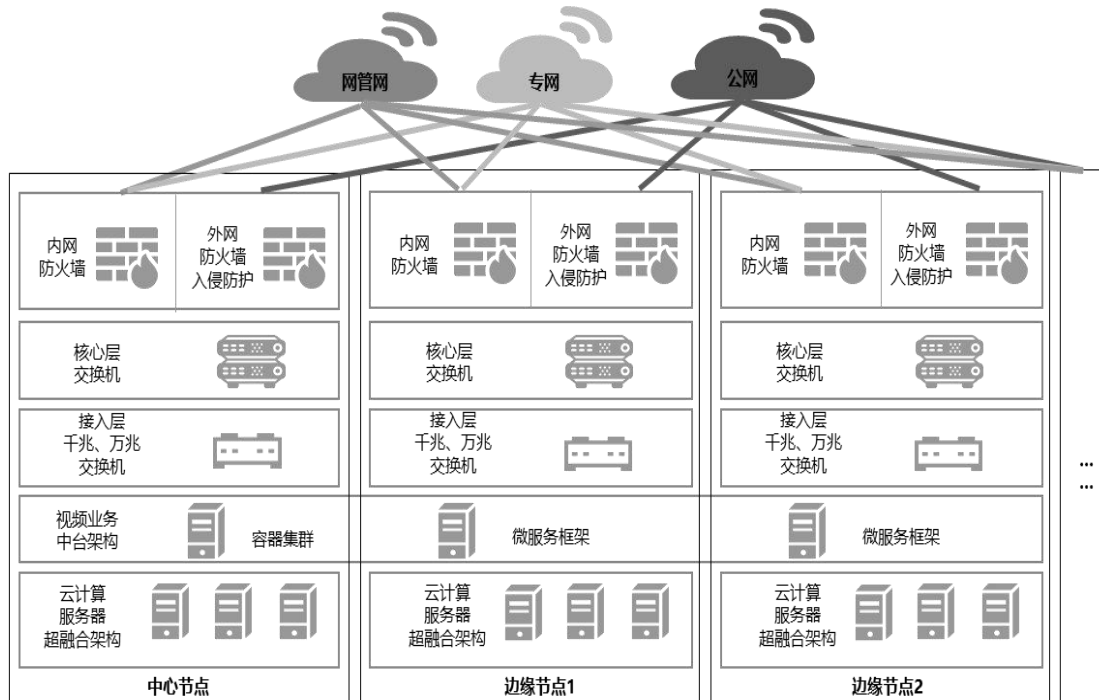


图 3 视频中台网络拓扑样例图

2.3 应用成效

上述案例项目在实际投产运行过程中，系统运行稳定、运维管理方便、开发升级迅速、用户体验良好，可为业主带来良好的直接经济效益与间接效益：

(1) 满足业务发展需求和用户需求

有效促进本地视频监控业务的发展，有利于改善、培育本省视频业务土壤，为客户提供更加完善的视频监控服务。

(2) 进一步降低运营成本

视频存储在各地市，减少网络传输成本，提高数据传输效率。且软硬解耦模式使得系统的升级与扩容变得灵活可控，可第一时间响应市场当前热点技术与需求，如视频可进行压缩或抽帧存储，有效降低视频运营成本等。

(3) 有利于提升用户体验，为公司创造价值

面对市场竞争日渐激烈的现状，尤其是面向低延时、大流量的视频监控需求日益迫切。“中心—边缘”的组织架构可以有效保障用户访问速度，进一步提升用户体验和满意度，从而为公司创造更大价值奠定良好的基础。

3. 结语

当前，新型视频业务的技术与需求正随着社会不断发展不停变化，但视频业务中台与边缘云技术的协同模式将是未来重点转型的方向，这一模式将是实现全覆盖、全高清、全联网、全智能的科技强国的重要目标。本文基于探索实践，提出一种较为通用的业务中台+边缘云的设计方案，从平台功能结构、组织结构和网络结构几方面进行详述，希望可以为投身信息化建设的同胞提供新的思路。

参考文献

- [1]杨茜泽,朱里越,汤新坤.基于智慧广电的视频中台建设[J].广播电视网络.2022, 388(4): 24~25
- [2]刘博斐,朱雨豪,刘伟,廖光文.视频云+边缘计算架构在新型智慧城市中的建设模式[J].电子技术与软件工程.2020年第10期: 141~142
- [3]黄智伟.云计算下的云资源池网络分析[J].通信电源技术 2019年第36卷第9期: 165~167
- [4]夏翠兰,施文丽.基于视频云计算技术打造云时代互动电视增值业务运营平台的策略[J].西部广播电视 2020年第17期: 232~234

智能算力赋能园区数字化转型

范永圣

(中鸿达信息科技有限公司)

摘要: 智慧园区是现代城市发展的一种重要模式,通过整合各种物联网设备和数字技术,实现园区内各项管理和服务的智能化。其中,智能算力作为重要的支撑技术之一,为智慧园区的精细化管理水平提供了强大的能力。本论文以智慧园区为背景,探讨了智能算力在提升智慧园区精细化管理水平方面的作用。

本文从智能算力的基础概念出发,介绍了智能算力和智慧园区的概念。其次,阐述了智能算力在智慧园区规划设计、设备管理维护、运营管理等领域的的作用,最后,本文剖析了智能算力在园区领域面临的问题和挑战。同时,本文对于智能算力在智慧园区应用中可能出现的安全性、隐私保护和资源利用等问题进行了探讨。

最后,本文总结了智能算力赋能智慧园区的优势和潜力,并提出了未来研究的方向和建议。通过深入挖掘智能算力的潜力和探索适合智慧园区的创新应用,可以进一步提升智慧园区的精细化管理水平,实现园区的可持续发展。

关键词: 智能算力、智慧园区、赋能

引言

1. 智能算力的概念

智能算力是指由基于 GPU、FPGA、ASIC 等 AI 芯片的加速计算平台提供的算力,主要用于人工智能的训练和推理计算,例如语音、图像和视频的处理。智能算力由人工智能的核心计算能力提供,注重单精度、半精度等多样化计算能力。

智能算力的原理是利用人工智能芯片进行高速计算,从而实现人工智能应用的训练和推理。GPU、FPGA、ASIC 等 AI 芯片都具有高效的计算能力和并行处理能力,可以快速处理大量的数据,并进行深度学习、机器学习等计算任务。智能算力主要用于支持人工智能领域的的应用,如语音识别、图像识别、自然语言处理等。

2 数字园区的概念

将传统的园区转变为高度集成的智能化、网络化、数字化、可持续发展的园区。智慧园区包含了人员管控、楼宇管理、节能减排、智慧交通、安防监控、企业服务、生态保护等多个方面,通过物联网、云计算、大数据、人工智能等技术的应用,实

现了数据的共享和资源的整合,进而提高了园区的综合素质和市场竞争力。

总之,智慧园区是一种利用现代技术手段实现高度集成、智能化、数字化、可持续发展的园区,具有高效、智能、数字化、可持续等特点,能够提高园区的综合竞争力和市场影响力。

3 智能算力赋能智慧园区的发展

智能算力在智慧园区的发展中起到了关键作用,可以涉及以下方面:

1. 智慧园区规划与设计:智能算力可以通过数据分析和模拟等手段,辅助制定和优化智慧园区的规划与设计方案。它可以模拟不同的布局和配置,评估各种因素对园区运营效率和用户体验的影响,以帮助设计出更高效、可持续和宜居的园区。

2. 智慧园区设施管理与维护:智能算力可以实时监测园区内各类设施的状态和性能,识别潜在问题并提供预警。通过与传感器、物联网设备等的结合,智能算力可以实现设施管理的自动化和智能化,提高设施维护的效率和准确性。

3. 智慧园区智能化运营与管理:智能算力可以

集成并分析园区内各类数据，包括人员流动、能源消耗、交通情况等，从而实现园区内各项运营活动的智能化管理。通过对数据的分析和预测，智能算力可以优化运营策略、资源利用和决策制定，提高园区的整体效率和可持续性。

总的来说，智能算力通过数据处理、模拟分析和决策支持等手段，为智慧园区的规划、设施管理和运营提供了强大的技术支持，促进了园区的智能化和可持续发展。

1 智能算力在智慧园区规划与设计中的作用

1.1 智能算力支持智慧园区的空间规划

智能算力可以在支持智慧园区的空间规划中发挥着重要作用：

1. 利用智能算力处理大规模的地理和空间数据，包括土地利用、建筑物分布、人口统计和用户行为等数据。通过数据分析和预测模型，可以识别潜在的发展趋势、热点区域和需求特征，为空间规划提供基础和指导。

2. 基于 GIS 的空间分析，地理信息系统 (GIS) 结合智能算力，可以提供强大的空间数据管理和分析能力。使用 GIS 技术，可以对园区内的地形、地貌、道路网络、建筑物和设施进行三维建模和可视化分析，以辅助空间规划决策和评估。

3. 通过智能算力和物联网技术，对园区内的建筑物、设施和基础设施进行智能化管理和布局。运用智能算力分析空间需求、能源利用、通风采光等参数，优化建筑和设施在园区中的位置和布局，提高资源利用效率和用户体验。

4. 借助智能算力，可以建立可持续发展模型和评估指标体系。通过分析环境、能源、社会和经济等方面的数据，智能算力可以支持园区的可持续发展规划，以平衡发展和环境保护的需求。

综上所述，智能算力可以进行智慧园区的空间规划。这将有助于优化资源利用、提升用户体验，实现智慧、可持续和宜居的园区发展。

1.2 智能算力推动智慧园区的绿色发展

智能算力可以在在推动智慧园区的绿色发展中发挥关键作用：

1. 利用智能算力处理实时能源数据，并结合

智能控制系统，实现智能能源管理和优化。智能算力可以分析能源消耗模式、预测负荷需求，并提供智能调度和优化控制，以最大程度地减少能源浪费和碳排放，提高能源利用效率。

2. 通过智能算力和物联网技术，对园区内的建筑和设施进行智能化管理。智能算力可以分析建筑物的能源使用情况、运行效率和室内环境质量，并利用智能控制系统进行优化。通过智能照明、智能空调、智能水管理等解决方案，实现能源节约和减排。

3. 利用智能算力对园区内的废物管理进行优化。通过智能传感器和数据分析，可以实时监测和管理垃圾和废物的产生和处理。智能算力可以优化垃圾收集和处理的路线规划、资源回收利用等，减少废物的产生和对环境的影响。

综上所述，智能算力将有助于减少能源消耗、废物排放和环境污染，提高资源利用效率，实现可持续、环保和绿色的园区发展。

2 智能算力在智慧园区设施管理与维护中的应用

2.1 智能算力辅助智慧园区设施监控与维护

智能算力在辅助智慧园区设施监控与维护方面可以发挥重要作用：

1. 智能算力可以通过物联网技术和传感器网络，实现对各种设施的实时监测。例如，通过智能传感器监测设备的运行状态、能耗情况、温度、湿度等参数，可以及时发现异常情况，并提供实时报警，并进行相应的维修和保养，从而避免设备的故障和停机时间的延长。

2. 智能算力可以处理大规模的设施数据，并利用机器学习和数据分析技术，识别设备的健康状况、寿命预测和故障潜在因素等。通过智能算力的辅助，可以进行设备故障预测和维护计划优化，提高设备的可靠性和运行效率。

3. 智能算力可以通过云平台和远程控制技术，实现对设施的远程监控和控制。这样，设备运行状态和维护需求可以随时被监测，并能够通过远程控制系统进行实时操作和维护。

4. 利用智能算力和自动化技术，可以实现设施的自动化维护与优化。例如，通过智能算力和自

自动化控制系统，设备可以自动调节运行模式、维持最佳运行状态，同时进行自动化的保养和故障排除，减少人工干预和维护成本。

5. 通过智能算力提供的大数据平台，可以集中存储和管理设施监控数据。这些数据可以被用于分析设施运行情况趋势、优化维护计划、预测设施性能以及支持决策制定。

综上所述，智能算力将有助于及时发现设备问题、提高维护效率、减少停机时间，实现设施的高效运行和延长使用寿命。同时，还能够提升用户体验和园区的可持续发展，进一步推动智慧园区的建设和发展。

2.2 智能算力提升智慧园区设施维护效率

智能算力在提升智慧园区设施维护效率方面扮演着重要的角色：

1. 智能算力还可以通过自动化技术和机器学习技术，实现设施的自动化运营和维护，减少人工干预和故障率。自动化运营可以大大提高工作效率，减少人为错误和延误，确保设施的稳定运行。

2. 智能算力还可以通过数据分析和优化运营，实现能源的合理分配和消耗，降低能源成本 and 环境污染。这样可以帮助园区管理机构实现可持续发展目标，同时也可以提高设施的运行效率。

3. 智能算力可以结合文本分析和自然语言处理技术，提供维修指导与故障排除支持。通过智能算力的辅助，设施维护人员可以快速获取设备的操作手册、维修指南、故障症状分析等信息，提高维护效率和准确性。

综上所述，智能算力有助于提升智慧园区设施维护效率，减少运营成本和故障率，实现更加智能化和高效的设施维护。

3 智能算力在智慧园区智能化运营与管理中的应用

3.1 智能算力支持智慧园区的智能化运营

园区的智能化运营中有着广泛的应用：

1. 智能算力可以通过人工智能技术，实现智能化管理和服务。例如，通过人脸识别技术，实现人员的快速进出和安全监控；通过自然语言处理技术，实现智能客服和语言翻译，提高服务质量和效率。

2. 智能算力还可以通过物联网技术，实现设备的自动化监控和运营。通过实时监测设备的运行状态和能源消耗，可以实现能源的优化分配和节约，提高运营效率。

3. 智能算力还可以通过安全监控技术，实现园区的安全管理和监控。包括视频监控、入侵检测、消防安全等，确保园区的安全和稳定。

综上所述，智能算力在智慧园区的智能化运营中有着广泛的应用，可以实现更加智能化和高效的运营。

3.2 智能算力提升智慧园区的精细化管理水平

智能算力可以通过以下方式提升智慧园区的精细化管理水平：

1. 智能算力可以通过大数据分析和挖掘，实现对园区内各种数据的精细化管理。包括人员流动、设备运行、能源消耗、安全监控等数据，可以进行深度分析和挖掘，为管理决策提供科学依据。通过数据分析，可以发现一些潜在的问题和机会，帮助园区管理机构实现精准决策和精细管理。

2. 智能算力还可以通过人工智能技术，实现智能化管理和服务。例如，通过人脸识别技术，实现人员的快速进出和安全监控；通过自然语言处理技术，实现智能客服和语言翻译，提高服务质量和效率。通过智能化管理和服务，可以提升园区的运营效率和质量，提高客户满意度。

3. 智能算力可以通过智能化运营平台，实现各种信息的整合和共享。通过智能化运营平台，可以将各种信息整合起来，实现信息的共享和协同，提高运营效率和质量。同时，还可以实现任务的自动化分配和跟踪，提高管理效率和准确性。

综上所述，智能算力可以提升智慧园区的精细化管理水平，实现更加精准、高效、可靠的管理和服务。

4 面临的问题与挑战

4.1 智能算力在智慧园区应用中面临的问题与挑战

智能算力在智慧园区应用中的使用可以带来很多优势和便利，但同时也面临一些问题与挑战：

1. 智慧园区应用中涉及大量的数据采集和处理，包括个人身份信息、交通数据、能源消耗等。

因此,保护数据的安全性和隐私成为一个重要问题。存在着数据泄露、未经授权的访问以及数据滥用的风险,需要通过严格的安全措施和有效的隐私保护机制来解决。

2. 智慧园区应用涉及多个系统和设备,需要进行技术整合与兼容性的处理。不同设备和系统的技术标准、通信协议等差异性,可能导致数据传输和信息协同的困难。为了实现智慧园区的整体优化和管理,需要解决设备的互联互通的问题。

3. 智慧园区的规模往往很大,涉及多个子系统和设备。在不断增加新的子系统和设备的过程中,系统的复杂性和可扩展性成为一个挑战。需要建立灵活的系统架构和强大的智能算力支持,以适应园区不断变化的需求。

4. 智慧园区应用中需要大量的能源供应和设备运行,可能会对环境造成影响。由于智慧园区应用的复杂性,需要更多的能源供给和处理,导致能源消耗问题。因此,研究和采用能源高效的智能算力解决方案非常重要。

5. 智慧园区应用需要涉及多个利益相关者,包括园区管理者、业主、居民等。在推动智慧园区应用过程中,需要解决社会接受度和治理问题。这包括信息透明、权益保护和合理治理机制的建立,以确保园区居民的参与和支持。

综上所述,智能算力在智慧园区应用中虽然带来了许多优势,但也面临一些问题与挑战。通过加强数据安全与隐私保护、技术整合与兼容性、系统复杂性可扩展性、能源消耗与环境影响以及社会接受度与治理等方面的研究和措施,可以有效应对这些问题与挑战,推动智慧园区的可持续发展。

4.2 如何解决智能算力在智慧园区应用中的问题与挑战

解决智能算力在智慧园区应用中的问题与挑战需要综合考虑技术、管理和政策等方面的因素:

1. 加强数据加密和访问控制,确保数据在传输和存储过程中的安全性。采用隐私保护技术,如数据脱敏、差分隐私等,保护个人隐私。建立合规性框架和政策,明确数据使用和共享的规则,加强对数据隐私的监管和保护。

2. 制定统一的技术标准和通信协议,促进设备之间的互联互通。采用开放式架构和接口,方便

不同系统和设备的集成。通过使用中间件和 API,实现系统之间的数据交换和信息共享。

3. 建立灵活的系统架构,采用模块化设计,以便于系统的扩展和升级。利用云计算和边缘计算技术,实现资源的弹性分配和管理。采用智能算法和自动化技术,提高系统的自适应性和自动化程度。

4. 采用能源高效的硬件设备和优化的算法,降低能源消耗。通过能源管理系统和智能控制策略,实现能源的优化利用。推广可再生能源的应用,减少对传统能源的依赖,降低环境影响。

5. 建立透明的沟通机制,与居民和业主进行充分的信息交流和参与。制定合理的治理机制,确保各方利益的平衡。加强社会教育和意识提升,提高公众对智慧园区的理解和接受度。

总之,解决智能算力在智慧园区应用中的问题与挑战需要综合运用技术手段、管理措施和政策支持。通过加强数据安全与隐私保护、技术整合与兼容性、系统复杂性可扩展性、能源消耗与环境影响以及社会接受度与治理等方面的工作,可以推动智慧园区应用的可持续发展和成功实施。

5 结论

智能算力在智慧园区建设中具有重要作用,可以推动园区管理创新和提高园区的运营效率、安全性和可靠性。但是,仍需要进一步研究和应用,以实现更加安全、高效、智能化的管理手段和服务模式。未来,智能算力将在智慧园区建设中发挥更加重要的作用,为园区的可持续发展和高效运营提供有力支持。

未来研究方向可从以下方面入手:

1. 进一步研究智能算力的应用技术和方法,探索更加安全、高效、智能化的管理手段和服务模式。

2. 加强对数据安全和隐私保护的研究和探索,提出更加安全和可靠的保护措施和方法。

3. 拓展智能算力在智慧园区建设中的应用领域,如智能交通、智能环保、智能安防等。

智能算力是智慧园区建设的重要技术支持,其应用将推动园区管理创新和提高园区的运营效率、安全性和可靠性。未来,智能算力将在智慧园区建设中发挥更加重要的作用,为园区的可持续发展和高效运营提供有力支持。

基于智算引擎的 AI 应用平台的关键技术与实现

傅 玮

(中国电信股份有限公司福建分公司)

摘 要：福建电信 2021-2023 年逐步建立了人工智能体系化的技术框架，依托泛在的网络资源和 AI 算力，不断研究并完善体系化中所需的理论、技术、机制、范式和框架，打造了云中心、边缘算力、SDC 摄像头三位一体的云边端协同的计算平台，融合数据接入、数据解析、数据治理、数据安全输出等能力的智算引擎实现视频图像解析、人脸聚类、以图搜图等功能，完成智能数据向数据智能的转化。智算引擎对 AI 能力进行灵活且高效的配置、调度、训练和部署，不仅能满足日益丰富的数智化业务需求，同时确保 AI 业务的可信可控安全。福建电信基于此架构，为客户打造集成数据存储、数据分析、数据应用于一体的 AI 应用平台，实现了应用层的展示。AI 应用平台为千行百业的客户提供了千人千面的行业解决方案，而客户也能借助人工智能技术实现价值增长。

关键词：CV、AI、SAAS

1 概述

1839 年 1 月法国摄影师达盖尔发明了照相机，自此人类学会了如何保存眼睛看到的图像，但这还不够，我们还想让机器自己去看，并且告诉我它看到了什么，这就是计算机视觉。计算机视觉比较严谨的定义如下：“基于感知图像做出对客观对象和场景有用的决策”（Sackman & Shapiro, 2001）。

在计算机视觉诞生之前，为创建一个程序来识别特殊的图像，需要耗费数小时的时间，来手动完成这类繁琐的工作。首先，需要核对一个相似图像的数据库，然后不得不人工分析、测量这类图像，当研究人员识别到目标时，标注相关的数据（如：颜色、测量值及形状）。而现在，计算机视觉采用深度学习这类机器学习方法，自动完成上述的所有流程。

视觉识别是计算机视觉的关键组成部分，神经网络和深度学习等技术的进展极大地推动了视觉识别系统的发展，帮助识别视频流中的事件和面孔、对视频进行分类、自动添加字幕以及使用超分辨率等技术增强视频和图像。目前视觉识别应用较为广泛的 AI 技术例如：包括图像分类、对象检测、目

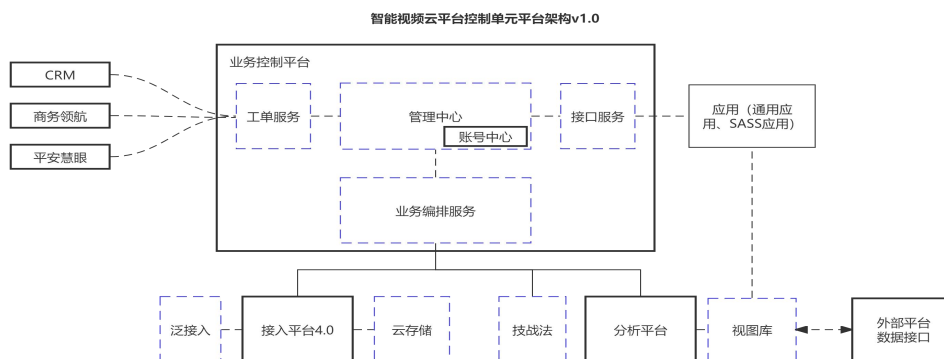
标跟踪、语义分割、实例分割等主要技术。

未来，人工智能技术会渗入到生活，化为无形，融于万物。AI 的智能化、数据化变革的背后既是时代发展的推动，也是政策、用户、技术共同推动的结果。首先，国家十三五规划、十九大报告等均提出加强视频监控智能化的目标，政府在智慧城市、雪亮工程等战略建设又为安防行业发展带来新契机；其次，随着城市化规模扩大，交通、楼宇、公安等场景需求广泛，拓宽了 AI 监控的覆盖面积；第三，大数据、云计算、生物识别、AI 等技术的发展与传统监控沉淀下来的大量图像、视频的融合，实现可视化、网络化、智能化为一体的监控架构将进一步打开和加速 AI 监控产品的扩大。

发展智能视频业务是运营商的必经之路，这不仅体现在国家层面将 AI 纳入新基建的政策支持，且从需求侧和供给侧，智能产业生态正在不断丰富，聚焦价值领域，获得越来越多元的应用场景和更大规模的受众。现在，不论是算法厂家、设备厂家还是系统集成商，都不具备全行业应用的开发能力。而运营商恰恰有面向全行业业务应用解决方案的能力：从 AI 技术到客户价值实现，需要一整套产品与解决方案，向客户最终交付的是应用，尤其是

不同行业、不同场景、不同用途的业务应用都不尽相同，需要针对性的集成与开发。

2 智算引擎平台



工单服务：包括 CRM、商务领航等多个渠道工单接收与处理；

业务编排服务：生产系统将能力注册到控制单元，控制单元将能力按项目编排到应用；

能力调用服务：应用系统调用生产系统能力时，通过控制单元实现统一认证。

智算引擎作为生产系统并不直接暴露在外部，业务人员通过控制单元进行业务配置，如将 AI 需求送到智算引擎，实现算法的任务启停、算法配置与监控。控制单元提供如下管理能力：

- 1、工单报表统计；
- 2、项目交付表管理；
- 3、AI 能力管理；
- 4、应用授权管理等

2.1 业务控制单元

业务控制单元是本文 AI 应用平台的智能调度中心，分为：工单服务、业务编排服务、能力调用服务三个部分组成，见下图：

2.2 智算引擎

福建电信 2022 年以 AI+为基础，结合大数据对原视频云平台的核心能力进行升级，逐步完善自研 AI 的品类，目前已接入人车非、区域入侵、明厨亮灶、垃圾屋监控、城市治理等多个大类算法。海量的 AI 事件经过视图大数据技术的加工后，生产出基于人脸、人体、车辆特征的时空轨迹等技战法应用，为行业客户定制智能化应用产品。因此，AI 产品化工作的核心是算法，本质是大数据。

智算引擎围绕算法引入标准化、算力管理集约化、AI 赋能便捷化三大核心功能，打造 AI 生态圈，为行业客户提供安全可信的智能视频数字化使能服务，支持轻量化部署，满足高可用特性，保障业务稳定性。其功能架构如下：



智算引擎具备跨域授权、潮汐调度、算力池化、多算法编排等核心能力，具体如下：

与行业主流算法厂商开展实质性合作，实现算法跨域授权、全网调度功能；

具备多厂家多算法动态授权能力，支持 License 申请、下发、回收等功能；

实现同一算力资源在不同时段支撑不同算法的分析需求；

采用算法任务+场景策略因子+多算法能力组合方式，充分利用计算资源，实现 AI 解析智能化、算力集约化，降低视频业务的分析成本；

适配多种 GPU 型号，采用容器化技术，结合基础配置的资源模版，快速生成 AI 算力单元，提供 AI 分析能力；

根据业务需求，具备高可用及集群负载能力，保障业务的快速扩容及稳定性；

多算法厂商多 AI 能力的自定义编排功能，实现同一视频源不同 AI 目标检测及解析，输出丰富的结构化数据。

2.3 计算平台

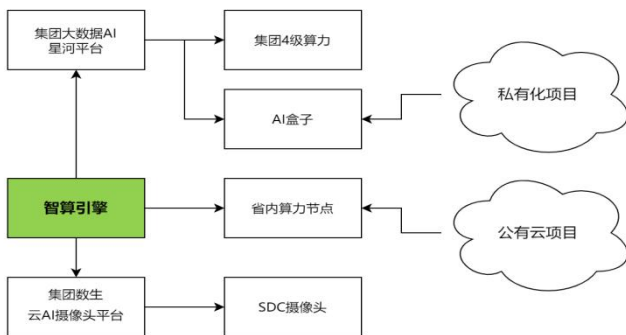
计算平台由云中心、边缘算力、SDC 摄像头构建而成，其中：

云中心：由英伟达和华为昇腾两类推理芯片组成；

边缘算力选用：集团大数据 AI 中心定制的 AI 盒子；

SDC 摄像头选用：集团数字生活公司云 AI 摄像头平台。

其中，由智算引擎完成多算力之间的任务调度，如下图：



3 AI 应用平台

3.1 一个案例

传统算法是：输入数据和规则，产生结果；

人工智能是：输入数据和结果，产生规则。

面向智慧农业、智能制造、智能交通、智慧物流、数字金融、数字商贸、数字社会、数字政府、其他数字化效率提升业等领域的机关、企事业单位，在发展过程中存在痛点、难点、堵点问题，都需要优秀数字技术、产品和解决方案。因此福建电信 AI 产品的思路是基于客户需求，打造标杆案例，以星火燎原之势，扩大规模发展。

以城市治理案例为例：

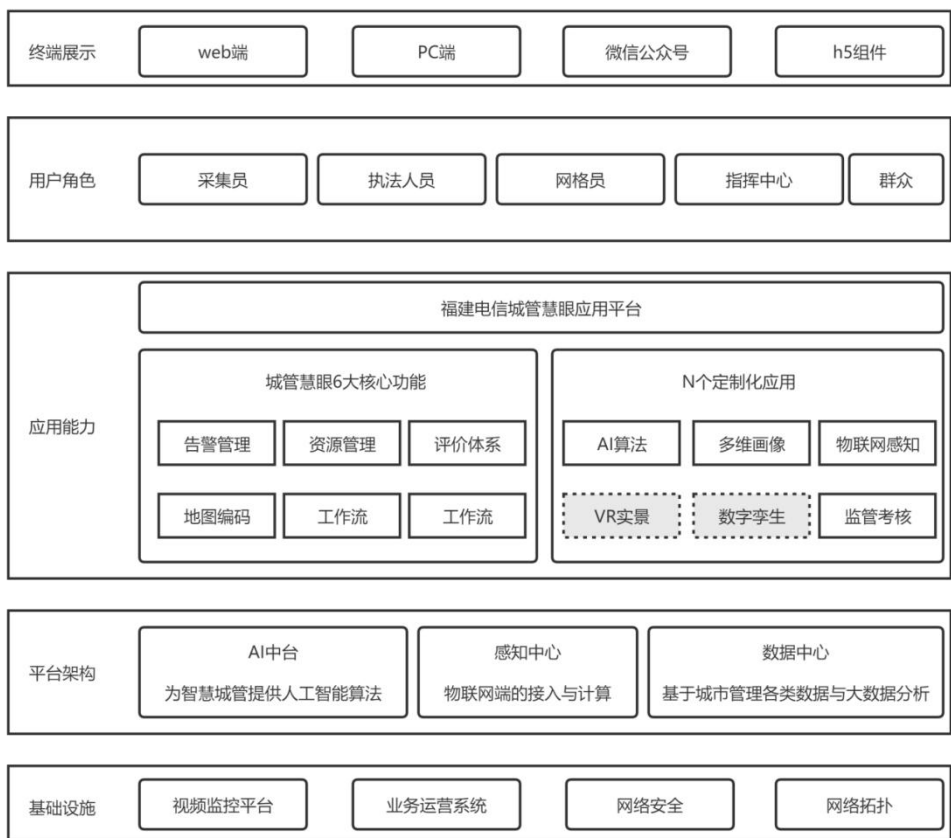
城管执法已有的三种模式分别为：

1、物业化模式：政府通过服务外包的形式交给具有一定资质的物业公司参与城市管理的社会治理方式，本质上呈现的是政府、市场与社会的角色重构，由深圳宝安区西乡街道首开先河。

2、进社区模式：以基层社区为基点，以创建整洁优美的城市环境为目的，整合社区力量，探索城市治理共治、共建、共享新格局，如深圳的 721 工作法、青岛的 1122 工作法等。

3、律师驻队模式：政府通过购买服务的方式与律师事务所签订法律服务协议，指派专职律师进驻城市管理部门，由深圳南山区沙河街道首创。

福建电信为客户提出第四种模式：**AI 智能监管模式**，以采集多元化、流程极简、运行智能化为目标，实现城市事件智能识别，充分利用网格员拍照、视频监控布控、物联网终端等方式，运用电信自研团队的 AI 能力，实现城市精细化、常态化管理；在应用上结合事件证据链、网格派单、知识图谱能力，实现事件智能分发派单，提升效率；重构数据中台，实现市一级的一网统管平台的数据入湖，构建城市管理联动指挥体系。系统架构图如下：



基于智算引擎，福建电信为客户提供机动车违停、非机动车乱停放、行人乱丢垃圾、违规占道摆

摊、违规占道晾晒、违规张贴广告等 AI 监管的城市治理应用，见下图：



3.2 AI 应用平台的开放性

要能支持千行百业的客户定制化的需求，基于城市治理应用，打造了开放服务体系的 AI 应用平台，包括：

开放环境：部署了 4 台训练机，2 台推理机搭建了算法样本仓，用于算法模型训练；

开放资源：提供知识库及数据，包括全省 9 地市架设了 20 路模拟环境的视频及标注数据等；

开放算力：在生产环境提供华为 atlas 芯片的算力资源和网络资源；

开放能力：提供视觉、语音、NLP 等通用领域 AI 能力；

开放应用模板：提供城管慧眼应用的基础模板，可二次开发定制组件，加载到页面上；

开放测评：提供测评系统，并通过人工筛查、系统计算的方式给出算法精度。

这套体系无疑是具有竞争优势的，体现如下：

电信运营商的资源整合优势：包括 5G+光网络的最后一百米的网络覆盖，触达十里八乡的装维队伍等等，都是无可比拟的。

属地化研发的优势：算法的泛场景化高可用性基于海量数据的训练，而在城市治理领域里，仅靠部分场景来训练算法，是很难提升其泛化性。此外，视频传输极占带宽资源，没有属地化研发能力的产品很难完成任务。

可定制的解决方案优势：AI 产品只有转换成应用后，才能为客户带来价值。在人车物识别的标准组件的基础上进行逻辑加工，满足客户的定制需求，才具备竞争优势。

3.3 创新推动发展

至 2023 年，福建电信已成功打造城管慧眼、智慧城中村、明厨亮灶三大标品，均已获得客户的高度认同并带动了视联网定制版业务的发展。AI 应用平台充分利用企业现有的能力，具备**系统稳定可靠、经济原则、标准性、共建共享原则、易操作原则**等先进的特性，当前端有新的需求出现时，后端可以用自动化的工具迅速迭代实现功能。目前，智算引擎已实现 100+算法的部署和使用，有力的支撑了省内智能视频业务的发展。在技术上也紧随业内头部企业的步伐，实现以下几个创新能力：

告警证据链：支持将同一摄像头下的所有告警信息进行事件归档，通过目标特征和告警坐标等多维度进行研判，生成多条告警证据链，避免同一告警事件多次重复告警，提高数据价值，并为事件追踪回溯提供有力支持。

高效抽帧：通过对视频流的高效抽帧技术（CPU+GPU 抽帧方式的结合），调用自研的人像模型，结合人脸特征判别规则，将检测框标在图片中，再拼合图片成视频流，通过流媒体服务器输出。

算力分级调度：支持根据任务的不同重要等级，对算力资源进行不同通道划分，针对一类任务，建立算力高架专道，保证任务高效运行；针对二类任务，建立算力共享资源池，充分利用资源的同时保障业务正常运行；针对三类任务，数据允许 T+1 告警。

坚持基于客户的行业属性不断创新，基于 AI 应用平台，为客户提供大屏可定制、告警可定制、处置可定制等灵动界面，助力业务发展。

4.运营体系

客户认同的 AI 应用，必定是 AI 能有效促进客户实际生产与管理的效率。因此，AI 是否好不好是最重要的一个指标。在运营体系构建中，分为装维体系、平台运营、业务运营三条支线同步推进，相辅相成。

4.1 开通、交付、维护纳入装维体系

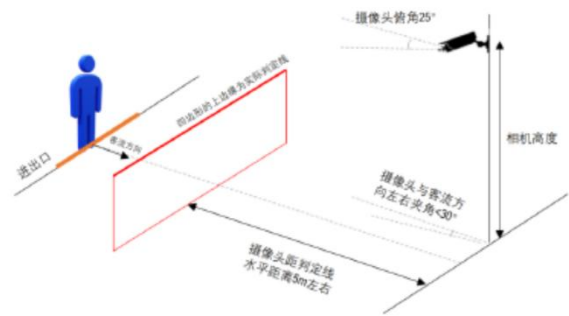
视频 AI 业务还未达到即装即用这么高的智能化，比如 AI 的效果与摄像头的安装、AI 识别区域的设置、AI 有效运行时间的设定等有很强的相关性。以客流统计为例，算法定义是在有效区域内，识别到人的进出。但如果需要算法给出满足期望的结果，至少要在**摄像头安装和可识别区域设置**两个方面按标准操作。

摄像头安装：

优先选择**狭窄通道**或**出入口**等人流方向单一的位置为目标区域，大密度或人流方向复杂，会由于相互遮挡导致误差增大；

安装高度 2.5-4.0 米，俯角推荐 20° 至 30°，摄像头距离判定线的 5m 左右，如高度低于 2.5 米人员密集时容易造成遮挡，影响检测精度；

安装角度正对客流方向，摄像头和客流方向左右夹角小于 30°。



可识别区域设置：

例 1：



红色线框问题分析：上边缘判定线和闸机虽然有夹角，但是因为此处空间不存在横向行人，所以也能满足闸机进出人员的识别统计。但是上边缘线过长，可能会误识别花瓶右侧的行人。正确为：蓝色画线。

上边缘判定线如果水平绘制，但因为相机和门有一定角度，导致判定线和门面（马路）有一定夹角，门口行人会误触线导致被误统计。正确为：黄色画线。

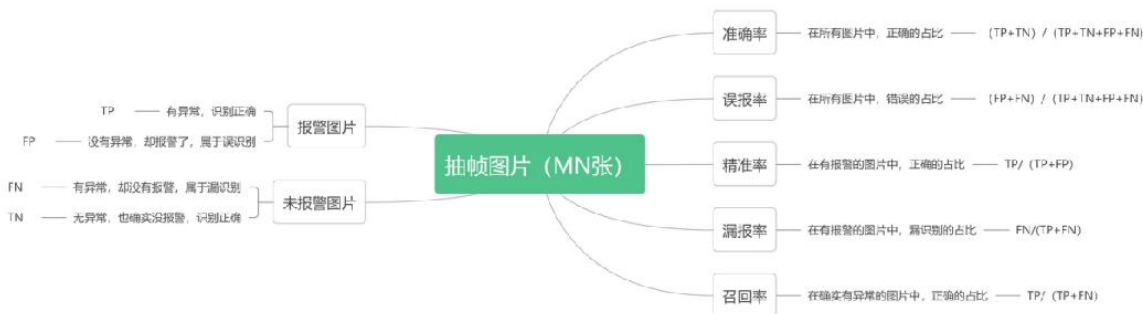
因此，AI 项目的开通、交付、维护纳入装维体系，才能持续提升客户对 AI 的满意度。



4.2 平台支撑体系

一个算法要达到可商用，依赖于海量的数据来训练模型，但自研算法过程中，最缺的就是可信的、泛化场景的数据，平台支撑体系主要从客户场景化的角度，从算法接入、告警回流、模型训练三个方面实现算法可运营体系，具体为：

在项目实景中，对算法进行全方面的测评，为项目客户筛选可靠、性价比高的算法；



通过应用反馈、抽样审查等，建立标准数据集管理；

对项目中回流的告警数据进行标注，再导入算法训练模型，提升特定场景下的算法准确率。

建立算法成熟度指标，当算法的准确率 95% 以上，召回率 90% 以上，定义为 L4，为可交付态。

4.3 业务运营体系

AI 分为项目和标品两类用途。其中，标品 AI 是通过 CRM 开通后可直接使用的 AI，如口罩识别，

在整个画面中，识别到有未戴口罩的人后，即给出告警图，这类 AI，只要摄像头像素足够高，不需要设置，基本都能正确捕捉。不同的是项目 AI，由于客户的需求不同，同一个算法在不同场景下，效果也相差很大。具体运营策略如下：

云边协同：边端采用低廉的国产 CPU 对视频流进行抽帧及目标检测，云端采用高端 GPU 对边缘上传的图片流进行识别解析，降低边到云的带宽消耗、提升云侧的解析效率；

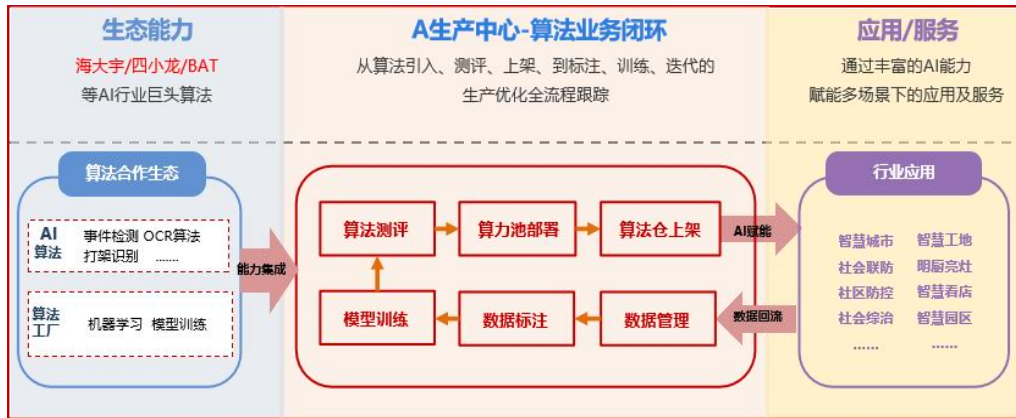
能力编排：拉取一路视频流，通过视频抽帧能

力+多图片能力的编排，实现一个设备赋能多种 AI 能力，降低流媒体资源消耗、利用图片 AI 代替视频 AI，降低算法软件成本；

算力集约化：通过算力容器化技术，实现对算力资源的集约化管理，将原本独立的服务器进行资

源池化，适配多种 GPU 型号，结合资源配置模版，快速生成 AI 算力单元；

算力动态扩容：在资源池的基础上，实现算力集群负载，根据业务需求，具备高可用及集群负载能力，保障业务的快速扩容及稳定性；



5 结束语

人工智能的解决方案和应用主要在语音识别、自然语言处理、图像识别三个方面，福建电信已经在图像识别领域上做了大量的产品规划、市场分析、应用落地的工作。省内 AI 发展战略也将更深度的融入到集团的视联网发展规划中，在控制单元上完成能力与业务分拆，快速定制出千人千面的 AI 应用，

为主航道提供轻交付的 AI 产品能力。未来也在多模态大模型上继续开展研究，例如，通过整合图像、文字、音频、传感器等内容，实现更加准确和全面的识别和检测结果。

参考文献：

人工智能白皮书(2022年)中国信息通信研究院

基于企业 MEC 部署视频算力网络边缘服务的方案研究

詹丰源 骆晓霞 林启笔 张浩宇 林泽剑 游思萍
(中国移动通信集团设计院有限公司福建分公司)

摘要: 随着 5G 时代到来,使得各行各业网络具有增强型移动宽带、大规模物联网、高可靠低时延的特征,对云网络未来有着更高的展望。作为支撑 5G 时代典型应用场景的关键技术——MEC 极大地提升了云数据的安全,降低了应用中的延时。随着运营商 5G 覆盖技术的建设,MEC 技术成为开发网络市场拓展各项业务的敲门砖,在这个技术应用下强化了各产业的融合性,指引产业向着 5G 进军。本文根据当前福建省内某运营商的行业视频中台能力及 MEC 应用试点基础,提出了基于企业 MEC 构建视频算力网络边缘服务的方案,充分发挥算网资源协同优势,满足企业视频监控以及其他融合应用场景的需求。

关键词: MEC, 多接入边缘计算, 视频算力网络。

引言

本方案研究在企业 MEC 上部署视频算力网络边缘服务,通过视频算力网络管理中心,实现视频监控服务的自动部署、企业账户及终端账户的在线快速开通、企业视频设备的快速接入等操作,并通过算网资源协同优势提供算力资源保障服务,同时还可以兼顾满足企业的融合场景应用需求。

1. 研究背景

1.1 视频通信市场规模现状

随着通信基础与云基础设施的成熟,以视频为代表的富媒体信息成为主流,视频与其他业态的融合发展将进一步加速,“视频+”跨界融合正当时。作为一种基础设施的视频将在多元化应用场景中迎来生态爆发,视频正在成为数字经济时代的新型基础设施,我国视频通信市场保持着快速发展的趋势。根据统计,2020 年我国视频通信行业市场规模达到 1021 亿元,同比增长 24.3%。近三年,中国视频通信市场将以年均复合增长率 20.75% 的速度持续增长,预计到 2022 年市场规模可以达到 1430.30 亿元。



图 1 中国视频通信市场规模预期 (亿元)
(注明来源)

目前,集传送语音、数据、视频于一体的视频通信行业成为通信领域发展的热点,以点到点或多点视音频通信为主要形式的视频监控、视频会议、远程医疗、远程教育等服务得到越来越多的使用,特别是视频融合的需求也越来越多。

1.2 MEC (多接入边缘计算) 的发展趋势

MEC(Multi-access Edge Computing)是 ETSI 标准组织提出的概念,即多接入边缘计算,是一种在相比中心 DC(Data Center) 更靠近终端用户的边缘位置提供用户所需服务和云端计算功能的网络架

构,通过业务靠近用户处理,以及应用、内容与网络的协同,为客户提供可靠、极致的业务体验。ETSI 定义了 MEC 应用的七大场景,包括企业分流、视频优化、视频流分析、增强现实、辅助敏感计算、物联网、车联网。

MEC 经过多年发展,国内三大运营商相继完成 MEC 试点工作,发布了 MEC 技术白皮书等,同时积极引进生态合作伙伴,共同推进商业应用落地。目前三大运营商以企业园区/专网场景应用为主,满足企业本地分流、算力卸载、能力开放的需求。

2. 视频算力网络边缘服务部署思路

本方案基于国内某运营商现有行业视频中台(大区控制节点+省级节点架构),以及企业园区 MEC 新增的边缘媒体节点构建大区控制节点+省级节点+企业 MEC 边缘节点组成的视频算力网络。实现企业视频监控就近接入视频算力网络的企业 MEC 边缘媒体节点,将不同制式、不同来源的多个视频资源统一整合并接入单一平台进行管理、调度、运营及其他应用操作,为行企客户提供多种视频服务,以及融合类视频业务应用,充分发挥算网融合协同优势,助力相关企业的数字转型。

视频算力网络边缘服务部署思路:

(1) 在福建省内企业 MEC 部署视频算力网络的边缘媒体节点,包括视频分发、存储和转码等视频能力,以及云资源与存储等基础算力设施,实现企业流量减少迂回,不出园区,媒体就近分发及存储;

(2) 基于视频算力网络的算力感知与自动编排能力,实现边缘媒体节点的自动编排部署,自动按需弹缩,充分发挥视频算力网络的一体编排调度优势,实现精准资源利用;

(3) 省内企业利用视频算力网络,实现业务协同与融合应用。通过算力网络资源的调度和就近部署,将通用的音视频能力服务化,并通过云网边缘融合,提供增强型视频融合服务,高效资源利用,统一极简的运维运营,实现一张网络同时承载和协同多种业务。

融合类业务主要可关注如下应用:

➤ 企业办公协同与智慧工厂远程监测应用的融合:在企业的智慧工厂中,可在车间内部署高清摄像头,用于远程监测车间内的机器运行情况 & 仪表盘数据等,尤其适用于某些具有危险性质的工业操作中。通过高清摄像头采集的视频,可就近接入到视频算力网络的企业园区 MEC 边缘媒体节点中,由该边缘媒体节点直接转发到企业内的监测端进行监测。同时,对于企业内需要多人、多部门协同的工作,可将企业内会议室的会议硬终端、各部门工作人员手机上的会议软终端,通过接入到视频算力网络的企业园区 MEC 边缘媒体节点,组建多方视频会议进行多人、多部门的协同,并可以在会议中拉入智慧工厂远程监测的视频画面,用于辅助会议协商,协同指挥操作等。

➤ 企业工作人员对讲与会议应用的融合:在企业中某些涉及多人协作的工作,由于需要日常性质的实时通报各方的情况,一般会通过 5G 对讲机,在各方之间通过对讲的方式进行实时通报。此时,5G 对讲机采集的语音流,可就近接入到视频算力网络的企业园区 MEC 边缘媒体节点中,由该边缘媒体节点直接转发给同一对讲频道的其他各方。但是,在进行到某些重要的操作节点时,各方通过半双工的对讲的方式已无法进行高效的沟通,此时可将终端与视频算力网络 MEC 边缘媒体节点之间半双工的对讲通讯方式,直接切换为全双工的实时音视频通讯方式,在业务层面将对讲的应用模式直接切换为会议的应用模式,通过视频算力网络在各方之间组建音视频会议,提高沟通效率。会议沟通清楚后,可将终端与视频算力网络 MEC 边缘媒体节点之间的通讯方式切换回对讲的半双工通讯方式,继续用于日常性的信息实时通报。

3. 技术方案研究与探讨

3.1 业务功能架构设计

基于国内某运营商现有行业视频中台(大区控制节点+省级节点两级架构),以及企业园区 MEC 新增的边缘媒体节点构建大区控制节点+省级节点+企业 MEC 边缘节点组成的视频算力网络,视频算力网络业务功能架构图如下所示:

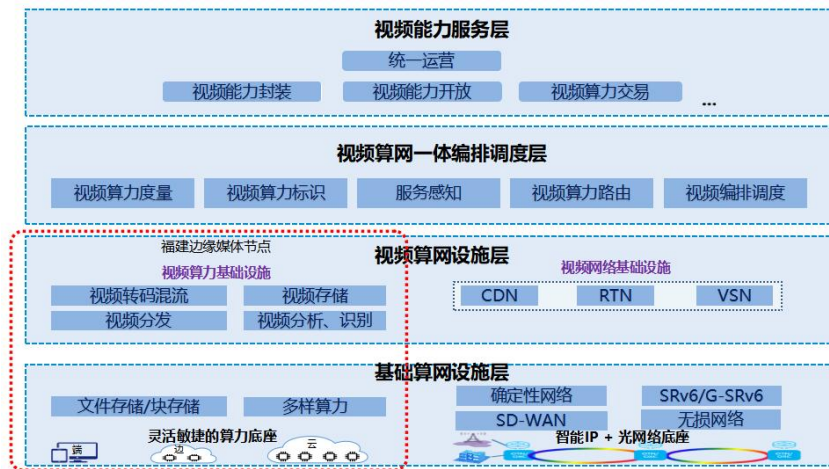


图 2 边缘媒体节点业务功能架构图

(1) 视频算力网络的能力服务、一体编排调度层：在全国大区控制节点；

(2) 福建省级媒体节点、企业 MEC 边缘媒体节点：主要包括视频分发、视频存储、视频转码混流等视频算力设施，以及底层的云资源算力、存储算力等基础算力设施；

(3) 由全国大区控制节点对福建省级媒体节点、企业 MEC 边缘媒体节点进行统一调度控制，并与其他省份的边缘媒体节点一起，共同构建全国组网的视频算力网络。

3.2 网络架构设计

视频算力网络组网方案如下图所示：

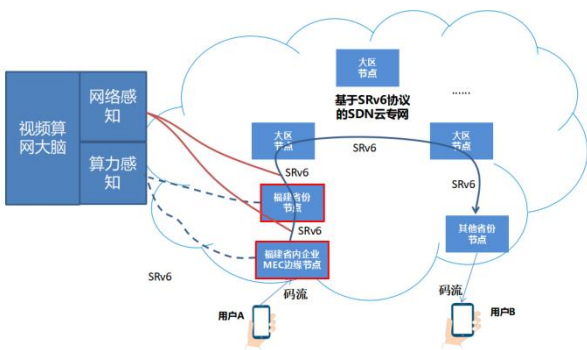


图 3 试点组网方案示意图

(1) 福建省级媒体节点、企业 MEC 边缘媒体节点用于提供媒体转发、转码、混流、剪辑等视频算力；

(2) 视频算力网络中福建省级媒体节点和企

业 MEC 边缘媒体节点之间，福建省级媒体节点与大区控制节点之间，福建省级媒体节点与其他省份省级媒体节点之间，均通过基于 SRv6 协议的 SDN 云专网实现互联互通。通过 SRv6 协议，依据视频算力网络上每个媒体节点之间的网络感知，实现媒体传输路径的智能路由规划，满足不同业务、不同用户的需求。

3.3 网络接入方案

在福建省内试点企业的园区内，部署 5G MEC 边缘云，并基于该 MEC 边缘云的算力资源，部署视频算力网络 MEC 边缘媒体节点，实现该试点企业内部媒体流的就近接入、就近处理、就近存储，并且对于某些业务，可以做到媒体数据不出园区。如下图所示：

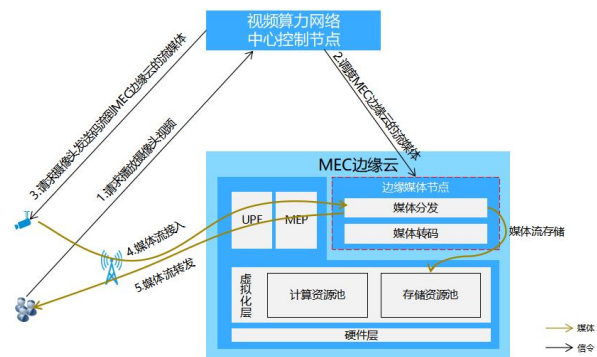


图 4 网络接入方案示意图

首先，通过视频算力网络全国大区控制节点，

将媒体流接入点调度到 MEC 边缘媒体节点上，同时依靠边缘 UPF 的路由转发,实现媒体流直接发送到 MEC 边缘内部的媒体处理相关服务，实现媒体流的就近接入、就近处理、就近转发，降低媒体处理的端到端时延，提升用户体验，同时降低媒体数据对骨干网络带宽的占用。

然后，利用 MEC 边缘云提供的存储算力，实现媒体流在本 MEC 边缘内部进行存储，尽可能将占用带宽较高的媒体数据流量在 MEC 边缘内终结，同时达到媒体数据不出园区的目的。

此外，在福建省中心部署视频算力网络福建省

级媒体节点，用于接入和处理福建省内、试点企业以外的其他端的媒体数据。

3.4 云边协同方案

视频算力网络全国大区控制节点与福建省级媒体节点、企业 MEC 边缘媒体节点之间：全国大区控制节点主要提供视频算力感知、视频网络感知、视频算力编排、视频算力调度等能力，而福建省级媒体节点、企业 MEC 边缘媒体节点则主要在大区控制节点的控制调度下，提供媒体处理的原子视频算力。大区控制节点和边缘媒体节点之间的云边协同如下图所示：

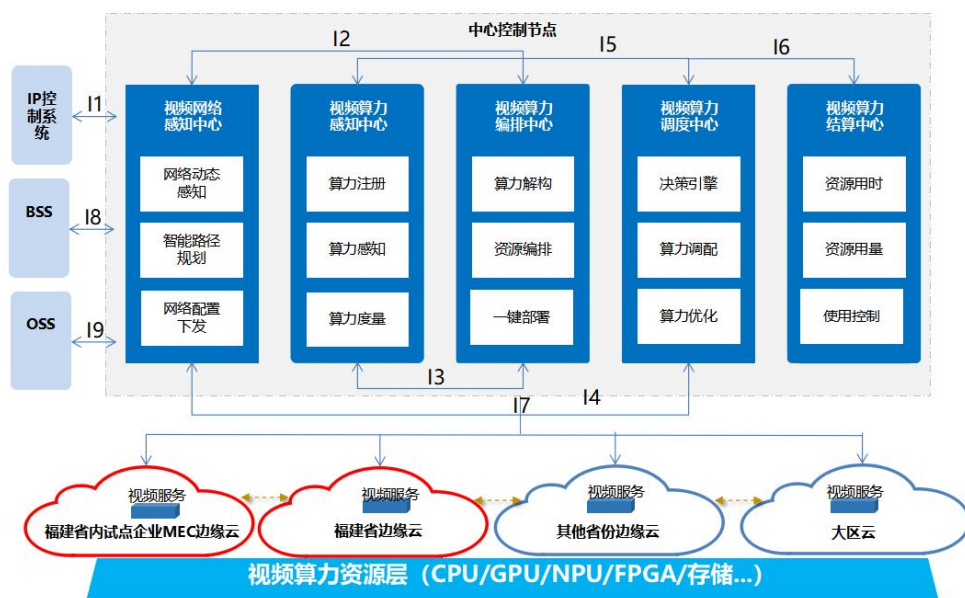


图 5 中心控制节点和边缘媒体节点之间的云边协同示意图

福建省级媒体节点和企业 MEC 边缘媒体节点之间：福建省级媒体节点可以为试点企业 MEC 边缘媒体节点提供媒体处理的容灾备份能力，当试点企业 MEC 边缘媒体节点出现故障或能力已用满时，可以在中心控制节点的调度下，将媒体相关业务调度到福建省级媒体节点上去处理。

福建省级媒体节点与大区边缘媒体节点之间：大区边缘媒体节点可以为福建省级媒体节点提供媒体处理的容灾备份能力，当福建省级媒体节点出现故障或能力已用满时，可以在中心控制节点的调度下，将媒体相关业务调度到大区边缘媒体节点上去处理。

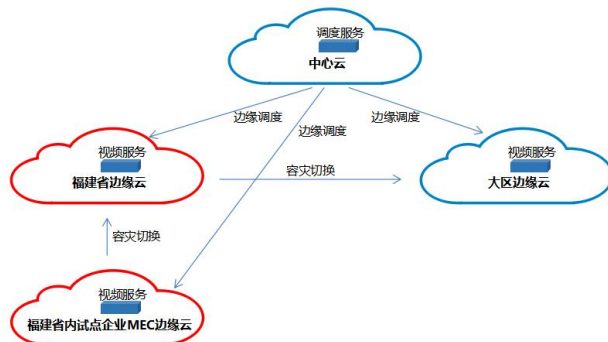


图 6 福建省边缘媒体节点云边协同示意图

3.5 算网大脑编排方案

福建省级媒体节点和企业 MEC 边缘媒体节点

需要支持通过算网大脑的编排能力，实现其自动化编排部署，从而满足视频算力网络福建边缘的自动弹性伸缩、边缘快速自动化开通等需求。

首先，福建省级媒体节点和企业 MEC 边缘媒体节点的视频微服务需要是基于云原生架构，通过算网编排实现异构云的统一自动化部署；

第二，福建省级媒体节点和企业 MEC 边缘媒体节点的异构云资源算力也均支持 K8S 原生框架，支持算网编排的统一调度；

第三，通过基础算力网络的算力感知，自动获取到福建省级媒体节点和企业 MEC 边缘媒体节点的云资源算力，然后通过算力网络的统一编排能力，将边缘媒体节点自动编排部署到边缘侧的云资源上。

视频算力网络边缘媒体节点的自动编排部署的主要流程如下图所示：

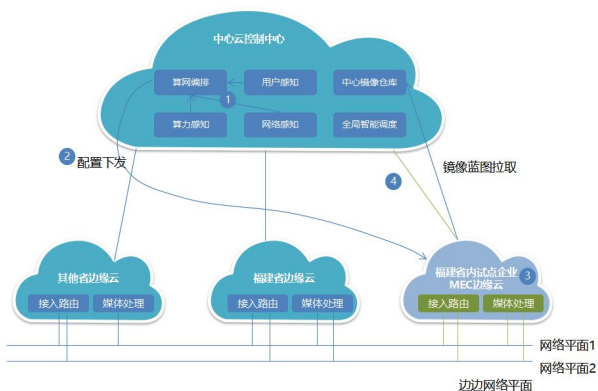


图 7 视频算力网络边缘媒体节点自动编排部署示意图

- (1) 基于用户感知，算力感知和网络感知发现视频算力网络瓶颈，找到需要新增的边缘云节点；
- (2) 通过算网编排中心选定合适的视频微服务及网络平面配置，向目标边缘云下发配置；
- (3) 目标边缘云根据配置信息从中心镜像仓库拉取镜像并自动化完成部署和网络平面配置；
- (4) 新增边缘云媒体微服务接受中心云的智能调度，分流视频业务。

3.6 资源配置方案

1、算力需求测算

以一个企业园区为例，部署一个 MEC 边缘侧

的边缘媒体节点，按照媒体处理的并发量计算所需算力资源，按照边缘媒体节点并发处理 200 路媒体流转发/分发、200 路媒体流存储（存储 30 天）、50 路媒体流转码/混流等处理，且每路媒体流平均带宽 2Mbps 计算。所有服务配置均按照 N+1 冗余配置。

➤ 底层 PaaS 需求：需要 3 台 32C/128G 内存的云主机部署 PaaS 控制节点，需要 2 台 32C/128G 内存的云主机部署负载均衡 SLB，需要 3 台 32C/128G 内存的云主机部署公共服务，即总共需要 3+2+3=8 个 32C/128G 内存的云主机算力资源；

➤ 实时媒体分发服务：每个 32C/128G 内存的实时媒体分发服务按照支持并发处理 450 路 2Mbps 的媒体流转发/分发计算，因此，总共需要的实时媒体分发服务的个数= $\text{ROUNDUP}(200/450)+1=2$ 个，即总共需要 2 个 32C/128G 内存的云主机算力资源；

➤ 媒体转码服务需求：每个 32C/128G 内存的媒体转码服务按照支持并发处理 32 路 2Mbps 的媒体流转码、混流等操作计算，因此，总共需要的媒体转码服务的个数= $\text{ROUNDUP}(50/32)+1=3$ 个，即总共需要 3 个 32C/128G 内存的云主机算力资源；

➤ vSBC 服务需求：每个 32C/128G 内存的 vSBC 服务按照支持并发处理 150 路终端接入，因此，总共需要的 vSBC 服务的个数= $\text{ROUNDUP}(200/150)+1=3$ 个，即总共需要 3 个 32C/128G 内存的云主机算力资源；

➤ 流媒体服务需求：每个 32C/128G 内存的流媒体服务按照支持并发处理 1000 路 2Mbps 的媒体流存储计算，因此，总共需要的流媒体服务的个数= $\text{ROUNDUP}(200/1000)+1=2$ 个，即总共需要 2 个 32C/128G 内存的云主机算力资源；

➤ 存储空间：总共需要的存储空间= $200 \text{ 路} * 2\text{Mbps} * 3600 \text{ 秒} * 24 \text{ 小时} * 30 \text{ 天} / 8(\text{b/B}) / 1024\text{M} / 1024\text{G} = 124\text{TB}$ 。

➤ 按照以上计算结果，共需要的云主机算力资源为 8+2+3+3+2=18 个 32C128G 的云主机算力资源，需要的存储空间为 124TB 空间。

2、网络需求测算

视频算力网络的中心控制节点与每个边缘媒体节点之间、每两个边缘媒体节点互相之间，均需要通过基于 SRv6 协议的 SDN 云专网互联互通。

边缘媒体节点的网络互联需要的带宽=并发
200 路媒体流转发/分发*每路媒体流带宽
2Mbps=400Mbps。

4. 结语

本文根据 MEC 现阶段发展与福建省内业务情况与基础, 提供了一种通过 MEC 建设朝着视频算力网络演进的方案。现阶段 MEC 虽然还存在较多的问题需要解决, 但在物联网、视频、大数据和人工智能等产业的发展还比较先进, 在这些数以万计的接入局所和边缘机房电信运营商中构建出的 MEC 边缘计算还是具有优势的, 并迎来其较大机遇。

总之, MEC 技术的应用为各行各业网络服务端带来巨大变革, 获得相关行业的关注。在商业发

展领域中可以结合 MEC 的相关应用, 多元化利用网络运营服务模式, 对 MEC 技术有更长远的发展部署。

参考文献

- [1] 范桂颀, 5G 与 MEC 边缘计算 [OL]. 博客园, 2019.09.24 <https://www.cnblogs.com/jmilkfan-fanguiju/p/11825026.html>
- [2] 何春霞, 陈刚. 多接入边缘计算(MEC)技术及业务发展策略[J]. 科技与创新, 2023(1):65-67. DOI:10.15913/j.cnki.kjycx.2023.01.018.
- [3] 栗蔚, 徐恩庆, 董恩然, 等. 云计算与边缘计算协同九大应用场景 [J], 云计算开源产业联盟, 2019 (07) .

基于信息间隙的虚拟能源枢纽 工厂响应模型的多目标优化

范林元 林耿 刘鲤晞

(闽江学院数学与数据科学学院(软件学院))

(中国福建省力得自动化设备有限公司)

摘要:随着多载体能源系统在当地能源市场(如热力市场)的出现,虚拟能源枢纽工厂在许多国家变得更加突出,以实现可持续能源供应网络。以往对虚拟能源枢纽工厂技术经济运行的研究没有考虑到在氢能市场推广 Power-to-X (P2X) 转换设施,以综合方式评估能源效率、温室气体排放和经济问题。为此,该文提出一种规避风险的多目标方法来解决虚拟能源枢纽工厂的自调度问题并解决提到的问题。所考虑的虚拟能源枢纽工厂基于不同的能源载体(即电力、供暖、制冷和氢气)、P2X 机组、电动汽车和发达的需求响应(DR)策略来运营。虚拟能源中心工厂运营商试图通过参与当地能源市场并考虑到不同的不确定来源来最小化二氧化碳排放率,从而最大限度地提高收入。热电联产(CHP)机组和 DR 策略被用作可再生能源输出不稳定的实用解决方案,并利用各种当地能源市场的机会。为了对所提出的策略进行建模,开发了一种非概率信息差距决策理论(IGDT)。将所提模型应用于样本测试用例,数值结果表明,考虑到最新的储能系统、发达的灾难恢复策略和 P2X 机组,虚拟能源枢纽工厂的利润提高了 17.54%,CO₂ 排放率降低了 10.24%。

简介

最近,不同能源载体的整合以及分布式能源发电的日益普及,导致了当地电网运行更加高效。考虑到这一点,能源枢纽系统的概念作为接口已被当地电网和其他能源基础设施广泛欢迎[1]。在技术经济问题领域^[2],使用综合能源服务提供商(即能源枢纽系统)满足多能源消费者的各种需求,将使当地的能源系统变得灵活,为系统运营商带来更多经济机会^[3]。能源枢纽系统运营商通过采取规避风险的策略来克服可再生能源等不确定来源的不利风险,从而强烈倾向于参与当地能源市场交易发电和能源消耗^[4]。风险规避意味着能源枢纽系统运营商倾向于基于以下因素的调度:低不确定性结果到高不确定性结果,即使第二种情况采用的调度结果的平均值在货币金额方面等于或高于更确定的结果。考虑的风险厌恶程度取决于然而,系统运营商有必

要使用稳健的方法来减轻不确定来源对能源枢纽系统自我调度的影响,以获得可靠的结果^[5]。在存在这种不可避免的不确定性的情况下,能源枢纽系统调度中考虑的经济指标(例如利润率或成本率)的改善受到多载体能源系统中各种设施和服务的参与的强烈影响。然而,以前关于能源枢纽系统短期自调度的研究没有考虑多载体能源转换设施的独特有效性、当地能源市场的现有机会以及能源生产消费者,例如插电式混合动力电动汽车(PHEV),以实现风险规避方法。使用多载波储能系统(ESSs)和功率转 X (P2X) 单元可以帮助能源枢纽系统运营商消除实现既定目标的技术和经济障碍。这些设施为能源中心系统提供更优化、更灵活的运行以及参与不同能源市场的能源枢纽系统运营商满足所有所需的能源需求^[6]。这虚拟能源中心工厂的理念是基于公共能源中心架构的概念,通过优化地自我调度各种发电、转换和存储单元以及管理多能源消

费者的行为，让所有现有的本地市场参与进来，从而获得更多收入。考虑到上述情况，特别是灵活性多运营商能源网络，关注虚拟能源枢纽工厂及其在灵活资源下的稳健性能，在风险规避优化方法下捕

获不确定性可再生能源发电是未来多载体能源系统的必需品之一^[7]。因此，必须解决的主要研究课题之一是制定可靠的策略，以满足多能源消费者和其他实体的需求，例如以综合方式满足 PHEV。

Nomenclature	
Abbreviations	
CES	Cooling energy storage
CHP	Combined heat and power
DR	Demand response
EES	Electrical energy storage
ESSs	Energy storage systems
HES	Hydrogen energy storage
IGDT	Information gap decision theory
P2C	Power-to-cooling
P2H	Power-to-heating
P2HY	Power-to-hydrogen
PHEVs	Plug-in hybrid electric vehicles
P2X	Power-to-X
RESs	Renewable energy sources
TES	Thermal energy storage
Indexes	
t	Index for time intervals
n	Index for CHP units corner operation points
m	Index for EVs
e	Index for type of EVs
Parameters	
$ID_t^e, ID_t^h, ID_t^c, ID_t^p$	Initial electrical, heating, cooling, and hydrogen demands [MW, MWt, MWt, kg]
N^{arr}, N^{dep}	Number of EVs arrived/departure to/from parking
$z_t^e, z_t^h, z_t^c, z_t^p$	Electric power, heat, and hydrogen purchasing price from energy markets [\$/MWh, \$/MWt, \$/MWt, \$/MWh]
$\pi_t^e, \pi_t^h, \pi_t^c, \pi_t^p$	Electric power, heat, hydrogen, cooling selling price to the end-users [\$/MWh, \$/MWt, \$/MWt, \$/MWh]
$\mu^{ees}, \mu^{tes}, \mu^{ces}, \mu^{hes}$	Operation cost of EES, TES, CES, and HES [\$/MW, \$/MWt, \$/MWt, \$/MW]
$\mu^{edr}, \mu^{tdr}, \mu^{cdr}$	Rate of incentive for electrical, thermal, and cooling demands variation
$\zeta^e, \zeta^h, \zeta^c$	Emission coefficient for electric and heat purchasing from
energy markets and CHP unit	
$\eta^{P2C}, \eta^{P2H}, \eta^{P2HY}$	P2X units efficiency
$\alpha^e, \alpha^h, \alpha^c$	Coefficient of participating electrical, heating, and cooling demands in integrated DR program
Variables	
P_t^{sell}, P_t^{buy}	Power sold/purchased to/from the electricity market [MW]
H_t^{buy}	Heating purchased from the district heating market [MWt]
HY_t^{buy}	Hydrogen purchased from the hydrogen market [kg]
$H_t^{es,db}, H_t^{es,ch}$	Discharging/charging amount by TES [MWt]
$C_t^{es,db}, C_t^{es,ch}$	Discharging/charging amount by CES [MWt]
$HY_t^{hes,db}, HY_t^{hes,ch}$	Discharging/charging amount by HES [MW]
$P_t^{ees,db}, P_t^{ees,ch}$	Discharging/charging amount by EES [MW]
$P_t^{ev,db}, P_t^{ev,ch}$	Discharging/charging amount by EVs [MW]
$D_t^{e,dr}, D_t^{h,dr}, D_t^{c,dr}$	Electrical, heating, and cooling demands after implementing integrated DR program [MW, MWt, MWt]
$D_t^{e,up}, D_t^{e,dw}$	Up/down electrical demand response
$D_t^{h,up}, D_t^{h,dw}$	Up/down heating demand response
$D_t^{c,up}, D_t^{c,dw}$	Up/down cooling demand response
P_t^c, H_t^c	Power/heat produced by CHP units [MW, MWt]
S_t^e	Energy capacity of EES, TES, CES, HES [MWh]
C_t^{P2C}	Cooling produced by P2C units [MWt]
P_t^{P2C}	Power consumed by P2C units [MW]
H_t^{P2H}	Heat produced by P2H units [MWt]
P_t^{P2H}	Power consumed by P2H units [MW]
HY_t^{P2HY}	Hydrogen produced by P2HY units [kg]
P_t^{P2HY}	Power consumed by P2HY units [MW]
S_t^{EV}	Energy capacity of EVs [MWh]
$S_t^{EV,arr}, S_t^{EV,dep}$	Energy capacity of EVs at arriving and departing times [MWh]
$\omega_{n,j}$	Variable related to CHP corners' points
$\alpha_t^{(e,db)}, \alpha_t^{(e,ch)}$	Binary variable for EES, TES, CES, and HES operation

1.1.文献综述

文献中已经提出了各种适用的能源管理方法，以实现虚拟能源中心工厂的最佳自我调度。^[8]的作者提出了一个确定性模型能源枢纽系统的经济运

行，使用热电联产（CHP）装置、ESSs、锅炉和可再生能源（RESs）作为本地发电/存储/转换设备。此外，风力发电被认为是主要的不确定来源，因为研究。在 Ref. 中提出了一种随机方法。

Table 1 Summary of the proposed models.

Ref.	Trading energy with local markets			Supplied demands by VEHP				Integrated DR program	PHEVs	Objectives	Optimization strategy
	Electricity	Hydrogen	District heating	Electricity	Thermal	Cooling	Hydrogen				
[8-11]	✓	×	✓	✓	✓	×	×	×	×	Operation cost	Deterministic or stochastic optimization method
[13-16]	✓	×	✓	✓	✓	×	×	×	×	Operation cost	Stochastic or IGDT optimization method
[17]	✓	×	✓	✓	✓	✓	×	×	×	CO2 Emission-Operation cost	Deterministic multi-objective optimization method
[18-20]	✓	×	✓	✓	✓	✓	×	×	×	Operation cost	Stochastic or robust optimization method
[22,23]	✓	×	✓	✓	✓	✓	×	×	×	Operation cost	Stochastic optimization method
[25-27]	✓	×	✓	✓	✓	×	×	×	×	Operation cost	Stochastic or robust optimization method
[28]	✓	×	✓	✓	✓	×	×	×	×	CO2 Emission-Profit	Hybrid robust-stochastic multi-objective method
[29,30]	✓	×	✓	✓	✓	×	×	×	×	Operation cost	Stochastic optimization method
[32]	✓	×	×	✓	✓	×	×	×	✓	Operation cost	Stochastic optimization method
Our study	✓	✓	✓	✓	✓	✓	✓	✓	✓	CO2 Emission-Profit	IGDT-based robust multi-objective optimization

Note: VEHP-virtual energy hub plant.

^[9]通过依赖于条件风险价值方法实现能源枢纽系统的最佳调度。在参考文献^[10]中,已经研究了国内能源中心的优化模型,旨在最小化总运营成本和峰值需求。能源枢纽的经济调度问题已在参考文献^[11]中得到解决,其中建立多载波能源系统在线优化框架。^[12]的作者提出了住宅规模能源中心能源管理的优化模型,其中包括一个 CHP 装置、一个热能储存装置 (TES)、一个 PHEV 和一个屋顶光伏 (PV) 系统,增强配电灵活性客户侧的电力网络以及管理 RESs 的波动性。本研究采用两点估计法将光伏电池板的发电量作为不确定来源进行处理。为了研究电热能流的相互联系,参考文献^[13]分析了可再生能源能源中心的经济行为。该系统是作为一个独立的单元建立的,无法与上游电网进行电力交易。在参考文献^[14]中,提出了一种混合整数线性规划 (MILP) 策略,用于能源枢纽系统的能源管理,旨在最大限度地降低总运营成本和商业风险。在同一工作中,信息差距决策理论 (IGDT) 被用作风险管理模型,以达到针对 RES 不确定性的鲁棒调度模型。能源枢纽系统,在参考文献^[15]中考虑,一直专注于通过从当地电力和供暖市场购买能源以及使用当地设施 (即 ESSs, RESs, 电锅炉) 来满足系统的电气和热需求。上述系统的不确定性已通过纯随机优化规划进行管理。在参考文献^[16]中,已经开发了一种强大的能源管理策略来管理基于氢的虚拟能源中心的行为,考虑集成的 DR 计划和基于燃料电池的储氢系统,最大限度地降低总运营成本。

还有一些关于能源市场中冷却、热电联产 (CCHP) 系统的优化和运行的研究。在参考文献^[17]中,提出了多目标确定性优化,以实现基于 CCHP 的能源轮毂系统的优化运行,旨在最大限度地降低总成本和二氧化碳排放。在参考文献^[18]中,已经为基于 CCHP 的能源枢纽系统基础随机优化方法提出了一种规避风险的能源管理策略。参考文献^[19]中采用了鲁棒优化方法来评估三联产系统 (即 CCHP 系统) 的有效性,以满足多载波能源系统中的电力,加热和冷却需求。机会约束编程已在 Ref 中使用。^[20]在基于 CCHP 的能源枢纽系统的背景下捕捉可再生能源发电的不确定性。^[21]的作者使用 CCHP 系统和 RESs 开发了一个虚拟能源中心工厂,以降低虚拟能源中心工厂的运营成本,特别强调随机优化程序。参考文献^[22]中采用了基于数据驱动的基于集的

方法,以优化考虑 ESSs 的可再生 CCHP 系统的可访问性和运营成本。在经济限制下,参考文献^[23]中提出了下行风险约束方法,以控制能源市场中的现有风险,当 CCHP 系统中的能源枢纽系统运营商试图满足电力,冷却和供暖需求时。作者^[24]考虑了需求响应 (DR) 策略,RESs 和 ESSs 在使用两级协调解决方案实现能源枢纽系统最低成本调度策略方面的重要性。本研究的主要目标是减轻 RESs 的不可预测性,并降低日前能源市场的运营成本。在参考文献^[25]中,已经开发了一种纯随机优化方法,用于局部电力和供热能源市场中能源枢纽系统的最佳协调能源调度。在参考文献^[26]中,提出了一种稳健的优化方法,以在能源枢纽系统能源管理计划的框架内管理可再生能源的不确定性。同样,在参考文献^[27]中,已经使用了一种稳健的方法,为能源枢纽系统制定了一个日前排放意识计划,用于综合参与区域供热和电力市场。参考文献^[28]中采用了一种混合鲁棒随机方法,为能源枢纽系统推导出基于利润的方案,以在电力,区域供热和天然气市场进行交易。该研究的主要目标是减少对能源枢纽系统的整体运营损失,并通过依靠热电联产装置、储能系统和可再生能源来减少二氧化碳排放。虚拟能源枢纽工厂的市场参与方法,由几个单独的能源中心组成,已在参考文献^[29]中概述,以满足客户的电气和供暖需求。其中,虚拟能源枢纽工厂使用热电联产机组、锅炉、电能存储 (EES)、TES 和当地能源市场来满足能源需求。^[30]的作者利用当地能源市场,DR 计划和 TESs 提供的潜力来构建基于绿色的能源枢纽系统,以促进 RESs 集成并节省运营成本。在参考文献^[31]中,提出了一种全面的优化方法,用于增加虚拟能源枢纽工厂的利润,该工厂试图实现电力,加热和数据中心的冷却需求。参考文献^[32]提供了用于氢基虚拟能源中心优化运行的和谐搜索优化算法,以便系统运营商可以从当地电力和燃气市场以最佳方式购买所需的能源,以满足能源转换设备的需求,即热电联产机组、锅炉、储能系统。此外,IGDT 技术已在 Ref 中使用。^[33]通过参与区域供热和电力市场来优化虚拟能源枢纽工厂的利润,同时考虑到 RESs 的波动性。

由于 PHEV 在多载波能源系统中的渗透率正在提高,本文提出的策略试图利用 PHEV 的能力来实现为虚拟能源枢纽工厂运营商设定的目标。在这种

自调度问题中使用 PHEV 的技术原因是, 通过使用 PHEV 的存储容量, 首先可以增加集成能源系统中的非旋转储备能量, 其次可以增强集成首先可以增加集成能源系统中的非旋转储备能量, 其次可以增强集成能源系统对不确定来源的灵活性^[34]。在这方面, 参考文献^[35]中提出了一个日前能源控制策略, 用于 PHEV 和能源枢纽系统的集成运行, 从而降低运营成本。在参考文献^[36]中, 随机编程提出了考虑 PHEV 的虚拟能源枢纽工厂最优运行模式的方法。此外, 在参考文献^[37]中提出了两级决策模型, 用于在 PHEV 存在的情况下, 能源枢纽系统关于经济约束的鲁棒调度。

1.2. 研究差距和贡献

关于综述的文献, 在与虚拟能源枢纽工厂的调度和运营相关的现有研究中可以强调以下缺点。

- 能源枢纽系统作为虚拟能源枢纽工厂参与

氢市场以利用经济机会的能力被忽视了[8-37]。

- 缺乏在虚拟能源中心工厂框架内管理电力、供暖和制冷需求的综合灾难恢复策略, 以减少对能源市场的依赖并更多地利用消费者能力[8-37]。

- 未考虑在集成系统中使用电制氢 (P2HY)、电制制冷 (P2C) 和电制热(P2H)能够提高虚拟能源效率的方式能源中心工厂, 降低二氧化碳排放率, 并增加系统利润[8-37]。

- 在参考文献[8-16, 18-27, 29-37]中, 在与当地能源市场进行双边交流以减少二氧化碳排放的情况下, 尚未研究在虚拟能源中心工厂中使用 PHEV 的影响。

- 在文献[8 - 16, 18 - 27, 29 - 37]中, 经济环境挑战, 即虚拟能源中心工厂的利润和碳排放率, 已被解决为单一目标优化问题。同时, 应利用多目标优化问题使数值结果更接近实际条件。

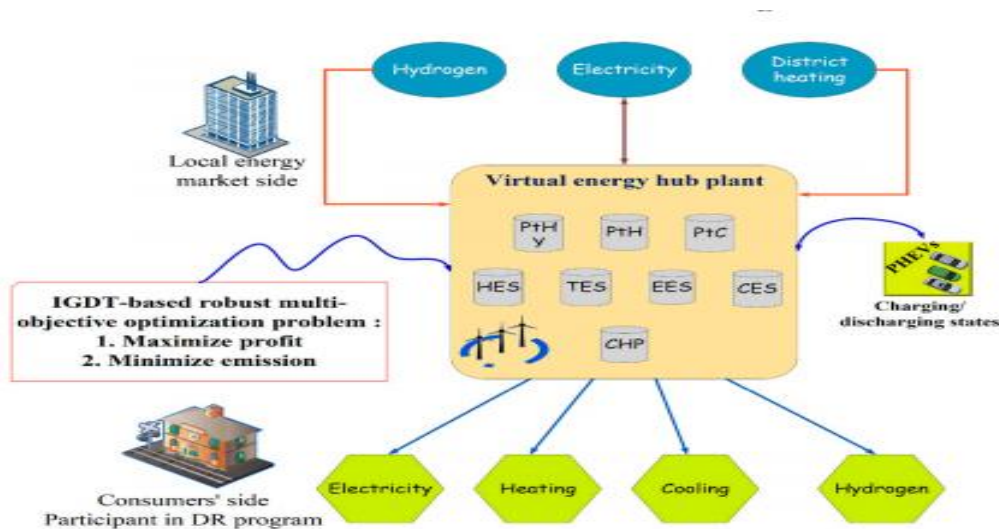


Fig. 1. A conceptual design of a sample virtual energy hub plant.

表 1 总结了虚拟能源枢纽工厂优化中提出的模型分类。为了克服上述研究差距, 本研究提出了一种规避风险的自我调度策略, 用于虚拟能源中心工厂, 以环保的方式与电力、氢气和区域供热市场合作。据作者所知, 没有研究考虑规避风险和稳健的收入自我调度虚拟能源中心工厂战略, 依托电转冷、电转氢等多维能源转换设施, 结合当地能源市场评估能源经济环境指标。特别是, 没有研究去评估 PHEV, 基于激励的多能源 DR 计划, PHEV, P2C 单元和冷却储能系统的情况下, 冷却负荷对与虚拟

能源中心工厂相关的温室气体排放的作用。为此, 开发了一个基于 IGDT 的鲁棒多目标优化问题, 以优化虚拟能源中心工厂的行为和满足电力、冷却、加热和氢气需求, 同时考虑基于激励的集成灾难恢复计划来管理最终用户的行为。开发的风险规避框架的主要目标是最大化虚拟能源中心考虑到可再生能源生产的固有不确定性, 工厂的利润和最大限度地减少碳污染排放。本文的主要贡献可以总结如下。

- 在当地电力、氢能和热能市场中交易虚拟能

源枢纽工厂，与 PHEV 通过基于多目标 MILP 的优化方法。

- 通过考虑氢市场对其他当地能源市场交易所的影响，并开发用于电力、热能、冷却和氢能生产的多样化平台，提高了虚拟能源枢纽工厂的灵活性水平。

- 在考虑综合灾难恢复程序以及基于 IGDT 的鲁棒方法的可再生能源发电预测值与实际值之间的误差的同时，实现了虚拟能源枢纽工厂的最大利润以及最低碳排放率。

- 现有的虚拟能源枢纽可以利用 P2X 单元的能力，即 P2H, P2C 和 P2HY, 作为环保能源转换设施，以利用可行的经济优势以及当地能源市场的环境机会。

1.3. 论文组织

本研究报告的其余部分组织如下。第 2 节介绍了虚拟能源枢纽工厂的架构，并介绍了用于虚拟能源中心工厂最优自调度的 IGDT 鲁棒优化模型的数学模型。数字结果在第 3 节中表示和讨论。最后，结论在第 4 节中给出。

2. 系统结构和问题表述

虚拟能源中心工厂的结构如图 1 所示，由 P2C, P2H, P2HY, EES, TES, 冷却储能 (CES), 氢储能 (HES), CHP, 风力涡轮机和 PHEV 组成，以满足考虑综合 DR 计划的电力，加热，冷却和氢气需求。目前，电力、供暖和制冷需求占多载波能源系统能耗的很大一部分[9]。目前，电力、供暖和制冷需求占能源消耗的很大比例多载体能量系统[9]。根据强有力的证据[16]，在不久的将来，氢需求将成为多载体能源系统最终用户的关键需求之一需要。因此，所提出的策略试图通过以虚拟能源枢纽系统的形式同时满足能源系统运营商的一些需求来解决能源系统运营商的一些技术-经济-环境挑战。拟议的虚拟能源中心工厂由当地供应供暖和氢能市场。此外，该工厂还可以进行电力交易，即在当地电力市场上进行购买和销售。换句话说，虚拟能源枢纽工厂运营商具有双向能源一方面与

电力市场进行交易，另一方面运营商可以只从其他市场购买能源。PP2HY 装置产生的氢能可以在 HES 内长时间保存，也可以用于满足氢气需求。这种循环可以帮助虚拟能源枢纽工厂的运营商达到参与多个能源市场的低碳经济潜力。此外，PHEV 所有者和虚拟能源枢纽工厂处于双向通信状态。业主可以通过从附近的虚拟能源中心工厂获取能源来为 PHEV 充电。虚拟能源枢纽工厂和 PHEV 所有者都可以从这种双向通信中获得经济利益。

本文提出的方法是一个多目标优化问题。开发的优化框架的目标是最大化虚拟能源枢纽工厂的利润并最大限度地减少二氧化碳排放。因此，所提出的优化方法试图优化虚拟能源枢纽工厂中涉及的每个设备的行为，以满足不同的能源需求，考虑基于集成激励的灾难恢复计划来管理最终用户的行为。此外，考虑采用风险规避框架来管理风能的不确定性发电，通过 IGDT 方法处理多目标优化问题。虚拟能源中心工厂数学公式和操作约束在以下小节。

2.1. 目标函数

本研究提出模型的主要目标是通过考虑能量转换设施的限制来最大化虚拟能源枢纽工厂电力的利润。虚拟能源枢纽工厂的财务利润可以用 (1) - (3) 计算，这是从收入和总运营成本减去中获得的。基于 (2)，虚拟能源中心工厂的收入来自向当地电力市场出售额外电力，以及根据虚拟能源中心合同向消费者出售电力，热能，制冷和氢气需求。虚拟能源枢纽工厂产生的总运营成本由三部分组成，如 (3) 所示。第一部分说明购买电力、热能和氢能的成本分别来自当地能源市场。第二部分表示 EES、TES、CES、HES 的运营成本，第三部分代表 DR 计划所产生的激励补偿成本。

$$\max: OFI \tag{1}$$

$$OFI = Rev - TOC$$

$$Rev = \sum_{t=1}^{NT} [\lambda_t^e P_t^{sell} + \pi_t^e ID_t^e + \pi_t^h ID_t^h + \pi_t^c ID_t^c + \pi_t^{hy} ID_t^{hy}] \tag{2}$$

$$TOC =$$

$$\sum_{i=1}^{NT} (\lambda_t^e P_t^{buy} + (\lambda_t^h H_t^{buy} + ((\lambda_t^{hy} HY_t^{buy} + (\pi^{ess} P_t^{ess,dis}) + (\pi^{tes} H_t^{tes,dis}) + ((\pi^{ces} C_t^{ces,dis}) + (\pi^{hes} HY_t^{hes,dis}) + \pi^{e,dr} (D_t^{e,up} + D_t^{e,dw}) + \pi^{h,dr} (D_t^{h,up} + D_t^{h,dw}) + \pi^{c,dr} (D_t^{c,up} + D_t^{c,dw}))) \tag{3}$$

其中 λ_t^e 、 λ_t^h 、 λ_t^{hy} 是从能源市场购买电力、热力和氢气的价格。 π_t^e 、 π_t^h 、 π_t^c 和 π_t^{hy} 是向客户出售的电力、热能、冷却和氢气价格。 P_t^{sell} 、 P_t^{buy} 、 H_t^{buy} 、 HY_t^{buy} 是向电力市场出售(购买)电力、从区域供热市场购买的供热需求以及从氢气市场购买的氢气。 ID_t^e 、 ID_t^h 、 ID_t^c 、 ID_t^{hy} 是初始电力、加热、冷却和氢气需求。 π^{ess} 、 π^{tes} 、 π^{ces} 、 π^{hes} 是 EES、TES、CES 和 HES 的运营成本。 $\pi^{e,dr}$ 、 $\pi^{h,dr}$ 、 $\pi^{c,dr}$ 是电、热和能量冷却需求变化的激励率。 $D_t^{(c),up}$ 、 $D_t^{(c),dw}$ 是增量或实施 DR 计划后能源需求减少。 $P_t^{ess,dis}$ 、 $H_t^{tes,dis}$ 、 $C_t^{ces,dis}$ 、 $HY_t^{hes,dis}$ 分别为 EES、TES、CES、HES 的发电量。

如前所述,拟议的自我调度策略旨在最大限度地减少碳排放率,同时最大限度地提高虚拟能源枢纽工厂的财务利润。因此,第二个目标函数(OF2)试图最小化碳排放,其表述为(4)。OF2 由三个术语组成,包括从当地能源市场购买电力和热力引起的碳排放,以及热电联产机组的排放。

min: OF2

$$OF2 = \sum_{t=1}^{NT} (\xi^e P_t^{buy} + \xi^h H_t^{buy} + \xi^r P_t^r) \dots \dots \dots (4)$$

其中 ξ^e 、 ξ^h 、 ξ^r 是从能源市场和热电联产单位购买电力和热力的排放系数。另外, P_t^r 是 CHP 装置产生的电力。

2.2.能源转换设施的限制

本节详细介绍了构成虚拟能源枢纽工厂结构以满足能源需求的所有能源转换设施的动态和数学建模。值得注意的是,所有这些方程都被视为优化模型的约束[38]。

2.2.1. P2X 转换单位建模

如前所述,本研究中使用了三种类型的 P2X 转换设施,是用于满足冷却、加热和氢气需求。电解槽是一种常见的电力制氢(P2HY)装置,其使

用为氢基消费者提供服务。表示消耗功率之间关系的数学模型(即 P_t^{P2HY})和产生的氢气(i.e., HY_t^{P2HY})(即 HY_t^{P2HY})由 P2HY 单元实现,如(5)

所示。不等式(6)说明了生成的氢气量由 P2HY 单元在每个时间间隔应在最小值内和最大金额。

$$HY_t^{P2HY} = \eta^{P2HY} P_t^{P2HY} \dots \dots \dots (5)$$

$$HY_t^{P2HY,min} \leq HY_t^{P2HY} \leq HY_t^{P2HY,max} \dots \dots (6)$$

电锅炉是常用的设施之一,可用作满足供暖需求的电热(P2H)装置。生成的 P2H 单位的热能(即 H_t^{P2H})可以由(7)计算。在此外,该装置的允许工作范围如下面的不平等式(8)。

$$H_t^{P2H} = \eta^{P2H} P_t^{P2H} \dots \dots \dots (7)$$

$$H_t^{P2H,min} \leq H_t^{P2H} \leq H_t^{P2H,max} \dots \dots (8)$$

其中 P_t^{P2H} 和 η^{P2H} 是 P2H 单元和本单元消耗的功率效率。

此外,电冷水机作为有效的冷却源被用于电源到冷却(P2C)单元的背景,以满足冷却需求。P2C 装置(即 C_t^{P2C})基于性能效率 η^{P2C} 的结果在(9)中给出。该值受到上限和下限的限制,如(10)所示。

$$C_t^{P2C} = \eta^{P2C} P_t^{P2C} \dots \dots \dots (9)$$

$$C_t^{P2C,min} \leq C_t^{P2C} \leq C_t^{P2C,max} \dots \dots \dots (10)$$

2.2.2.储能系统建模

由于忽略了储层的物理结构,因此同样定义了用于实施所有类型存储的能源管理策略的数学建模。首先,每个容量水库(即 S_t^X)在时间间隔 t 可以通过(11)计算,即基于周期 t-1 的能量容量,加上时段 t 的带电能量(电能、热能、冷却能量和氢气),并减去放电能量这一时期的能量[39]。根据式(12),每个存储的能量每个时期的水库必须在允许的范围内。水库的初始水位(即 S_{ini}^X)必须等于存储的能量最后一个时期(即 S_{NT}^X)的自调度问题

(NT),如上所述在(13)中。 $z_t^{(\cdot),dis}$ 和 $z_t^{(\cdot),dis}$

是二元变量,是用于防止每个水库同时充电和放电,即公式为(14)。最后,利用(15)和(16)限制各储层的最小和最大放电和充电能量。

$$S_t^X = S_{t-1}^X + \eta^{X,ch} Y_t^{X,ch} - \frac{Y_t^{X,dis}}{\eta^{X,dis}}; \dots\dots\dots (11)$$

$$X \in \{ees, tes, ces, hes\}; Y \in \{P, H, C, HY\}$$

$$S_t^{X,min} \leq S_t^X \leq S_t^{X,max} \dots\dots\dots (12)$$

$$X \in \{ees, tes, ces, hes\}$$

$$S_{ini}^X = S_{NT}^X; X \in \{ees, tes, ces, hes\} \dots\dots\dots (13)$$

$$z_t^{X,dis} + z_t^{X,ch} \leq 1; X \in \{ees, tes, ces, hes\} \dots\dots (14)$$

$$\begin{cases} Y_t^{X,dis,min} z_t^{X,dis} \leq Y_t^{X,dis} \\ Y_t^{X,dis} \leq Y_t^{X,dis,max} z_t^{X,dis} \dots\dots \end{cases} \dots\dots (15)$$

$$X \in \{ees,tes, ces, hes\}; Y \in \{P, H, C, HY\}$$

$$\begin{cases} Y_t^{X,ch,min} z_t^{X,ch} \leq Y_t^{X,ch}; \\ Y_t^{X,ch} \leq Y_t^{X,ch,max} z_t^{X,ch}; \dots\dots \end{cases} \dots\dots (16)$$

$$X \in \{ees, tes, ces, hes\}; Y \in \{P, H, C, HY\}$$

2.2.3. 热电联产单元建模

热电联产机组是至关重要的热电联产设施，在输入端口使用天然气同时产生电力和热量。在产生的热量和功率之间建立相互连接的可行操作区域指定了每个 CHP 单元的最佳性能[40]。为了制定每个热电联产单元的凸工作面积，应根据热电生产的最小和最大可能值考虑四个角点 (k = 1, ..., 4)。

基于可行操作区域的每个角点(即 $\omega_{n,t}$)，CHP 装置产生的电量(即 $P_{r,t}$) 和热量(i.e., $H_{r,t}$)分别由(17)和(18)表示。基于(19)和(20)，每个热电联产机组的运行点应始终在其可行范围内操作区域。

$$P_t^r = \sum_{n=1}^{NN} \omega_{n,t} p_n^r \dots\dots\dots (17)$$

$$H_t^r = \sum_{n=1}^{NN} \omega_{n,t} h_n^r \dots\dots\dots (18)$$

$$0 \leq \omega_{n,t} \leq 1 \dots\dots\dots (19)$$

$$\sum_{n=1}^{NN} \omega_{n,t} = 1 \dots\dots\dots (20)$$

2.2.4. EVPL 建模

电动汽车可以通过停车场运营商参与虚拟能源枢纽工厂的自调度问题。存储的能量量每个时间间隔的停车场取决于当时到达/离开停车场的电动汽车的充电状态的增加/减少。(21)-(25) 中的一组限制说明了这一点。各时段停车场充放电电量应根据电动汽车的允许范围来定义，其中规定如(26)-(29)。需要注意的是，本研究中使用蒙特卡罗模拟确定 EV 到达和出发时间以及每辆 EV 到达和离开时的充电状态。

$$S_t^{EV} = S_{t-1}^{EV} + S_t^{EV,arr} + \eta^{EV,ch} P_t^{EV,ch} \Delta t - (S_t^{EV,dep} + \frac{P_t^{EV,dis} \Delta t}{\eta^{EV,dis}}) \dots\dots (21)$$

$$S_t^{EV,max} = \sum_{m=1}^{NM} \sum_{e=1}^{NE} (S_{m,t}^{max} w_{e,t}) \dots\dots (22)$$

$$S_t^{EV,arr} = \sum_{m=1}^{NM} \sum_{t=1}^{N^{arr}} (S_{m,t}^{max} - S_{l,m,t}) \dots\dots (23)$$

$$S_t^{EV,dep} = \sum_{m=1}^{NM} \sum_{t=1}^{N^{dep}} S_{m,t}^{max} \dots\dots (24)$$

$$0 \leq S_t^{EV} \leq S_t^{EV,max} \dots\dots (25)$$

$$0 \leq P_t^{EV,ch} \leq P_t^{EV,ch,max} \dots\dots (26)$$

$$0 \leq P_t^{EV,dis} \leq P_t^{EV,dis,max} \dots\dots (27)$$

$$P_t^{EV,ch,max} = \sum_{m=1}^{NM} \sum_{e=1}^{NE} p_{m,t}^{EV,ch} w_{e,t} \dots\dots (28)$$

$$P_t^{EV,dis,max} = \sum_{m=1}^{NM} \sum_{e=1}^{NE} p_{m,t}^{EV,dis} w_{e,t} \dots\dots (29)$$

2.3. 集成灾难恢复程序约束

虚拟能源枢纽工厂运营商可以与消费者签署合作协议，以执行基于激励的 DR 计划，消费者可以根据虚拟能源中心工厂运营商的合理报销来调整其能源使用概况。在发达的决策框架中，假设电力、加热和冷却需求可以参与基于激励的灾难恢复计划。基于 (30) - (32)，根据每个需求的参与率，分别对电气，加热和冷却需求执行综合 DR 计划。综合 DR 方案实施后最终电气 (即 $D_t^{e,dr}$)、热

量(即 $D_t^{h,dr}$)、冷却(即 $D_t^{c,dr}$)可以分别由(33)-(35)计算。

$$\begin{cases} D_t^{e,up} \leq \alpha^e ID_t^e \\ D_t^{e,dw} \leq \alpha^e ID_t^e \end{cases} \quad (30)$$

$$\begin{cases} D_t^{h,up} \leq \alpha^h ID_t^h \\ D_t^{h,dw} \leq \alpha^h ID_t^h \end{cases} \quad (31)$$

$$\begin{cases} D_t^{c,up} \leq \alpha^c ID_t^c \\ D_t^{c,dw} \leq \alpha^c ID_t^c \end{cases} \quad (32)$$

$$D_t^{e,dr} = ID_t^e + D_t^{e,up} - D_t^{e,dw} \quad (33)$$

$$D_t^{h,dr} = ID_t^h + D_t^{h,up} - D_t^{h,dw} \quad (34)$$

$$D_t^{c,dr} = ID_t^c + D_t^{c,up} - D_t^{c,dw} \quad (35)$$

2.5.基于 IGDT 的多目标鲁棒算法

IGDT 是一种对严重不确定源进行建模的非概率方法。由于 IGDT 不需要创建大量场景,因此与随机编程相比,计算时间显著减少。事实上,从不确定参数的预期值推导出其预测误差是解决 IGDT 方法优化问题的根本目标。在这项研究中,IGDT 用于处理与风力发电相关的不确定性。在开发的自调度模型中,包络绑定模型用于证明关于不确定参数 P_t^w 的先验知识,其表示为(40)。

$$U(\xi, \tilde{P}_t^w) = \{P_t^w : (1 - \xi)\tilde{P}_t^w \leq P_t^w \leq (1 + \xi)\tilde{P}_t^w, \xi \geq 0\} \quad (40)$$

其中 ξ 表示决策的不确定性范围参数 \tilde{P}_t^w 为预测风力发电量, P_t^w 表示不确定参数的实际值。本研究采用基于 IGDT 的抗风险策略形式的鲁棒策略来解决风力发电的不确定性问题,保护虚拟能源枢纽工厂免受风力发电与预测值的不利偏差导致的低利润影响。基于 IGDT 的鲁棒模型公式为(41)。

$$\tilde{\xi}(X, OF_1) = \max\{\xi : (OF_1(X, P_t^w) \geq (1 - \tau)OF_{1,pre})\} \quad (41)$$

其中 $OF_{1,pre}$ 是基于预测风力发电量的虚拟能源枢纽工厂的利润确定性优化问题, τ 为基值的利润偏

差因子。基于所提出的模型,虚拟能源枢纽工厂运营商尝试在满足所需操作限制的情况下,在风险规避方法中最大化不确定性范围参数虚拟能源中心工厂基于 IGDT 的鲁棒自调度问题给出为(42)-(46)。很明显,在(44)中定义的虚拟能源枢纽工厂的最低财务利润将在不确定性 ξ 的范围内,在基于 IGDT 的鲁棒方法中获得最低允许风力发电。因此,在自调度问题中只声明不等式(40)的左侧,

即: $(1 - \xi)\tilde{P}_t^w = P_t^w$ 在这种情况下,虚拟能源中心工厂操作员将面临单级优化问题。这个问题在功率平衡方程中得到了强调,即(45)。

$$\tilde{\xi}(X, OF_1) = \max \xi \quad (42)$$

$$\min : OF_2 \quad (43)$$

Subject to:

$$\min\{Revenue - TOC\} \geq (1 - \tau) OF_{1,pre} \quad (44)$$

$$P_t^{sell} - P_t^{buy} + (1 - \xi)P_t^w + P_t^r + P_t^{ees,dis} - P_t^{ees,ch} + P_t^{EV,dis} - P_t^{EV,ch} - P_t^{P2HY} - P_t^{P2H} - P_t^{P2C} = D_t^{e,dr} \quad (45)$$

$$(5)-(35),(37)-(39) \quad (46)$$

最后,虚拟能源中心工厂运营商试图找到解决多目标问题的最佳解决方案,其中既定目标不一致。为了解决这个问题,虚拟能源枢纽运营商可以采用 ε -约束方法。在这方面, OF_1 是被视为 ε 约束中的主要目标函数方法,而 OF_2 应作为额外的约束添加到主要目标函数。因此,所开发的基于 IGDT 的多目标鲁棒优化问题可以转化为单目标问题,表述为(47)。需要注意的是,所提出的自调度策略的性质是基于基于 IGDT 的鲁棒模型,而 ε 约束方法在制定所提出的策略时没有建设性,仅用于处理基于先前研究的多目标优化问题[28]。

$$\begin{aligned} \max: & \xi S. t. \\ & OF_2 \geq \varepsilon \end{aligned} \quad (47)$$

更改 ε 约束方法的参数,以确定多目标优化问题的最佳布局。对于多目标问题,帕累托集可以通过选择合适的 ε 值来实现。从得到的帕累托集中

选出正确的解决方案，使用基于模糊的决策技术进行选择。因此，每个解决方案都在 (0, 1) 中进行调整。每个目标函数的模糊隶属关系特征为 (48)：

$$\hat{f}_k = \begin{cases} 1 & f_k \leq f_k^{low} \\ \frac{f_k^{max} - f_k}{f_k^{max} - f_k^{min}} f_k^{Low} & \leq f_k \leq f_k^{\phi} \\ 0 & f_k \geq f_k^{\phi\phi} \end{cases} \quad (48)$$

其中 \hat{f}_k 是第 k 个目标函数的归一化量， f_k^{min} 和 f_k^{max} 是第 k 个目标函数的最小值和最大值。多目标优化问题的最优解为在所有解决方案中使用最大最小方法选择如(49)所示。

$$Optimalsolution = \max\{\min[\hat{f}_1, \hat{f}_2, \dots, \hat{f}_m]\}$$

3. 结果和讨论

在本节中，首先给出了处理开发模型所需的输入参数，然后分析了不同案例研究的结果。

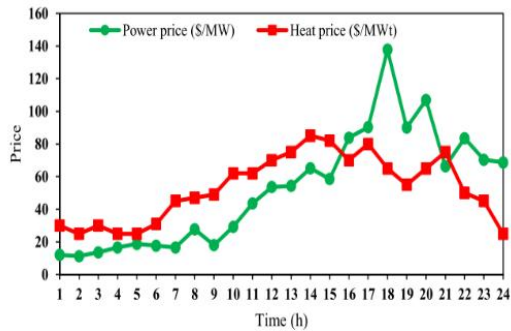


Fig. 2. Predicted power and heat market prices.

Table 2
Technical details of the energy storage systems.

	X = ess Y = P	X = hes Y = HY	X = tes Y = H	X = ces Y = C
$S^{X,min} / S^{X,max}$	2/0.1 (MW)	1/0 (MW)	1/0 (MWt)	1/0 (MWt)
$\gamma^{X,ch,min} / \gamma^{X,ch,max}$	0.5/0.02 (MW)	0.25/0 (MW)	0.2/0 (MWt)	0.25/0 (MWt)
$\gamma^{X,dis,min} / \gamma^{X,dis,max}$	0.5/0.02 (MW)	0.25/0 (MW)	0.2/0 (MWt)	0.25/0 (MWt)
$\eta^{X,ch} / \eta^{X,dis}$	95/95 (%)	100/100 (%)	90/90 (%)	90/90 (%)

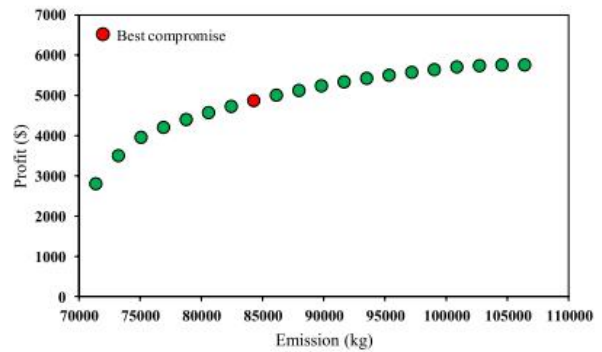


Fig. 4. Optimal Pareto front.

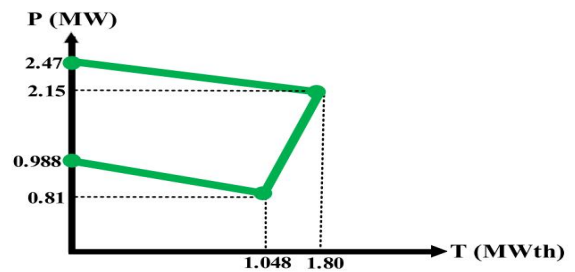


Fig. 3. Technical details of the CHP unit.

Table 3
Technical details of the energy conversion facilities.

$HY^{P2HY,min} / HY^{P2HY,max}$	0/1 (MW)	η^{P2HY}	0.85
$HY^{P2H,min} / HY^{P2H,max}$	0/1 (MWt)	η^{P2H}	1.5
$HY^{P2C,min} / HY^{P2C,max}$	0/2 (MW)	η^{P2C}	0.9

3.1. 模拟设置

所提出的日前稳健经济环境调度在虚拟能源中心进行了测试，如图 1 所示，所考虑的测试床包括一个热电联产单元，风力涡轮机，EVPL，四个存储系统，包括 EES，HES，TES 和 CES，以及能量转换，包括 P2HY，P2H 和 P2C。虚拟能源枢纽工厂的运营商通过操作这些技术必须提供电力、加热、冷却和氢气负载。此外，运营商可以与电网交换电力，并分别从区域供热网络和氢气供应商公司购买热能和氢气。最终消费者使用集成的需求响应参与调度，并有资格将其电气、加热和冷却负荷转移高达 5%。功率和热负荷的负载数据来自参考文献[41]，而氢负荷数据取自参考文献[42]。每小时预测日前电力和供热市场价格如图 2 所示[43]。表 2 和表 3 分别显示了存储系统和能量转换技术的技术细节。图 3 显示了从参考文献[44]获得的 CHP 单元的可行操作区域。此外，用于电动汽车的数据可在参考文

献[45]中找到。为了全面提供结果，定义了两个仿真场景。第一种情况集中在虚拟能源枢纽工厂的调度上，以确定性的方式提供详细信息，并且本案例研究中排除了不确定性的影响。然而，第二种情景研究将侧重于风电功率预测带来的不确定性的影响。需要注意的是，虚拟能源中枢的技术细节在优化之初就设置过一次，在所有场景中都保持不变。所有仿真均使用标准 PC 上的 CPLEX 求解器执行。

3.2. 仿真结果

方案一. 建议的经济环境调度以确定性的方式运行，在这种情况下不考虑不确定性。忽略利润的虚拟能源枢纽的最大可实现利润为 5760.821 美元，忽略利润的最小可获得排放为 71384.233 公斤。寻求妥协在利润和排放之间，运行了多目标问题使用 ϵ 约束方法进行 20 次迭代，步长为 2308.5。图 4 说明了最优帕累托前沿。可见，通过增加利润，排放量上升了。根据该图，在仅考虑虚拟能源枢纽工厂的经济性以达到最大利润的情况下，例如[8-16]，将向环境中释放 106401.044 千克的排放量。然而，通过采用模糊隶属函数，最佳折衷方案为\$4870.337 和 84285.551 kg，这是第二天考虑经济和环境方面的最佳操作点。因此，此时报告剩余结果。

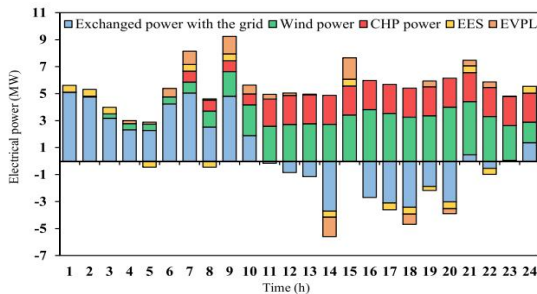


Fig. 5. Optimal electrical power dispatch.

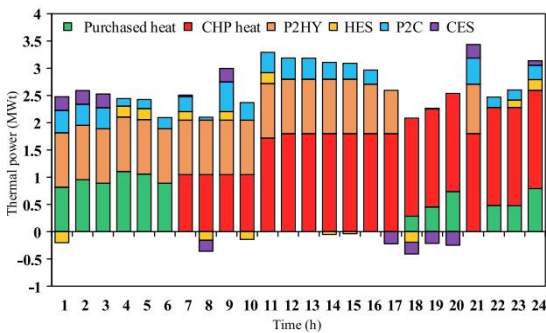


Fig. 6. Optimal thermal power dispatch.

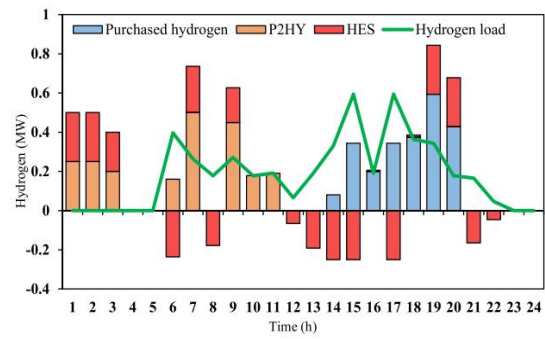


Fig. 7. Optimal hydrogen dispatch.

图 5 描绘了虚拟能源枢纽工厂的最佳电力调度。仔细观察交换的电力会发现与电价有明显的相关性。在低电价（如 $t = 1-10, 22$ ）期间，运营商购买电力，而随着电价上涨，运营商向电网出售电力。由于电价的运营成本较低，热电联产单元的发电量从 $t = 11$ 增加到 $t = 24$ 。此外，当电价较低时，ESS 适用于充电模式，例如 $t = 1-3, 7, 9, 15, 21, 24$ ，然后通过达到高电价（例如 $t = 5, 8, 14, 17-20, 22$ ）应用于放电模式。由于电价的局部极值，在 $t = 5, 7$ 处放电和在 $t = 21$ 处充电。另外，在一天的最后时刻充电的主要原因是满足与初始条件相同的 SoC 水平。虽然电动汽车的充电/放电取决于其到达和离开的时间，运营商试图从可用的 V2G 能力中受益高峰电价时段的电动汽车。就此而言，在 $t = 14, 18$ 期间，20 日，当电价达到峰值时，电动汽车被放电。

图 6 表示虚拟能源的最佳热调度枢纽工厂。由于可行操作区域的特点热电联产机组在运行时也会产生热量。因此，在 $t = 7-24$ 期间，供应加热负载主要依赖于 CHP 单元。P2H 组件的主要运行发生在非高峰电价时段，在高峰时段，即 $t = 18-20, 22-24$ ，运营商从区域供热网购买电力。运营商还在一天的最初几个小时购买了电力以满足热负荷，因为 CHP 装置在此期间关闭。除了 CHP 和 P2H 装置外，运营商还试图从中受益通过在低电价时段充电并放电来从 HES 获取在高峰时段，如 $t = 8, 10, 18$ 。关于冷调度，唯一的选择是全天运行的 P2C 技术 $t = 17-20$ 除外。在此期间，操作员将 CES 放电到满足高峰电价带来的冷负荷。此外，图 7 说明了最佳氢气调度。如图所示，直到 $t = 11$ （此时电价较低），电力才会转换为氢气。主要的在此期间 P2HY 的目标是给 HES 充满电，以便电

价较高时段放电。此外,采购氢气是操作员在 $t = 14 - 20$ 期间使用的另一种选择以满足氢气需求。请注意不同的图案关于能源调度(图 5-7)与能源价格以及需求相关,并且模拟设置仍然与 3.1 节中介绍的相同。

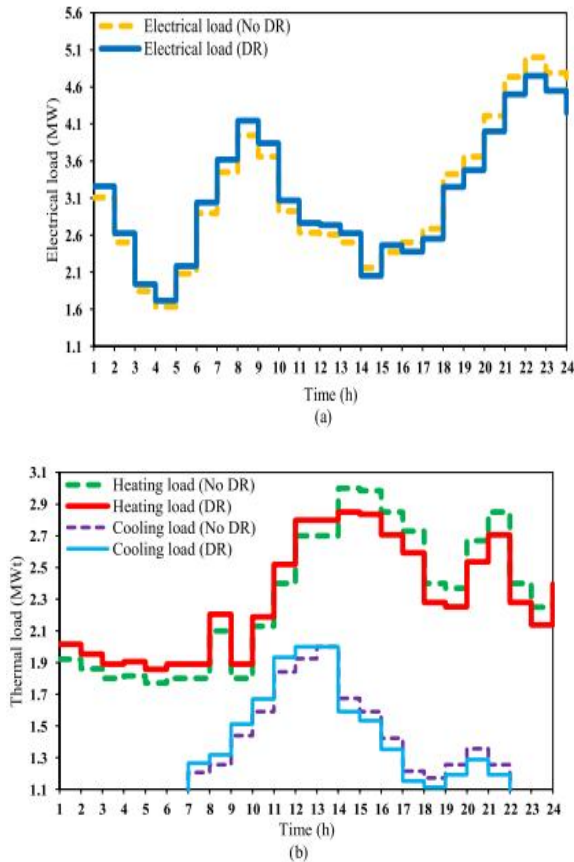


图 8 显示了实施综合需求响应的影响。

从图 8-(a) 可以看出,电力负荷变化和电价之间的负相关关系。错峰电价期间小时,即 $t = 1 - 13$,负载向上移动,而从 $t = 14 - 24$,负载被向下移动。对于加热和冷却来说,类似的模式是显而易见的负载。这是因为 P2H 和 P2C 转换技术,它们

Table 4
Impact of the flexible technologies on economic of the virtual energy hub plant.

	No flexible technology	Storage systems	Storage systems + P2HY + P2H	Storage systems + P2HY + P2H + integrated DR	Storage systems + P2HY + P2H + integrated DR + V2G
Exchanged power cost (\$)	1159.845	972.626	1466.720	1385.745	1315.846
Purchased heat cost (\$)	1829.245	1823.516	1131.282	1041.607	1023.979
Purchased hydrogen cost (\$)	272.448	244.315	141.029	141.029	141.029
Profit (\$)	-4143.774	4342.693	4618.7103	4813.187	4870.337
Emission	93902.532	89958.648	87025.996	84676.294	84285.551

的运行取决于电价。最后给出表 4 为了展示所实施的灵活技术对虚拟能源中心工厂的经济性的影响。总体而言,灵活的技术对利润和排放显示出积极的影响。通过考虑所有技术,利润增加了 726.896 美元 (17.54%)排放量减少 9616.981 公斤 (10.24%)。在这些技术中, P2HY 和 P2H 对利润的影响最大,下降了约 6.3%,而存储系统对排放的影响最大,减少了 4.19%。这些通过灵活运用统计数据验证了该提案的优越性与之前的作品(如 [18-27])相比,这些技术忽略了 P2HY 和 P2H 设施, [8-37] 忽略了集成的 DR 计划, [8-16] 忽略了 PHEV。

灵活运用利润增加的原因是通过仔细检查这些技术的调度来揭示这些技术。如前所述,存储系统和电动汽车都是当能源价格低时充电,当能源价格低时放电价格高,运营商可能会拒绝购买,

Table 5
Economic analysis of the virtual energy hub plant under the uncertainty.

	$r = 0.8$	$r = 0.6$	$r = 0.4$
ξ	0.31	0.63	0.95
Exchanged power cost (\$)	1919.570	2535.754	3318.232
Purchased heat cost (\$)	1715.001	2522.437	2522.437
Purchased hydrogen cost (\$)	192.188	192.188	192.188
Profit (\$)	3896.269	2922.202	1948.134

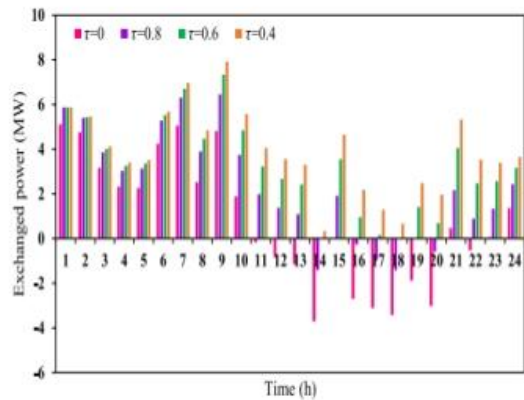


Fig. 9. Impact of the uncertainty on the exchanged power.

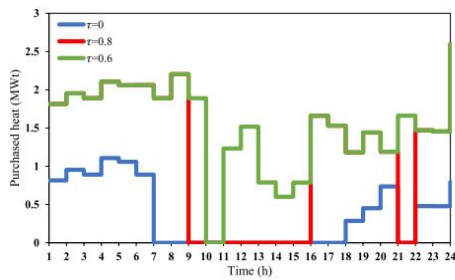


Fig. 10. Impact of the uncertainty on the purchased heat.

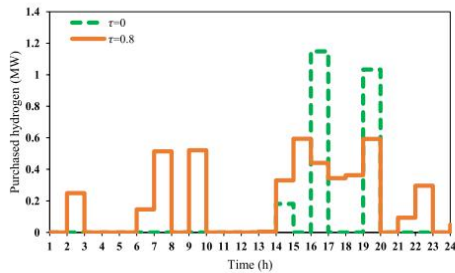


Fig. 11. Impact of the uncertainty on the purchased hydrogen.

能源或出售多余的能源以获得金钱利益。此外，当电价低于氢气和热量时，P2HY 和 P2H 可以通过将电力分别转换为氢气和热量来增加收益。在此过程中运营商无需购买氢气和热量，而是使用 P2HY 和 P2H 技术来节省成本，从而增加利润。此外，通过将负荷从高能源价格小时数转移到低能源价格小时数的集成灾难恢复可以降低高峰能源价格时段的供应成本，从而增加利润。

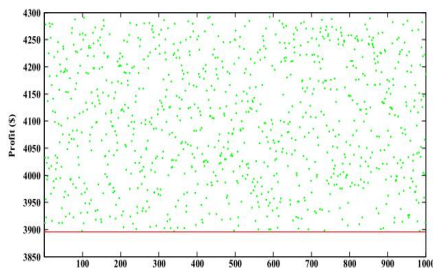


Fig. 12. Robustness validation of the proposed IGDT method.

场景二。在这种情况下，考虑了风的不确定性研究了考虑因素及其对调度的影响。在折衷点处获得的确定性利润，即 4870.337 美元，为 IGDT 技术中的基值。表 5 列举了虚拟能源中心工厂在不确定情况下的结果。根据从该表中可以看出，通过减小 τ 的值， ξ 的值会增加。这意味着考虑较少的利润，风电预测误差较多可以处理。换句话说，对于特定量的 τ ，例如， $\tau = 0.8$ ，保证最大风功率预测误差为 31%，利润不会低于 \$3896.269。图 9

描绘了不同 τ 值的不确定性下的交换功率。这是明确指出，在不确定性的情况下，向电网出售电力下降，购买力上升。这是由于坚固的强制调度调度的 IGDT 技术的行为风力发电较少。因此，经营者必须弥补不足通过从电网购买电力。此外，通过减少派送风力发电，P2H 容量减少，随之而来的是如图 10 所示，购买的热量增加。同样的趋势是氢的影响很明显，如图 11 所示。

此外，为了验证所提出的不确定性的稳健性管理层假设不确定性已经实现并且风电偏差 30%。然后，通过不确定参数的不同实现，生成了 1000 个场景正态分布（标准差为 30%）。图 12 说明了每个场景的虚拟能源中心工厂的总利润。按原样，显然，所有场景的总利润都没有低于考虑 IGDT 方法的利润，这意味着该方法是对不确定性相当稳健。另外，事后分析还展示了 IGDT 方法在不确定性中的有效性管理。为此，在分析中提出了上述偏差（30%）。假设运营商补偿缺乏实时购电风电调度市场[46]。这样，总利润下降至 3681.109 美元在作品中获得的确定性方式[8,10,11,17]；然而，通过考虑 IGDT 方法，总利润为 3935.509 美元（增加 6.9%）。这是因为 IGDT 方法考虑了缺乏风电实行日前调度，以便购电不需要实时市场。

4. 结论

引入的能源管理模型有助于虚拟能源中心工厂提前一天进行稳健的调度。拟议的模型根据多目标函数考虑了经济和环境方面。此外，还考虑了信息差距决策理论（IGDT）来解决风电的不确定性。通过数值结果验证了模型的成功性能。根据结果，考虑存储系统利润增长 4.8%，在排放中下降 4.19%。此外，能源转换设施显示出显著的利润增加约 6.3% 并减少排放量减少约 3.26%。实施的综合需求应对计划产生了积极影响，约占 4.2% 利润增加，排放量减少 2.7%。此外，该能力对 IGDT 技术进行了评估。进一步评估了 IGDT 技术的可利用性。结果表明，IGDT 方法保证了利润不会小于特定风电功率误差的特定阈值。还表明，考虑到不确定性，由于风力减少，运营商从市场上购买了更多的能源，导致利润下降。但是，减少比情况多 6.9% 在不确定性实现后没有不确定性。

参考文献（略）

关系型数据库智能算力实现智能化终端感知画像

潘嘉 杨慰民

(中国移动通信集团福建有限公司)

摘要: 移动通信用户增长迅速, 在网络智能化终端与应用众多, 网络接口及业务信令复杂, 依靠传统的分析方法难以从海量数据中快速提取有用的价值数据, 难以及时准确发现终端自身问题, 难以全面、准确评估终端与网络、业务之间匹配相关性, 无法实现通过匹配分析来主动指导终端测试、市场营销和客户服务工作。通过调度关系型数据库智能算力, 可以高效实现智能化终端感知画像。

关键词: 智能算力; 终端感知

引言

移动通信用户增长迅速, 在网络智能化终端与应用众多, 网络接口及业务信令复杂, 依靠传统的分析方法难以从海量数据中快速提取有用的价值数据, 难以及时准确发现终端自身问题, 难以全面、准确评估终端与网络、业务之间匹配相关性, 无法实现通过匹配分析来主动指导终端测试、市场营销和客户服务工作。使用旧的数据库存储和计算都要大量占用系统资源。

面临的难题 1: 成本高、效率低

采用传统人工测试, 评估终端端到端性能, 存在覆盖不全、成本高、效率低的缺点, 难以满足 4G 时代数据量指数级增长需要。

面临的难题 2: 多部门协调困难

终端测试需要各专业、各部门协同, 从而实现终端、网络、业务之间的匹配分析, 提升基于端到端整体自动化感知管控能力。

面临的难题 3: 使用原始 DPI 数据需要大量数据库资源, 容易出现存储、算力不足的问题

通过调度关系型数据库智能算力实现智能化终端感知画像, 可以高效解决上述问题。

1. 智能算力终端感知画像体系

1.1 数据库架构选择

目前最先进的数据库技术可以总结为以下几个方面:

(1) 分布式数据库: 传统的关系型数据库在处理大规模数据时容易出现性能瓶颈, 而分布式数据库可以在多个物理节点上存储和处理数据, 提高了性能和容量。一些流行的分布式数据库包括 Apache Cassandra、MongoDB 等。

(2) 内存数据库: 传统数据库通常将数据存储于磁盘上, 而内存数据库则将数据存储于内存中, 因此具有更快的读写速度和较低的延迟。一些流行的内存数据库包括 Redis、Memcached 等。

(3) 图数据库: 图数据库是为了处理图结构数据而设计的数据库, 可以高效地进行复杂的关系查询, 适用于社交网络、推荐系统等场景。一些流行的图数据库包括 Neo4j、ArangoDB 等。

(4) 时间序列数据库: 时间序列数据库特别适用于存储和处理按时间顺序排列的数据, 如传感器数据、日志数据等。一些流行的时间序列数据库包括 InfluxDB、TimescaleDB 等。

(5) 新 SQL 数据库: 新 SQL 数据库是一类结

作者简介:

潘嘉: 中国移动通信集团福建有限公司网络部网管中心员工

杨慰民: 中国移动通信集团公司首席专家; 中国移动通信集团福建有限公司网络部网管中心副总经理

合了传统关系型数据库和 NoSQL 的特性的数据库，旨在提供可伸缩性和高性能的同时保持 ACID 特性。一些流行的新 SQL 数据库包括 CockroachDB、TiDB 等。这些数据库技术在不同场景下具有各自的优势和适用性，选择最适合的数据库技术应该根据实际需求 and 业务特点进行综合考虑。

在现代的数据驱动世界中，智能算力是推动各种应用的关键因素，包括但不限于人工智能、大数据分析、云计算等。关系型数据库在这方面扮演着重要角色，它们通过提供可靠、高效和灵活的数据存储和查询服务，为智能算力提供了坚实的基础。

关系型数据库的核心是“关系模型”，这是一种用于描述数据和数据之间关系的高级抽象。在关系模型中，数据被组织为一系列的表，每个表都有一组定义好的列和任意数量的行。这种数据组织方式可以有效地支持各种数据操作，包括插入、删除、更新和查询等。

使用关系型数据库实现智能算力的一个主要优势是它们的查询能力。通过使用结构化查询语言

(SQL)，用户可以快速、准确地从大量数据中提取所需信息。这对于智能算力来说至关重要，因为它需要快速处理和分析大量数据以生成有用的洞察。

另一个优势是关系型数据库的事务处理能力。事务是一系列操作，它们被视为一个单一的工作单位。关系型数据库通过提供原子性、一致性、隔离性和持久性 (ACID) 的保证，确保事务的可靠执行。这对于需要保证数据一致性和完整性的智能算力应用来说非常重要。

关系型数据库还支持索引和优化器等高级功能，可以进一步提高查询性能和数据处理能力。

此外，许多关系型数据库还支持存储过程和触发器等程序化功能，可以在数据库服务器端执行复杂的数据处理任务，从而减轻客户端的负载。

2. 系统应用与实践

2.1 应用场景

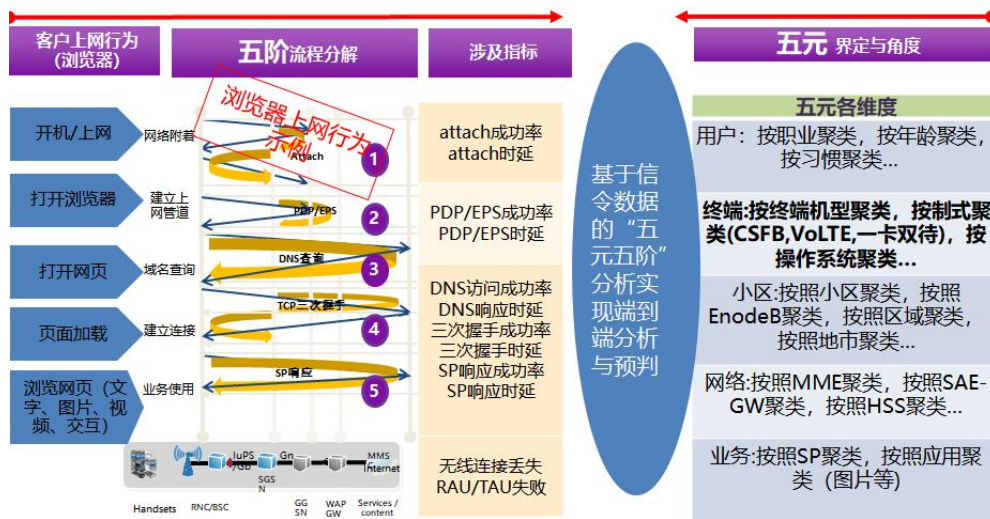
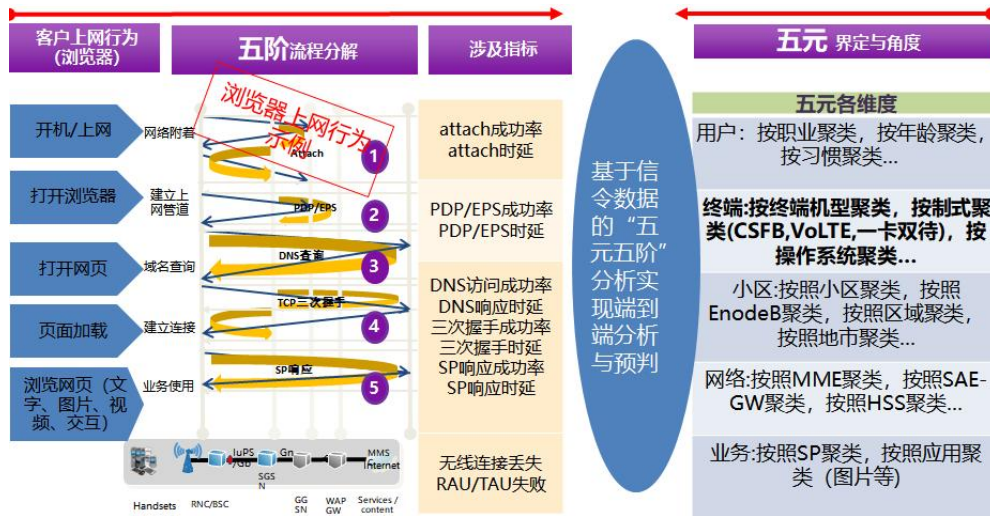


图 1 检测任务流程

终端与网络匹配分析主要是基于“五元五阶分析法”，通过捕捉上网业务全流程完整记录，借助指标量化、模型分析和客户画像等手段，从终端、网络和业务三个维度全面感知网络，从而协助客服

部门减少服务响应时间和提高关怀针对性与准确度、协助终端公司提高终端投放精准度、协助端到端问题快速定位与解决，促进终端、网络、业务协同发展。



2.2 终端信息建模方法

建立“终端信息表”、“终端机型信息表”、“小区信息表”、“基站邻接关系信息表”以及“用户偏好信息表”用于协助终端、网络、业务匹配分析。

“终端信息表”：IMEI 与终端机型关联，协助终端维度聚合分析

“终端机型信息表”：终端机型与终端特征关联，协助终端维度更深入的聚合分析

“小区信息表”：小区/TAC/基站/无线厂家/小区场景关联，协助小区维度聚合分析

“基站邻接关系信息表”：基站邻接关系，协助小区维度聚合分析

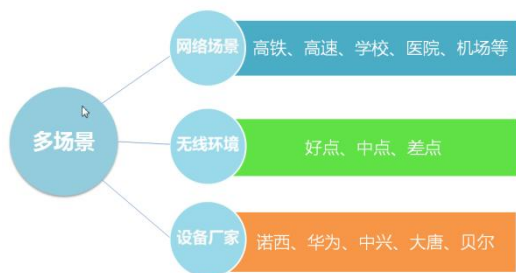
“用户偏好信息表”：描述用户业务偏好，协助适配分析

<p>终端信息表</p> <p>终端IMEI * 终端机型 是否测试终端 开始使用时间</p> <p>补充说明： * 为key属性； 灰色字体表示功能 尚未实现。</p>	<p>终端机型信息表</p> <p>终端机型 * 终端品牌 终端制式 是否支持TDS 是否支持WCDMA 一卡双待 CSFB VOLTE 上市时间 价格区间 分辨率 ...</p>	<p>小区信息表</p> <p>小区号 * TAC 基站号 无线厂家 软件版本 基站类型</p> <p>基站邻接关系信息表</p> <p>基站号 * 邻接基站号 *</p>	<p>用户偏好信息表</p> <p>用户IMSI * 浏览业务占比/流量 经常访问的SP1 经常访问的SP2 经常访问的SP3</p> <p>视频业务占比/流量 ... 用户主要活动区域 ...</p>
--	---	--	--

2.3 指标引导与场景分类

通过在福建省集中性能管理平台对终端各项指标的提取，发现终端在注册、接通、保持、切换等方面的问题。针对终端出现问题的业务类型，进行该业务在多场景下的轮换测试，最后找出问题较为集中发生的网络场景，解决终端在该场景下的网络适配性问题。

- 学校：用户集中，业务量集中于某些时段；
- 高铁：用户集中，移动速度极快，专网，无线环境变化快、发生重选，切换较多；
- 高速：用户集中，移动速度快，无线环境变化快，发生重选，切换较多；
- 医院：用户集中，话务量大；
- 机场：用户集中，进行大量附着去附着。



2.4 智能算力搭建

1. 挑选合适的关系型数据库管理系统 (RDBMS): 根据业务需求、数据规模以及性能要求等因素, 基于福建移动特点, 选择 Microsoft SQL Server 数据库。

2. 硬件需求评估: 分析数据大小、并发用户数、读写频率等因素, 评估所需的硬件配置, 包括主机、存储、网络等。本期使用 3 台物理机, 3TB 存储资源和物理机全部内存。

3. 搭建数据库环境: 根据硬件需求, 搭建数据库服务器, 安装和配置所选的 RDBMS。

4. 数据库设计和优化: 设计关系型数据库的表结构 (共计 45 张表), 确保良好的数据组织和关系, 并建立索引 (12 个)、分区 (2 个) 等优化措施来提升性能。

5. 水平和垂直扩展: 根据业务需求和负载情况, 规划后续进行水平扩展 (增加服务器节点 2 个)、垂直扩展 (增加服务器性能) 来提升数据库的智能算力。

6. 负载均衡和高可用性: 使用负载均衡技术可以将流量分发到多个数据库服务器上, 提高系统的并发处理能力。

7. 监控和调优: 建立合适的监控系统, 跟踪数据库的性能指标, 及时发现潜在问题并进行调优。

2.5 系统应用情况

终端画像输入数据:

1. 设备信息: 手机终端感知可以收集用户设备的相关信息, 如设备型号、操作系统版本、屏幕分辨率、网络类型等, 这些信息可以用于了解用户使用的手机类型和环境条件。

2. 应用使用情况: 手机终端感知可以监测和记录用户在手机上使用的应用程序的情况, 包括应用的使用时间、使用频率、应用的版本信息等。通过分析用户的应用使用情况, 可以了解用户的兴趣爱好、偏好等。

3. 位置信息: 手机终端感知可以利用手机的 GPS 定位功能或网络基站定位功能获取用户的位置信息。通过分析用户的位置信息, 可以了解用户的地理位置、移动轨迹、常去地点等, 用于提供个性化的位置服务或广告推荐。

4. 传感器数据: 手机终端感知可以利用手机内置的传感器 (如加速度传感器、陀螺仪、光线传感器等) 收集用户的环境信息, 如手机的运动状态、光线强度等。这些信息可以用于实现基于用户行为的智能功能, 如摇一摇、自动调节屏幕亮度等。

基于上述数据, 进行打分加权, 每日自动形成终端感知评分结果。

已在省内广泛推广使用, 与日常运维工作紧密结合, 发现的案例提交集团公司终端管理平台, 用户群体包括省市业务管理人员、市场部和终端公司, 该系统已全面融合到日常的生产管理工作中。

3. 结束语

利用现网 DPI 海量数据, 通过全量用户数据采集, 以手机用户作为天然样本, 充分利用样本空间置换方法, 通过关系型数据库智能算力实现智能化终端感知画像, 突破了传统分析方法技术瓶颈, 实现了对终端问题的快速、准确分析和定位。通过捕捉上网业务全流程完整记录, 借助指标量化、模型分析和客户画像等手段, 从终端、网络和业务三个维度全面感知网络, 从而协助客服部门减少服务响应时间和提高关怀针对性与准确度、协助终端公司提高终端投放精准度、协助端到端问题快速定位与解决, 促进终端、网络、业务协同发展

参考文献:

- [1] 中国移动 5G 终端测试规范.
- [2] 中国移动通信集团. 中国移动统一 DPI 设备技术规范[S]. 北京

基于智能算力的智慧园区中智能楼宇管理和节能优化

范永圣

(中鸿达信息科技有限公司)

摘要: 随着智能技术的迅猛发展,人们对于智慧园区和节能优化的需求日益增加。智能化的楼宇管理和节能优化成为智慧园区中的重要组成部分。本文旨在探讨基于智能算力的智慧园区中的智能楼宇管理和节能优化方法。本文首先介绍了智慧园区和智能楼宇的概念,并阐述了智能楼宇管理和节能优化的关键技术,包括数据采集、数据分析和智能控制等。然后,本文分析了智慧园区中节能优化的需求和挑战,在此基础上提出了基于智能算力的智慧园区中智能楼宇管理和节能优化的策略。通过对楼宇内部各个子系统进行集

关键词: 智慧园区, 智能楼宇管理, 节能优化, 智能算力, 大数据分析

引言

随着城市化进程的加速和可持续发展的要求,智慧园区的建设已成为现代城市发展的重要方向。在智慧园区中,智能楼宇管理和节能优化是关键领域。智能算力作为智慧园区建设的重要支撑技术,可以为智能楼宇管理和节能优化提供强大的计算能力和智能决策。本论文旨在探讨基于智能算力的智慧园区中智能楼宇管理和节能优化的方法和策略。

1. 智慧园区中智能楼宇管理技术概述

1.1 智慧园区和智能楼宇的概念

智慧园区是指利用先进的信息技术和通信技术,将传统的城市或园区转变为智能化、数字化和可持续发展的现代化城市空间。智慧园区以提高居民生活质量、优化资源利用和环境保护为目标,通过整合智能化的基础设施、服务和管理,实现智能化的城市运营和管理。

而智能楼宇是智慧园区中的重要组成部分,指的是在建筑物内部融入先进的信息技术和自动化设备,实现楼宇内部各个系统的智能控制和互联互通。智能楼宇通过感知、通信和智能决策等技术手段,实现楼宇的高效运行和舒适环境的提供。

智慧园区和智能楼宇的共同目标是通过应用物联网、云计算、大数据分析等先进技术,实现城市和建筑物在能源消耗、环境保护、安全管理、交通流动等方面的智能化和优化。通过实时监测和数据分析,可以实现楼宇能效的提升、节能减排的目标,提高住户、企业和园区的运营效率和舒适度。

智慧园区是以智能化、数字化和可持续发展为特点的现代化城市空间,而智能楼宇作为智慧园区的一部分,通过融合先进的信息技术和自动化设备,实现对楼宇内部系统的智能化管理和优化,为居民提供高效、舒适、可持续的居住和工作环境。

1.2 智能楼宇管理的关键技术

智能楼宇管理依赖于以下关键技术来实现数据的采集、分析和智能化控制:

1) 数据采集技术

传感器和监测设备: 通过安装在楼宇内部的传感器和监测设备,收集楼宇各个系统(如电力、照明、空调、水暖等)的运行数据,包括能耗、温湿度、光照等。

无线通信技术: 利用无线传感器网络和物联网技术,实现对传感器数据的无线采集和传输,从而实现楼宇内部数据的实时监测和远程访问。

2) 数据分析技术

大数据分析：通过对大量的楼宇数据进行收集、存储和处理，运用数据挖掘、机器学习和统计分析等方法，获取有价值的信息，如能耗模式、设备故障预测等，从而优化楼宇管理。

数据可视化：采用图表、仪表盘和报表等可视化工具，将数据转化为易于理解和分析的形式，帮助管理者快速识别问题、找到优化策略。

3) 智能控制技术

自动化控制系统：通过集成楼宇设备和系统，利用自动化控制技术实现对楼宇运行的自动化和智能化管理，如自动调节照明、空调和电力设备等。

智能决策和优化算法：基于数据分析结果，采用智能决策和优化算法，自动化地调整楼宇各个系统的运行参数，以达到节能、舒适和性能优化的目标。

响应式控制：通过与外部环境的互动，如能源价格、天气预报等，实时调整楼宇的运行策略，以提高能源利用效率。

数据采集技术、数据分析技术和智能控制技术作为智能楼宇管理的关键技术，可以实现对楼宇能源消耗、运行效率和舒适性的监测、分析和调控，从而实现节能优化和智能化管理的目标。

2. 智慧园区中节能优化的需求和挑战

港口卸船机落料口智能停车定位系统 1、节能优化的重要性

智慧园区中节能优化的重要性不容忽视，它对于可持续发展和环境保护有着重要的影响。以下是具体描述智慧园区中节能优化的重要性：

1) 资源保护

节能优化可以最大限度地减少能源的使用和消耗，从而保护有限的能源资源。智慧园区通过应用先进的技术和策略，能够实现能源的高效利用，减少能源浪费，从而降低对能源供应系统的依赖性，促进资源的可持续利用。

2) 环境保护

能源消耗是导致温室气体排放的主要原因之一。智慧园区中的节能优化措施有助于减少温室气体的排放，降低碳足迹，对于环境保护和应对气候变化具有重要意义。这有利于改善空气质量、减少大气污染，保护生态环境，营造更加宜居和健康的园区环境。

3) 经济效益

节能优化带来的能源成本的节约是显而易见的。

通过降低能耗和提升能源利用效率，智慧园区可以减少其能源开支，增加经济效益。此外，通过引入节能技术和设备的更新，还可以促进新兴产业的发展，创造新的经济增长点，推动园区可持续发展。

2、节能优化的需求

1) 节能减排

随着能源资源的日益紧缺和环境压力的增大，智慧园区对节能减排的需求越来越迫切。节能是实现可持续发展的关键，因此，智慧园区需要通过优化能源消耗，减少能源浪费，降低碳排放。

2) 能源成本控制

能源成本是园区运营的重要成本之一，因此，园区希望通过节能优化来降低能源消耗，减少能源费用支出，并提升经济效益。

3) 提升园区形象

智慧园区采用节能技术和策略，可提升园区绿色、可持续的形象，提升了园区的竞争力和吸引力，更好地满足居民和企业对环境友好的需求。

3、节能优化的挑战

1) 多样化的能耗模式

智慧园区内存在多种能耗模式，如办公楼、商业中心、住宅区等。这些区域对能源消耗的需求和要求不同，因此，需要针对多样化的能耗模式开展节能优化工作。

2) 复杂的系统互动

智慧园区内部存在复杂的系统互动，如照明系统、空调系统、电力系统等。这些系统之间的相互影响和耦合会给节能优化带来挑战，需要综合考虑系统的整体性能与效益。

3) 技术和设备更新

随着科技的不断发展和进步，节能技术和设备也在不断更新。智慧园区需要不断跟进最新的技术和设备，同时面临技术更新带来的投资成本和技术更新的管理挑战。

4) 行业标准和规范缺乏

尽管智慧园区的节能优化是重要的，但行业标准和规范仍然相对缺乏。确保节能优化的标准化和规范化是一个挑战，需要建立相应的准则和指南，以便在整个智慧园区中实施节能优化策略。

综上，智慧园区中的节能优化既面临着重要的需求，如节能减排、能源成本控制和形象提升等，也面临着多样化的能耗模式、复杂的系统互动、技

术和设备更新以及行业标准与规范缺乏等挑战。有效应对这些挑战,实现智慧园区的节能优化,是智慧园区可持续发展的关键之一。

3. 基于智能算力的智慧园区中智能楼宇管理策略

1、能耗监测与仿真模拟

1) 能耗监测

通过安装传感器和监测设备,实时采集各个楼宇系统(如电力、照明、空调等)的能耗数据。这些数据包括电量、功率、温度等指标,利用数据采集技术实现对能源消耗行为的监测。

将采集的能耗数据进行处理、分析和统计,利用大数据分析和机器学习技术,生成能耗报告和分析结果。这些结果能够帮助管理者了解能源消耗的情况,发现潜在的节能机会和问题。

2) 仿真模拟

基于实际楼宇的数据和参数,利用建模工具和算法,创建楼宇能耗的仿真模型。该模型能够模拟楼宇系统的运行情况,并预测能耗和效果。

通过在仿真模型中进行不同的能耗优化方案的模拟和比较,帮助管理者评估不同策略的效果,并提供决策支持。这可以包括调整设备运行参数、改变供电策略、优化照明和空调系统等。

通过能耗监测与仿真模拟,智能楼宇管理能够实时监测能耗情况,及时掌握能源消耗的状况,发现异常行为和um设备故障,精确分析能耗数据,发现潜在的能源浪费问题,并提出相应的节能改进措施,在仿真模型中预测不同策略的能耗效果,辅助管理者做出正确的决策和优化方案选择。

总而言之,在智慧园区的智能楼宇管理中,能耗监测与仿真模拟是重要的策略,它们利用智能算力和数据分析的能力,帮助园区管理者实时监测能耗、发现节能机会,并通过仿真模拟评估和优化能源管理策略,从而实现可持续发展和高效能耗的目标。

2、智能照明系统管理

1) 智能照明系统的原理和应用

智能照明系统是一种基于先进的技术和控制策略的照明系统,它具有自动化、智能化和节能化的特点。

智能照明系统的原理是通过集成传感器、控制器和通信设备,实现对照明设备的自动化调控和智能化管理。传感器可以感知周围环境的变化,如光

线、人体活动等,根据感知结果发出信号给控制器。控制器根据接收到的信号,通过智能算法和预设的规则,自动调节照明亮度、开关状态和色温,从而满足不同场景和需求的照明需求。

智能照明系统广泛应用于各种场景和领域,包括居住区、商业办公楼、公共场所等。

居住区:智能照明系统可以根据居民的活动和时间,自动调节室内照明亮度,提高居住舒适度,并节约能源。

商业办公楼:智能照明系统可以根据办公楼内部的人流量和光线情况,自动调节照明亮度和开关状态,提供舒适的工作环境,并节约能源。

公共场所和道路照明:智能照明系统可以根据光线变化和人流量,实现智能控制和动态调节,提供安全可靠的照明服务,并减少能源浪费。

2) 照明系统的智能控制策略

照明系统的智能控制策略是为了实现节能、舒适和智能化的照明管理而设计的。

光线感知:照明系统配备光线传感器,能够感知周围环境的光照强度。根据感知结果,系统可以自动调节照明亮度以保持适宜的光照水平。

周期调光:照明系统在不同时间段设定不同的亮度水平。例如,在白天使用较低的亮度,而在晚上或需要更明亮的场景下使用较高的亮度。

活动感知:使用人体红外传感器等技术感知人体活动。当没有人在区域内时,照明系统自动降低亮度或关闭灯光,以节约能源。

区域划分:将照明区域划分为不同的部分,例如工作区、休息区、过道等。通过分别控制每个区域的照明,可以更精确地满足不同区域的照明需求,并避免不必要的能源浪费。

个性化设置:根据用户的偏好和需求,提供个性化的照明设置。例如,用户可以调整亮度、色温等参数来获得最适合自己的照明效果。

数据分析:通过采集和分析照明系统的使用数据,了解能源消耗情况和用户行为模式。根据数据分析的结果,优化照明策略,减少能源浪费并提升能源利用效率。

自适应学习:照明系统可以通过机器学习算法学习用户的行为模式和喜好,自动调节照明设置,从而提供个性化和智能化的照明体验。

这些智能控制策略可以结合使用,根据实际需求和场景进行组合和调节,以实现节能、舒适和智

能化的照明管理。从用户角度来看,智能控制策略可以提供个性化的照明体验,提高舒适度和效率。从能源管理角度来看,智能控制策略可以减少能源消耗,降低运行成本,并对能源浪费进行监测和优化。

3、智能空调系统管理

1) 智能空调系统的原理和应用

智能空调系统的原理是通过结合先进的传感器、智能控制器和通信技术,实现对空调设备的智能控制。智能空调系统集成了各种传感器,例如温度传感器、湿度传感器、人体感应传感器等。这些传感器可以感知室内外的环境参数变化,如温度、湿度、人员活动等。

传感器将感知到的数据传输给智能控制器。智能控制器通过对数据的采集和分析,能够了解当前环境的状况,如室内温度、湿度等,并预测未来的变化趋势。基于数据的分析结果和预设的控制策略,智能控制器会自动调节空调设备的工作模式、温度、湿度和风速等参数,以提供舒适的室内环境。智能空调系统通常与互联网相连,可以实现远程控制和监测。通过通信技术,用户可以使用手机应用或者 web 界面对空调进行控制和调节。

智能空调系统的应用非常广泛,包括居家生活、商业办公等领域。

居家生活:智能空调系统可以根据居住者的习惯和需求,自动控制室内温度和湿度,提供舒适的居住环境。例如,当居住者离开家时,系统可以自动关闭空调以节能,当居住者回家时,系统可以自动启动空调提前调整室内温度。

商业办公:在办公场所,智能空调系统可以根据工作时间和员工数量,自动调节空调工作模式和温度,提供舒适的办公环境。此外,系统还能实时监测能耗情况,提供能源管理的报表和优化建议。

智能空调系统的原理和应用通过传感器感知、数据分析、智能控制和通信技术实现了对空调设备的智能管理和优化。这种系统能够提供舒适的室内环境,节约能源,提升用户体验

2) 空调系统的智能调控策略

自适应控制:根据室内外温度差异和人员活动情况,自动调节空调温度。例如,在人员较少或无人时,可以适当提高温度以节能;而在人员较多或需要更舒适环境时,可以降低温度。

预测控制:通过分析历史数据和外部环境变化

趋势,预测未来的温度需求,并提前调整空调工作模式和温度,以避免过度冷却或加热。

人体感应:利用人体传感器或摄像头等技术,感知人员的活动情况。当没有人员在特定区域时,可以自动降低或关闭该区域的空调,以节约能源。

时间段调节:根据不同时间段的需求,设定不同的空调参数。例如,在白天可以设置较高的温度,而在夜间可以降低温度以提供更舒适的睡眠环境。

定时控制:通过预设定时器,自动调节空调的开关、温度和风速等参数。例如,在人员离开办公室后自动关闭空调,或在人员到达办公室前自动开启空调。

数据采集:通过传感器和监测设备,采集室内外温度、湿度、二氧化碳浓度等数据,以及空调运行状态和能耗数据。

数据分析:对采集到的数据进行分析,识别能耗模式、舒适性问题 and 潜在的节能优化机会。

通过智能调控策略,空调系统可以根据实时数据和预测分析,自动调整工作参数,提供舒适的室内环境,并实现能耗的优化和节约。

4.结论

本文主要阐述了基于智能算力的智慧园区中智能楼宇管理和节能优化。通过综合运用智能传感器、控制器和数据分析等技术,提出了一种综合性的智能楼宇管理方案,并实施节能优化措施。智能算力在智慧园区中的智能楼宇管理和节能优化方面具有重要的应用价值。通过合理而有效地应用智能算力技术,能够提高楼宇能耗效率,减少能源浪费,增加舒适性,降低运营成本。

未来,基于智能算力的智慧园区中智能楼宇管理和节能优化仍面临一些挑战和发展方向。首先,虽然智慧园区中的智能算力技术已经取得了一定进展,但仍需进一步发展和创新。例如,深入探索更先进的传感器技术,提高数据的准确性和实时性,以更好地支持智能楼宇管理和节能优化。

此外,随着人工智能和大数据技术的发展,智慧园区中的智能楼宇管理和节能优化将进一步发展。应用机器学习和数据挖掘技术,从大量的数据中挖掘隐藏的模式和规律,在整个园区范围也将进行更加精准的能源管理和优化,为园区的可持续发展和绿色环保做出更大的贡献。

港口卸船机落料口智能停车定位系统

陈璟 高伟铭

(福州港罗源湾码头有限公司)

摘要: 在散货码头中, 货船上的煤炭、金属等矿石材料主要用卸船机抓取, 再通过集卡从卸船机落料口装运后运输到港区堆场。在传统的作业中, 集卡司机需要通过现场调度员以对讲机的形式提醒司机集卡是否对准卸船机落料口, 作业效率低, 安全性低, 一旦未对准落料口可能出现司机驾驶舱被压或货物散落集卡外部的情况。港口卸船机落料口智能停车定位系统是针对于散杂货码头开发的一款卸船机倒场车的智能安全辅助定位系统。系统通过综合三路视频动态分析车辆的位置和行进方向, 利用人工智能技术结合 AI 算法, 判断车斗相应位置是否对准落料口, 以及是否存在车头停留落料口下方区域的危险情况, 实时定位倒场车位置, 通过落料口倒场车区域的引导屏幕, 辅助倒场车司机安全停车。以此逐步减少现场指挥员数量并最终实现智能停车定位对运输车辆的无人化安全自动引导作业, 降本增效。

关键字: 落料口; 停车定位; 人工智能; AI 算法

引言

伴随着科技进步, 我国港口现代化要求不断提高, 港口卸船机落料口设备逐渐向大型化、自动化的方向发展。在发展过程中, 改善港口货运质量以及提高落料口精准落料工作安全性能一直是个重要问题。

通过对市场调查, 目前并未有类似的港口卸船机落料口智能停车定位系统应用于市场上。本论文设计的港口卸船机落料口智能停车定位系统通过测试平台进行测试, 测试成功后, 将部署于福州港罗源湾码头。

1. 相关工作

1.1 系统背景

卸船机落料口智能停车定位系统是针对于散杂货码头开发的一款卸船机倒场车的智能安全辅助定位系统。系统通过视频检测车头、车身和车尾的位置, 预先标定上方落料口在图像中的位置, 基于现有作业停车合规的位置, 并综合三路视频动态分析车辆的位置和行进方向。完成判断车斗相应位置是否对准落料口, 以及是否存在车头停留落料口下方区域的危险情况。根据倒场车前、后、侧面三个

视角的三路同步视频, 实时定位倒场车位置, 辅助倒场车司机安全停车。本系统预留与卸船机开斗放料操作联动功能^[1-3]。

利用人工智能技术, 通过落料口倒场车区域的引导屏幕, 告知和提醒倒场车司机倒车停车是否安全到位。逐步减少现场指挥员数量并最终实现智能停车定位对运输车辆的无人化安全自动引导作业, 降本增效。同时避免因停车不到位造成的落料碾压车头, 造成驾驶司机伤亡。同时降低员工粉尘吸入、矿物腐蚀等职业病的伤害可能。提升装卸精准度, 提升漏泄料对环境的影响污染。提升装卸车辆作业效率, 降低车辆作业的碳排时长, 节能减排。

智能系统应保证服务器系统的可靠性、稳定性, 确保 AI 算法及呈现展示系统正常稳定工作, 满足各项功能技术指标。

- 1、本系统实施对象包含 1 台卸船机 (指定 3 号机);
- 2、卸船机落料口智能停车定位系统主要包含核心部分: 运输车辆停车定位智能安全监控系统 (含车头、车身、车尾 3 路视频实施检测分析);
- 3、卸船机卸船机落料口智能停车定位系统计算运行服务器 1 台 (部署在 3 号卸船机机房);

4、利用当前 3 号机现有的视频前端设备、局域网组网网络设备、视频存储设备及机房环境配套设施；

5、综合现场调研情况和环境，选购安装 1 台倒塌车专用引导显示 LED 屏。设备及防护应满足内陆和沿海户外使用环境，在设计及工艺上充分考虑防腐、防尘、防雷电、防漏电、防暴雨、防台风，充分考虑海港城市的盐雾天气、灰尘等恶劣环境影响；

6、各项系统主要技术指标满足系统要求。

1.2 系统设计要求

1、安全要素检测核心量化指标：算法模型完成持续深度学习及优化后，纯视觉引导车头停放位置在非危险区域（落料无碾压车头风险）综合准确率 $\geq 90\%$ ；

2、效率要素检测核心量化指标：算法模型完成持续深度学习及优化后，视觉引导车厢停放位置误差值 $\leq 10\text{CM}$ ；

3、本地引导屏幕呈现时效量化指标：车辆动态变化与本地屏幕显示内容的同步刷新状态时间 $\leq 1\text{S}$ 。

表 1 视频流模块说明

模块	子模块
视频流接入服务	车前角度视频流接入
	车身后侧视频流接入
	车身侧面视频流接入
场景算法服务	视频流解析（车辆位置判断）
数据源接入方式	
数据接入	支持通过 RTSP 协议接入网络摄像机视频流，或通过国标协议对接第三方平台获取视频流。
数据格式支持	
视频分辨率	1080P
视频编码格式	H.264、H.265

1.3 系统难点

在传统的停车辅助定位场景中，激光雷达是一种简单有效的方法，但在散货码头下，由于需要在各种天气场景下 24 小时作业，传统的采用激光雷达的方法不仅成本高，而且在雨天工作效果较差，

本项目采用基于纯视觉的方案又有以下几个难点：

需要考虑各种场景，不同光照条件下对车辆的定位识别，需覆盖雨天、雾天、夜间、清晨强光直射作业场景；

受现场安装条件限制，在摄像头未正对车辆的情况下如何判断是否发生偏移是项目的另一个难点。

在行车作业中可能存在车辆遮挡的情况，如何通过车辆跟踪，解决车辆遮挡问题。

1、如何在恶劣天气、强光照影响下使系统能够稳定运行，并且精准检测车辆位置。

2、如何精准判断倒塌车辆位置偏移，引导车辆下一步操作。

1.4 系统创新

根据卸船业务作业排班，人工本地化启动系统后自动化开始 AI 在线检测应用。

1、多路视频实时融合采集现场数据达到全方位实时捕捉车辆位置。

2、采用视觉处理，训练新的目标检测模型实现在恶劣天气、强光照的情况下达到高准确率实时检测。

3、使用边缘 AI 算法（前后帧坐标差值算法、单应矩阵车辆偏移定位算法）实现多车辆跟踪与车辆偏移分析。

4、在引导显示层上无误差实时根据车辆位置来引导车辆操作，流程如下：

1) 目标检测阶段，视频流检测，ROI 区域均无目标时，纯文字显示“无车辆”状态。

2) 发现有作业车辆后，启动对倒塌车的实时位置定位 AI 解析，当车辆位置靠后。车头、车斗均不满足要求时，屏幕显示“请倒车”文字，同时颜色+图文引导方式：红色车体+箭头+文字。

3) 当车辆位置靠前。车头满足，车斗不满足要求时，屏幕显示“请前移”文字，同时颜色+图文引导方式：红色车体+箭头+文字。

4) 当车头、尾、身的三路视频解析检测到车辆均已移动到位，保证车辆位置到位。头、箱均满足要求情况时，屏幕显示“停车落料”文字，同时颜色+图文引导：绿色车体+圆圈+文字。

5) 根据作业进度，本地化关闭检测应用。系统应用结束。

6) 卸船机落料口智能停车定位系统主要包含核心部分: 运输车辆停车定位智能安全监控系统(含车头、车身、车尾 3 路视频实施检测分析)。

2. 系统方案论证与设计

通过对市场调研发现, 目前并未有类似的港口卸船机落料口智能停车定位系统应用于市场上。综合成本、实时性、可靠性等方面考虑, 对比激光雷达、毫米波雷达的性能以及目标检测能力, 最终选择能够支持通过 RTSP 协议、支持 1080P 接入网络摄像机视频流来实现三路同步实时融合策略采集现场数据。

本论文设计的卸船机落料口智能停车定位系统采用边缘端运算模式, 摄像机采用独立供电, 网线 RTSP 直连取流。该系统整体逻辑是通过使用配备 Yolov5 功能的边缘设备, 对车辆目标进行检测, 具体检测目标有车辆车头, 车身和顶部到料的漏斗。Yolov5 模型使用的是检测精度较高的 Yolov5X^[4-7]。整体检测方案部署通过使用 TensorRT 技术, 配合 c++ 接口, 直接进行网络结构的搭建与部署。后期进行算法逻辑分析, 对于检测到的目标, 比照车身与倒料口区域的位置, 进行判断并在定制化显示引导层上进行显示提醒。实现实时检测车辆倒场车位置以及落料漏斗精准落料的功能。卸船机落料口智能停车定位系统的整体布局逻辑设计将在第 3 章以及第 4 章进行详细。

当车辆到达到车点时, 系统会动画提示“请倒车”, 此时车辆应该倒车, 此时的数据不断更新, 主要保留数据为车厢上角的坐标, 动画显示请倒车。当车辆到达到料区域下方时, 检测前 5s 的车厢右上角坐标信息, 确保前 5 秒的坐标都在到料区域以内, 就动画提示停车落料。到料区域被设定为车厢 x 坐标与漏斗 x 坐标差值的绝对值 30 以内。当车辆驶离到料区域时, 动画更新为带停车, 清空所有内部存储的数据。等待下一辆车到来, 引导内容如图 1 所示。

Yolov5 检测时, 并不是每一帧都能检测出数据, 比如画面遮挡, 多辆车遮挡的时候, 就会出现检测失败^[5-6]。检测失败的这帧应该进行丢弃, 不能记录, 不然会导致数据的误差较大。这种情况需要有保护机制, 我们设定 100 帧为极限, 如果前一帧有数据, 接下来的最多 100 帧数据全部有问题, 系统将进行

数据清空, 认为当前没有车辆, 等待下一辆车。如果 100 帧内出现了一帧有数据, 就不做状态丢弃, 继续保留原来数据。以上方法经过实测可以解决该问题^[8-12]。

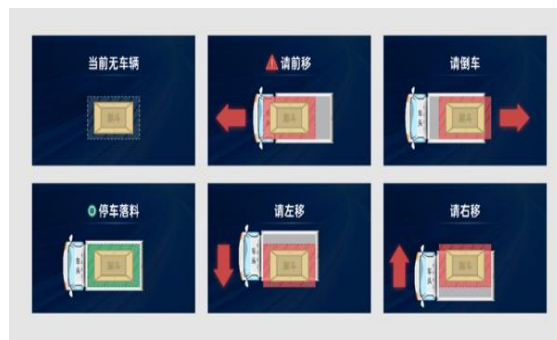


图 1 系统引导显示内容

在实验现场的的车前、车身后侧、车身侧面部署能够支持 RTSP 协议的网络摄像机视频流的海康摄像机。对于车前角度的摄像头对获取的视频流能够通过算法识别倒场车车头关键目标信息识别, 系统能够实现对车辆车头位置的 AI 解析定位^[13-20]。车身后侧摄像头采集车身后侧视频信息, 系统能够对倒场车车位关键目标信息识别并能够实现对车辆车身尾部位置进行定位解析。车身侧面的摄像头获取倒场车侧部视频流信息, 系统能够对对车辆车头、车身两部分进行定位解析, 具体摄像头安装如表 2 所示。

3. 系统原理与总体布局







3.1 系统原理

本系统对于落料口倒场车设定倒车区域, 倒车区域设定为 x 的坐标范围 (700-1800), y 的坐标范围为 (590-650), 为一个矩形区域。区域右侧的头部, x 的范围 (1700-1800), y 的范围 (590-650) 设定为倒车点。只有当车辆到达倒车点时, 系统给的检测逻辑才会启动, 此时车辆的倒车状态会保持更新。通过对 yolov5 检测出来的目标, 进行筛选, 只选择出落在倒车区域的车辆进行追踪, 当车辆进行倒车时, 动画提示请倒车。为了达到加速检测, 将部署 TensorRt 来对检测进行加速处理。当车辆倒车到到料位置, 检测车厢的右上角坐标, 当该坐标的 x 值与漏斗右侧 x 坐标的差值的绝对值在 30 个

像素以内，就进行动画提示停车落料。车辆停车落料 5s 后，当车辆驶离时，动画更新为待停车。具体情况在第 4 章系统软件设计与流程、第 5 章系统测试与分析章节进行详细描述。当车辆到达到车点时，系统会动画提示“请倒车”，此时车辆应该倒车，此时的数据不断更新，主要保留数据为车厢上角的坐标，动画显示请倒车。当车辆到达到料区域下方

时，检测前 5s 的车厢右上角坐标信息，确保前 5 秒的坐标都在到料区域以内，就动画提示停车落料。到料区域被设定为车厢 x 坐标与漏斗 x 坐标差值的绝对值 30 以内。当车辆驶离到料区域时，动画更新为带停车，清空所有内部存储的数据。等待下一辆车到来。

表 2 摄像头安装位置

编号	识别功能	安装位置采集内容	位置（仅供参考）
1	识别对倒场车车头关键目标信息识别，系统实现对车辆车头位置的AI解析定位。		
2	识别对倒场车车尾关键目标信息识别，系统实现对车辆车尾位置的AI解析定位。		
3	识别对倒场车车身关键目标信息识别，系统实现对车辆车头、车身位置的AI解析定位。		

Yolov5 模型由骨干网络、颈部和头部组成^[8]。

Yolov5 模型的骨干网络主要由 Focus、BottleneckCSP 和 SSP 网络构成，其中主要包括 Focus、Conv 卷积块、BottleneckCSP 和 SSP 等模块。

Yolov5 使用的 BottleneckCSP1 和 BottleneckCSP2 能在保证准确的同时，提高网络速度。YOLOv5 的颈部采用 BottleneckCSP2 模块减少模型参数量，通过上采样操作 80*80*512 大小的特征图，上采样过程由 2 组 BottleneckCSP2、大小

步长为 1 的 Conv、Upsample 和 Concat 连接完成，其中 Concat[i,j]中 j 代表网络中第 j 层操作得到的特征图。

在三个不同尺度特征图上生成候选框，由于本文对烟雾的检测只有 2 类，即烟雾和非烟雾，对 YOLOv5X 的 3 个尺度特征图使用 3 种大小不同的候选框对烟雾进行预测，因此对于第三章自制数据集，一共有 21 个候选框。最后输出采用加权非极大值的方式对目标框进行筛选对生成的 21 个候选

框进行筛选，输出目标分类和边框回归。

TensorRT 是可以在 NVIDIA 各种 GPU 硬件平台下运行的一个 C++推理框架。TensorRT 是一个高性能的深度学习推理 (Inference) 优化器，可以为深度学习应用提供低延迟、高吞吐率的部署推理。TensorRT 可用于对超大规模数据中心、嵌入式平台或自动驾驶平台进行推理加速。TensorRT 通过组合层和优化内核选择来优化网络，从而改善延迟 (latency)、吞吐量 (throughput)、能效 (power efficiency) 和内存消耗 (memory consumption)。

在对目标检测模型性能评价中，最常采用的是精确率 (Precision) 和召回率 (Recall) 又称为查准率和查全率，精确率衡量检测结果中正样本的概率，而召回率衡量数据样本中正样本被检测出来的概率，计算方式如式 3.1 和式 3.2 所示^[1]。

$$Precision = \frac{TP}{TP + FP} \# \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \# \quad (2)$$

其中，TP 表示正数据样本被模型检测为正确的数量，FP 表示负数据样本被模型检测为正确的数量，FN 表示正数据样本被模型检测为负的数量。

在模型验证时都希望精确率和召回率越高越好，但是两者是互相博弈的关系，一个指标上升的同时往往带来另外一个指标的下降。一个好的模型会使两者均处于较高的水平，而差的模型会牺牲一个指标换取另外一个指标的提升。为了综合评价模型的性能，学者们引入了 P-R (Precision-Recall) 曲线和 F1 值。P-R 曲线是模型的检测精确率随着召回率的增加而变化的曲线，它可以直观的显示算法整体的精确率和召回率^[1]，当一条曲线完全将另一条曲线包围住时，我们认为前者的精确率和召回率均高于后者，或者曲线下方的面积越大代表的模型性能更优。F1 值是精确率和召回率的调和平均数，值越大代表模型性能越好，计算方式如式 3.3 所示^[1]。

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall} \# \quad (3)$$

在多类别目标检测任务中，对模型评价时常用到 AP (Average Precision) 指标，代表某一个类别的平均精确率，mAP (mean Average Precision) 则表示多个类别的平均精确率，计算方式如式 3.4 和 3.5 所示。

$$AP = \int_0^1 P(r) dr \# \quad (4)$$

$$mAP = \frac{1}{C} \sum_{C_k \in C} AP_{(C_k)} \# \quad (5)$$

其中 p 、 r 分别表示精确度和召回率， $p(r)$ 表示精确度随召回率变化的 P-R 曲线。对目标检测速度的评价采用 FPS (Frames Per Second) 指标，表示每秒钟内模型检测图像的帧数。

NMS 算法是最早应用于候选框筛选的方法，该算法旨在选取候选框置信度的局部极大值，并且依据候选框之间的交并比 (Intersection over Union, IoU) 抑制局部非极大值候选框^[2]。

NMS 算法是一种硬判决算法，依据冗余框与最高置信度框的重叠度来做决策，将大于阈值的冗余框的类别置信度置为零，否则，保留类别置信度不做修改，其数学表达式如公式 3.6^[2]。

$$s_i = \begin{cases} s_i, Iou(B_k, B_i) < N_i \# \\ 0, Iou(B_k, B_i) \geq N_i \end{cases} \quad (6)$$

其中 s_i 表示候选框的置信度值， B_k 为极大值候选框， B 为冗余框， N 为重叠度阈值， $Iou(B_k, B_i)$ 表示候选框之间的交并比，其表达式如式 3.7 所示^[2]。

$$Iou(B_k, B_i) = \frac{B_k \cap B_i}{B_k \cup B_i} \# \quad (7)$$

对于一个锚框样本而言，若其输入 IoU 较高，而输出 IoU 较低，这部分样本大概率是属于正样本的，直接丢弃会对模型收敛造成影响。基于此，定义了匹配度的概念，如公式 3.8 所示，其中， sa 表示生成的预测框和真实框的 IoU，称其为输入 IoU； fa 表示回归之后的预测框与真实框的 IoU，称其为输出 IoU； u 表示锚点框回归前后的差异，其表示了锚框被正确划分的可能性， u 的定义如公式 3.9 所示； α 是权重系数，用来衡量输入 IoU 和输出 IoU 的重要程度； r 用来衡量锚框回归前后 IoU 差异的重要程度。md 为计算锚框的匹配度。

$$md = \alpha * sa + (1 - \alpha) * fa - u^r \# \quad (8)$$

数据集训练评价指标：交并比 (intersection over union, IoU) 在目标检测的性能评价时用的非常多，表示的是预测的边框和原图片标注的真实边框的交叠率，是两者交集与并集的比值。当比值为 1 的

时候则说明预测的效果达到最佳。如图 2 样本数据集训练标注所示。

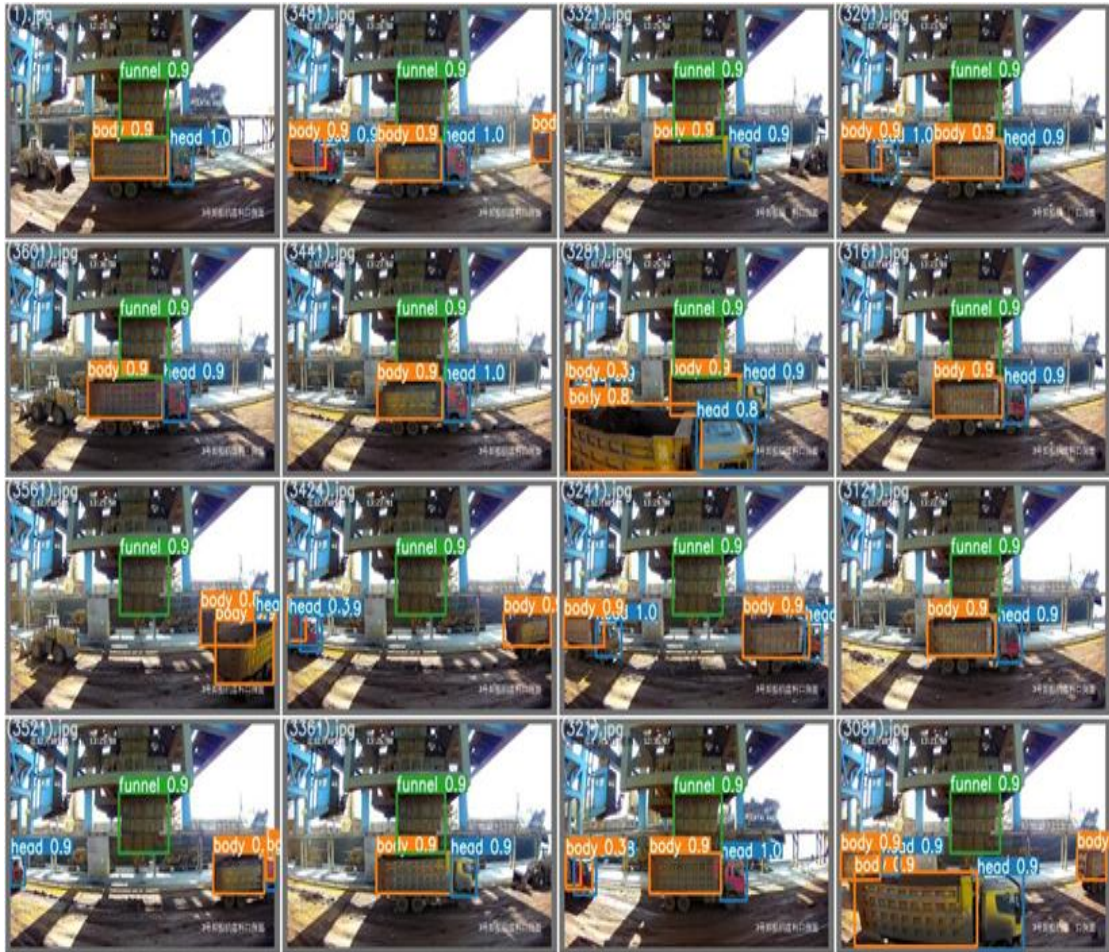


图 2 样本数据集训练标注

3.2 系统全局架构

本论文设计的港口卸船机落料口智能停车定位系统采用边缘端运算模式，每个摄像机采用独立供电，通过网线 RTSP 协议直连拉取视频流传输到服务器端进行处理。服务端与本地引导显示屏的软硬件设备由产品供应方调试。全局架构图及系统相关建设内容（红框内）如图 3 所示。在原有的层级架构上使用边缘设备算法下层对获取到的视频流进行相应的到场车定位信息进行检测处理。该系统

的 AI 赋能作业逻辑如图 4 所示。

本系统的 AI 赋能作业分为三个模块，即场景层、定制化显示引导层、能力算法层。首先，场景层又划分为目标检测、过程解析、结果输出三个阶段。系统通过边缘设备将从网络摄像头获取的视频流信息，对预先划定的 ROI 区域进行目标检测，将检测后视频数据传输到系统进行分析判断车辆位置是否正确，是否满足要求并进行输出，在定制化显示引导层上进行显示引导，其内容如图 1 所示。

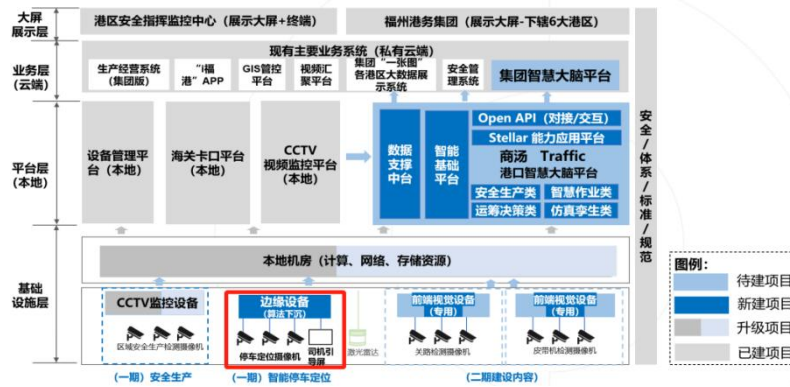


图 3 系统架构图

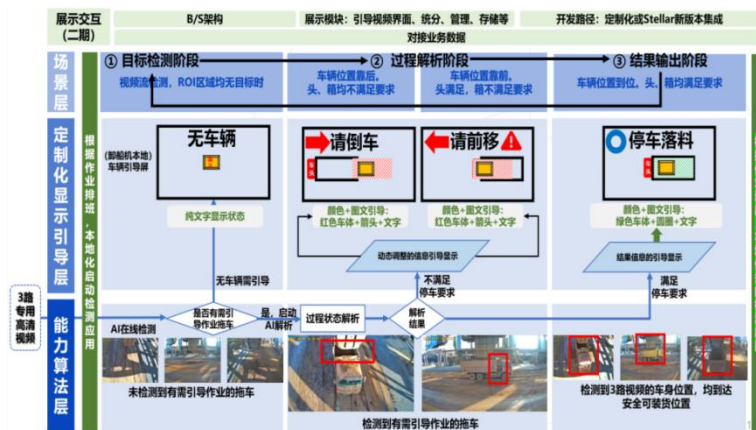


图 4 AI 赋能作业逻辑

在系统内部的能力算法层通过 AI 算法进行目标实时在线检测，对获取的视频流信息进行判断分析是否有需要对倒场车进行引导。若有则进一步 AI 解析，最终判断结果在系统的显示引导层模块的显示屏进行展示。现场显示引导屏安装如图 5 所示。



图 5 现场系统引导显示屏安装

根据系统摄像头检测是否有无车辆，系统判断

车辆所处位置是否满足正确落料要求。当不满足正确要求时，本地系统引导屏将提示司机进行前移、倒车、左移、右移等操作，当车辆满足要求时，则停车落料。当车辆到达到车点时，系统上的本地引导屏会动画提示“请倒车”，此时车辆应该倒车，此时的数据不断更新，主要保留数据为车厢上角的坐标，动画显示请倒车。当车辆到达到料区域下方时，检测前 5s 的车厢右上角坐标信息，确保前 5 秒的坐标都在到料区域以内，就动画提示停车落料。到料区域被设定为车厢 x 坐标与漏斗 x 坐标差值的绝对值 30 以内。当车辆驶离到料区域时，动画更新为带停车，清空所有内部存储的数据。等待下一辆车到来

在系统内部的能力算法层通过 AI 算法进行目标实时在线检测，多获取的视频流信息进行 AI 视频分析判断分析是否有需要对倒场车进行引导。当

车尾部在 C1 与 C2 界外时, AI 系统会判断车身是否摆正并判断车身应向左移还是向右移动; 倒塌车车身 AI 视频分析如图 6 所示。当车身车头部在 D1 与 D2、D3 界外时, AI 系统判断车身是否摆正并判断车身应向前移动还是向后移动。倒塌车车身 AI 视频分析如图 7 所示。当车身车头部在 D1 与 D2、D3 界外时, AI 系统判断车身是否摆正并判断车身应向前移动还是向后移动。

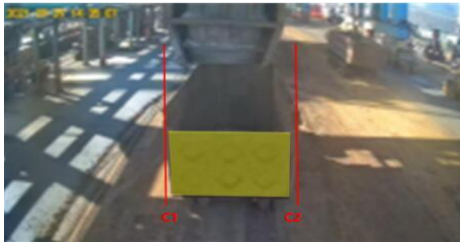


图 6 倒塌车车身 AI 分析

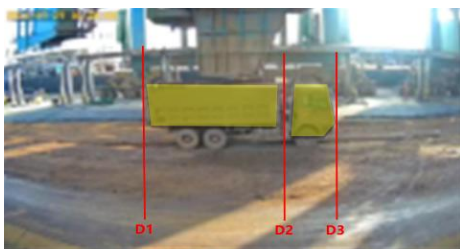


图 7 倒塌车尾部 AI 判断分析

为了判断后视图中翻斗车斗是否对准落料口, 我们根据车辆的朝向和左右偏移作为判断准则, 如图 8 所示。首先提取车斗顶部四个关键点 A、B、C、D (红色表示, 此处 CD 和 EF 在一条直线上, 否则计算直线 CB 与 EF 的交点, 直线 AD 与 EF 的交点), 人工标注的落料口的点 E、F (蓝色表示)。然后根据关键点和落料口的点位置关系作为判断的准则, 以下分别讲述尾部朝向和左右偏移的判断方法。



图 8 车辆朝向判断

车斗关键点检测: 采用神经网络提取车辆车斗关键点, 如若被遮挡则提取可见车斗的顶点。参考 CenterNet^[9]的架构输出目标中心点和关键点。需要标注图像关键点数据集, 修改网络输出头 (output head) 为四个关键点输出, 重新训练网络模型。

车辆尾部朝向: 尾部的关键点 A、B 连线为 l_1 , 落料口点的连线记为 l , 若 l_1 与 l 的夹角在一定可接受的范围 T_θ 内, 则认为车辆基本与落料口平行。

车辆左右偏移: 由于摄像机安装的角度和位置原因, 不同视角和位置观测到的图像都不一样, 车斗和落料口在图像中的位置关系也不同。为此, 为了有统一的判断阈值和不随视角改变的图像, 我们首先将图像相对于落料口平面做图像矫正。然后根据映射之后的车斗顶点 C、D 和落料口顶点 E、F 的关系, 具体说就是车斗左侧距离左侧落料口顶点和车斗右侧距离右侧落料口顶点应该差不多接近。如果 C 到 E 和 D 到 F 之间的距离之差小于一定阈值, $||C - E| - |D - F|| < T$, 则认为左右对齐, 反之则反。

论文设计的港口卸船机落料口智能停车定位系统的硬件安装拓扑结构如图 9 所示。

4. 系统软件设计与流程

本论文设计的港口卸船机落料口智能停车定位系统软件设计与流程如图 10 所示。



图 9 硬件安装拓扑结构

本论文设计的港口卸船机落料口智能停车定位系统软设定倒车区域设定为 x 的坐标范围 (700-1800), y 的坐标范围为 (590-650), 为一个矩形区域。区域右侧的头部, x 的范围 (1700-1800), y 的范围 (590-650) 设定为倒车点。当车辆未到达倒车点时, 系统始终显示待停车, 只有当车辆到达倒车点时, 系统给的检测逻辑才会启动, 此时车辆的倒车状态会保持更新。车辆到达倒车点, 系统显示请倒车, 此时车辆会倒车, 进入倒车区域, 在倒车区域时, 使用前后帧坐标差值法追踪目标车辆, 防止因为多辆车问题导致的坐标数值不准确, 具体做法是当车辆第一次进入倒车区域时, 记录当前帧的车厢右上角坐标, 当下一帧来临时, 如果画面有多辆车, 系统算法能力层将所有车辆的右上角坐标与前帧的数据进行计算差值, 只取前后帧差值 x 在 20 像素以内, y 在 10 像素以内的车辆, 将认为是当前追踪的车辆, 达到车辆实时跟踪。车辆追踪过程中, 对于每帧的数据, 系统都会进行一个暂存, 系统会实时保留 10 秒的数据, 并且取平均值, 防止因为某些特殊情况造成数据的误差。根据所计算出来的坐标数据, 到落料口区域被设定为车厢 x 坐标与漏斗 x 坐标差值的绝对值 30 以内。当车辆驶入落料口区域, 检测其前五秒的数据, 只有车辆都位于落料口区域, 并保证车辆已经停稳, 这时候系统显示停车落料, 并且落料机打开落料标志位。此外, 为了防止数据在倒料区域的临界区跳变导致系统显示跳变, 当车辆达到该区域时, 系统显示落料料标志, 这时候倒料区域适当扩张, 扩大为车厢 x 坐标与漏斗 x 坐标差值的绝对值 33 以内, 来解决当前问题。如果车辆位置在落料口区域的左侧, 则显示车辆向右, 位置在右侧, 则显示车辆向左移动, 若车头正对落料口区域, 则显示向前, 落车身与车位在落料口前方, 则显示倒车。当车辆在落料口区域并且打开了落料标志位时, 等其倒料完成驶出落料区域, 无论前后, 系统均显示待停车并且清空数据, 认为当前车辆已经检测完毕, 等待下一辆车到来。

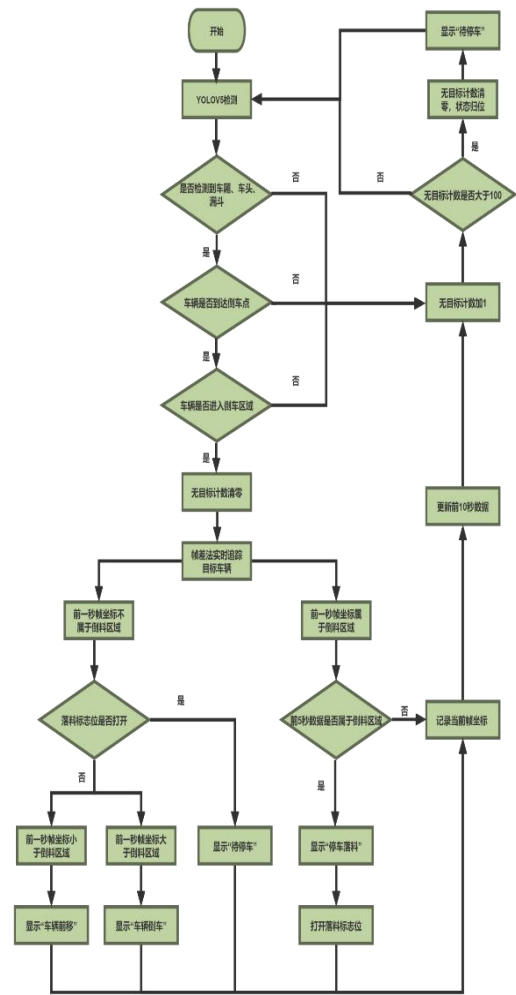


图 10 系统算法流程图

5. 系统测试与分析

5.1 系统测试平台

本论文设计的港口卸船机落料口智能停车定位系统开发实验环境测试平台如下: Ubuntu 18.04、Intel Core i7-7700、NVIDIA GeForce RTX 2060。YOLOv5x 在该平台下, FP32 精度下可达到 20+FPS, FP16 精度下可达到 50+FPS。该测试平台安装依赖项如下: NVIDIA Driver (510.68 版本)、CUDA (11.3.1 版本)、CUDNN(8.2.1 版本)、TensorRT (8.0.1.6 版本)、FFmpeg (4.4.1 版本)、OpenCV (4.0.0 版本)、Qt (5.9.5 版本)、yaml-cpp (0.5.2 版本)。不同环境下模拟测试如图 11、图 12 所示。



图 11 系统模拟测试一



图 12 系统模拟测试二

核心逻辑：设定倒车区域，通过对 yolov5 检测出来的目标，进行筛选，只选择出落在倒车区域的车辆进行追踪，如果车辆位置在落料口区域的左侧，则显示车辆向右，位置在右侧，则显示车辆向左移动，若车头正对落料口区域，则显示向前，落车身与车位在落料口前方，则显示倒车。当车辆倒车到到料位置，检测车厢的右上角坐标，当该坐标的 x 值与漏斗右侧 x 坐标的差值的绝对值在 30 个像素以内，就进行动画提示停车落料。车辆停车落料 5s 后，当车辆驶离时，动画更新为待停车。

本论文设计的港口卸船机落料口智能停车定位系统操作如下。

第一步设定倒车区域：倒车区域设定为 x 的坐标范围 700–1800， y 的坐标范围为 590–650，为一个矩形区域。区域右侧的头部， x 的范围 1700–1800， y 的范围 590–650 设定为倒车点。只有当车辆到达倒车点时，系统给的检测逻辑才会启动，此时车辆的倒车状态会保持更新。

第二步系统显示内容动画检测：当车辆到达到

车点时，系统显示屏会动画提示“请倒车”，此时车辆应该倒车，此时的数据不断更新，主要保留数据为车厢上角的坐标，动画显示请倒车。当车辆达到到料区域下方时，检测前 5s 的车厢右上角坐标信息，确保前 5 秒的坐标都在到料区域以内，就动画提示停车落料。到料区域被设定为车厢 x 坐标与漏斗 x 坐标差值的绝对值 30 以内。当车辆驶离到料区域时，动画更新为待停车，清空所有内部存储的数据。等待下一辆车到来。

5.2 问题分析及解决方案

问题一：通过不断测试，发现系统运行时，如果通过一帧一帧的数据实时进行画面更新的话，动画的实时性效果会很好，但是会出现问题。比如车辆停止时，明明车辆不动，但是 yolov5 检测出的每帧的结果总会有一两个像素的差别。这个问题将会导致，如果车辆正好停在到料区域的左右临界区域，就会出现动画出现“请倒车”、“停车落料”两个动画跳变。

解决措施一：为了解决上述出现的问题，将对所有的保存数据，取 30 帧的数据进行平均，取平均值，可以减小这个问题，通过改动逻辑，当车子进入到料区域（倒车车坐标和漏斗坐标绝对值差值 30 个像素以内），等待 5s，进入到料时，此时的到料区域适当扩张（扩张为车辆坐标和漏斗坐标绝对值差值 33 个像素以内），经过测试，该方案能够完美解决临界区域跳变问题。

问题二：当倒车区域只有一辆车时，以上逻辑都可以适用。但是根据现场的实际情况，很多时候倒车区域里面都是有多辆车，会对逻辑分析进行很大的干扰。需要对多辆倒车车进行目标检测跟踪。

解决措施二：为了解决问题二，我们通过对系统 AI 算法引入前后帧坐标差算法，当第一辆车进入时，系统会保留车辆前一帧的右上角车辆坐标。当下一帧的数据来时，如果画面有多辆车，取每辆车右上角坐标与之前那一帧的数值做差值，在算法上设置前后帧差值 x 在 20 像素以内， y 在 10 像素以内的车辆，就为当前追踪的车辆。实现对目标检测区域内的多辆倒车车位置的跟踪。该措施能够很好的解决问题二，测试效果良好。

问题三：在运行 yolov5 进行目标检测跟踪时，

并不是每帧都能检测出数据，比如画面遮挡，多辆车遮挡的时候，就会出现检测失败的情况。

解决措施三：检测失败的这帧应该进行丢弃，不能记录，不然会导致数据的误差较大。这要保护机制，我们设定 100 帧为极限，如果前一帧有数据，接下来的最多 100 帧数据全部有问题，我们将进行数据清空，认为当前没有车辆，等待下一辆车辆。如果 100 帧内出现了一帧有数据，我们就不做状态丢弃，继续保留原来数据。经过测试，该方法能够完美解决因遮挡导致检测失败的问题，测试情况良好。

问题四：在系统测试过程中由于落料口抖动、阳光直射、逆光、遮挡等干扰因素会造成检测效果较差的情况。

解决措施四：利用不同天气不同环境下的大量数据集来进行训练，对恶劣天气数据集进行图片预处理（灰度化、直方图均衡化）再进行模型训练。如图 13 所示。通过训练新的数据模型，能够达到良好的检测效果。

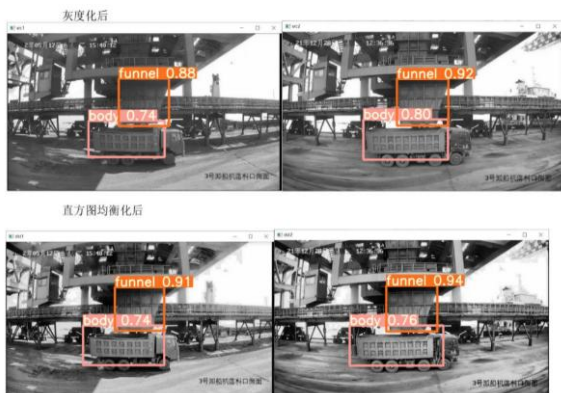


图 13 数据集图像预处理

6 总结

通过一系列的测试以及部署，本论文设计的港口卸船机落料口智能停车定位系统能够精准判断车身相应位置是否对准落料口，能够定位车辆与位置，通过定位分析判断是否需要位置移动，系统显示引导屏会精准显示，提醒司机进行相应的位置移动操作，能够极大程度上车头停留落料口下方区域的危险情况。根据倒塌车前、后、侧面三个视角的三路同步视频，实时定位倒塌车位置，辅助倒

场车司机安全停车。在极大程度上降低了人工成本。同时避免因停车不到位造成的落料砸压车头，造成驾驶员伤亡。还能降低员工粉尘吸入、矿物腐蚀等职业病的伤害可能。提升装卸精准度，提升漏泄料对环境的影响污染。提升装卸车辆作业效率，降低车辆作业的碳排时长，能够起到节能减排的作用。在测试过程中遇到的问题也进行了相应的处理，能够达到预期效果。此外本系统预留与卸船机开斗放料操作联动功能。

参考文献

- [1]陶佳,黄贤英,高钰澜.融合视觉信息的协同知识注意力网络推荐模型[J/OL].小型微型计算机系统:1-8[2023-07-31].DOI:10.20009/j.cnki.21-1106/TP.2022-0410.
- [2]黄健,张钢.深度卷积神经网络的目标检测算法综述[J].计算机工程与应用,2020,56(17):12-23.
- [3]茅智慧,朱佳利,吴鑫等.基于 YOLO 的自动驾驶目标检测研究综述[J].计算机工程与应用,2022,58(15):68-77. Technology,2023,131.
- [4]Xingang M,Shuai L,Xiao Z. YOLO-FR: A YOLOv5 Infrared Small Target Detection Algorithm Based on Feature Reassembly Sampling Method[J]. Sensors,2023,23(5).
- [5]Brett L,Emerson L D,Hyuk C. UAS-Based Real-Time Detection of Red-Cockaded Woodpecker Cavities in Heterogeneous Landscapes Using YOLO Object Detection Algorithms[J]. Remote Sensing,2023,15(4).
- [6]Haiying L,Xuehu D,Haonan C, et al. DBF - YOLO: UAV Small Targets Detection Based on Shallow Feature Fusion[J]. IEEJ Transactions on Electrical and Electronic Engineering,2023,18(4).
- [7]Xinyi L,Baofeng Z,Na L. CAST-YOLO: An Improved YOLO Based on a Cross-Attention Strategy Transformer for Foggy Weather Adaptive Detection[J]. Applied Sciences,2023,13(2).

算力驱动下的精准医疗发展

李 晨

(中国电信股份有限公司福建分公司)

摘 要：本论文旨在研究智能算力驱动下的精准医疗发展，并探讨其在癌症诊断中的具体应用。首先，介绍了精准医疗的概念和重要性，并阐述了智能算力在精准医疗中的作用。其次，通过一个具体的案例，展示了智能算力在乳腺癌早期检测和诊断中的应用效果。该案例利用深度学习模型分析医学影像数据，实现了自动识别异常区域和肿瘤病灶，并提供了癌症风险评估和预后预测。结果表明，智能算力技术显著提高了乳腺癌的早期发现率和诊断准确性，为患者提供了更早的治疗机会。然后，分析了算力驱动下的精准医疗所面临的挑战，包括技术、隐私和伦理等方面的问题，并提出了相应的解决方案。最后，展望了智能算力在精准医疗中的未来发展趋势，包括技术创新、数据共享与合作、法律法规与政策支持以及多方合作与交流。通过本研究的探索和讨论，我们认识到智能算力对于精准医疗的推动具有重要意义，有望为患者提供更好的医疗服务和健康管理。

关键词：智能算力、精准医疗、临床应用

引言

随着科技的飞速发展和医疗需求的日益增长，精准医疗作为一种新兴的医疗模式正在逐渐引起全球的关注。精准医疗以个体化诊断、个性化治疗和预防控制为核心理念，旨在为患者提供更精确、有效的医疗服务。在精准医疗的实践中，智能算力作为一项重要的技术手段和支撑平台，正发挥着不可或缺的作用。

智能算力在精准医疗中扮演着多重角色。首先，智能算力可以快速处理和分析大规模的医学数据，例如基因组学数据、医学影像等，从而实现个体化的疾病诊断和治疗方案的制定。其次，智能算力可以挖掘医疗数据中的潜在关联和模式，帮助医生和研究人员发现新的生物标志物和疾病机制，促进疾病的早期检测和预防。此外，智能算力还可以支持医疗资源的合理分配和管理，提高医疗服务的效率和质量。

本研究旨在深入探讨智能算力在精准医疗中的作用和应用，既从理论角度分析智能算力与精准

医疗的关系，又通过实际案例研究验证其在医疗领域的效果和优势。具体而言，本文将首先介绍精准医疗的概念和原理，阐述其在改善医疗效果和推动医疗进步方面的重要性。然后，重点探讨智能算力在精准医疗中的应用领域和技术手段，包括基于大数据分析的个体化诊断、智能辅助决策系统等。最后，通过案例研究和实践经验总结，总结智能算力驱动下精准医疗发展的现状和挑战，并展望其未来发展的前景和潜力。

通过本文的研究，我们旨在加深对智能算力在精准医疗中的应用和意义的理解，为推动医疗技术的创新和医疗服务的提升提供参考和启示。同时，我们希望通过智能算力驱动下的精准医疗的深入探索，为医疗机构和政策制定者提供相关建议和决策支持，为人们的健康福祉贡献力量。

1. 精准医疗概述

1.1 智能算力与精准医疗的关系

智能算力在精准医疗中扮演着至关重要的角

色。它是指利用先进的计算机技术和人工智能算法,处理和分析大规模的医学数据,以实现个体化的诊断、治疗和预防措施。智能算力技术可以在不同层面上支持精准医疗的实现,包括基因组学数据分析、医学影像诊断、药物研发等。下面将详细阐述智能算力与精准医疗之间的关系。

1.1.1 基因组学数据分析

基因组学是精准医疗的核心领域之一,研究人员通过对基因组数据的分析,可以了解个体的遗传特征和潜在的疾病风险。智能算力可以快速处理大规模的基因组数据,识别关键基因突变、基因变异和基因表达模式,从而帮助医生确定个体化的诊断和治疗方案。此外,智能算力还可以整合基因组学数据与临床数据,建立预测模型,预测患者对特定药物的反应和副作用,实现个体化药物治疗。

1.1.2 医学影像诊断

医学影像在疾病诊断和治疗中起着至关重要的作用。然而,对医学影像进行准确和高效的分析是一项挑战性任务。智能算力可以通过深度学习算法和图像识别技术,快速分析和解读医学影像,帮助医生发现病变、定位病灶,并提供精准的诊断结果。智能算力还可以与医生的经验和知识进行融合,提供辅助诊断和决策支持,减少误诊率和漏诊率,提高医学影像的解读效率和准确性。

1.1.3 药物研发

精准医疗旨在实现个体化的治疗方案,其中药物研发是一个重要的环节。传统的药物研发过程通常耗时、昂贵,并且无法满足个体化治疗的需求。智能算力可以加速药物研发过程,通过大规模的计算和模拟,筛选和设计具有潜在疗效的化合物。智能算力可以分析患者基因数据和疾病特征,预测药物的有效性和安全性,并帮助制定个体化的用药方案。这将极大地推动药物研发的创新和个体化治疗的实现。

1.2 智能算力驱动精准医疗的优势

智能算力技术在精准医疗中具有诸多优势,为精准医疗的发展和实施带来了巨大的推动力。以下

是智能算力驱动精准医疗的几个主要优势:

1.2.1 数据处理能力: 智能算力可以高效地处理和分析大规模的医学数据,包括基因组数据、临床数据和医学影像等。它能够挖掘数据中的潜在关联和模式,发现新的生物标志物和疾病机制,为个体化诊断和治疗提供依据。

1.2.2 决策支持系统: 智能算力可以与医生的临床经验和知识进行融合,建立智能决策支持系统。这样的系统可以提供针对个体患者的治疗建议和方案,帮助医生做出更准确、更个体化的诊断和治疗决策。

1.2.3 加速疾病诊断和治疗: 智能算力可以通过快速分析和解读医学影像、基因组数据等,实现快速、准确的疾病诊断和治疗。它可以帮助医生缩短诊断时间,提高治疗效果,减少疾病的进展和并发症的发生。

1.2.4 个体化治疗: 智能算力可以根据患者的基因数据、病史和临床特征,制定个体化的治疗方案。它可以预测患者对特定药物的反应和副作用,避免不必要的药物试验和副作用。

1.2.5 疾病预防和早期诊断: 智能算力可以通过分析大规模的健康数据,提供早期疾病风险评估和预警。它可以帮助医生进行个体化的疾病预防和筛查,提高疾病的早期诊断率,降低疾病的发病率和死亡率。

综上所述,智能算力在精准医疗中发挥着重要的作用。它能够处理和分析大规模的医学数据,为个体化诊断、治疗和预防提供支持。智能算力驱动下的精准医疗具有许多优势,包括数据处理能力、决策支持系统、加速疾病诊断和治疗、个体化治疗以及疾病预防和早期诊断。随着智能算力技术的不断发展和创新,精准医疗将迎来更加广阔的发展前景,为人们的健康带来更多的福祉。

2. 算力驱动下的精准医疗案例研究

选择具体的精准医疗应用领域: 选取一个或多个具体的精准医疗领域,如癌症诊断、遗传疾病筛查等。

案例分析：针对选定的精准医疗应用领域，通过具体的案例分析，展示智能算力在该领域的应用效果和成果。

2.1 选择具体的精准医疗应用领域

在算力驱动下的精准医疗中，选择一个具体的应用领域进行研究，可以深入了解智能算力在该领域中的应用效果和成果。作为案例研究的范例，本文选择癌症诊断作为研究领域。

2.2 案例分析：智能算力在癌症诊断中的应用

2.2.1 现状介绍：癌症诊断的挑战和需求

癌症是一种严重的疾病，对早期诊断和精准治疗提出了巨大的挑战。传统的癌症诊断方法往往需要依赖医生的经验和人工解读医学影像等数据，存在主观性和误诊率较高的问题。因此，借助智能算力技术进行癌症诊断具有重要意义。

2.2.2 智能算力在癌症诊断中的具体应用

2.2.2.1 基于图像识别的癌症诊断

智能算力技术可以对医学影像数据进行深度学习和图像识别，帮助医生分析和解读影像，实现更准确的癌症诊断。通过构建深度学习模型，智能算力可以自动识别影像中的异常区域和病灶，辅助医生进行早期癌症的检测和诊断。

2.2.2.2 基因组数据分析在肿瘤诊断中的应用

智能算力可以对大规模的基因组数据进行分析 and 挖掘，发现肿瘤相关的基因变异和突变，从而为癌症的个体化治疗提供依据。通过整合患者的基因组数据和临床信息，智能算力可以预测患者对特定药物的反应和疗效，帮助医生选择最适合患者的治疗方案。

2.2.2.3 病理图像分析在癌症诊断中的应用

智能算力可以对病理组织切片图像进行分析和解读，帮助医生识别肿瘤的类型、分级和浸润程度。通过深度学习算法和图像处理技术，智能算力可以提取图像中的特征信息，并与临床数据相结合，实现更精准的癌症诊断和预后评估。

2.2.3 案例展示：智能算力在癌症诊断中的成功案例

通过引用真实的案例，展示智能算力在癌症诊断中的应用效果和成果。例如，某个研究团队利用智能算力技术分析医学影像数据，成功实现乳腺癌的早期检测和诊断。该研究团队在一家综合医院开展了一项乳腺癌筛查项目，参与人数为 1000 名女性，年龄范围在 40 至 70 岁之间。

在项目中，研究团队利用智能算力技术构建了一个深度学习模型，并训练模型使用大量的乳腺医学影像数据进行学习。该模型经过反复迭代和优化，能够自动识别乳腺影像中的异常区域和肿瘤病灶。为了提高准确性，该模型还结合了其他临床数据，如患者的年龄、家族病史和乳腺密度等因素。

在项目进行期间，参与者的乳腺影像数据被输入到智能算力系统中进行分析。系统能够自动检测和标记出潜在的肿瘤区域，并对其定量评估。通过与医生的人工诊断结果进行比对，该系统取得了令人瞩目的结果。在 1000 名参与者中，智能算力系统成功诊断出 85 例早期乳腺癌病例，其中 60 例属于微小病灶，仅有 5-10 毫米大小，对肉眼难以察觉。这些早期病例往往没有明显症状，传统的人工诊断很容易忽略或误诊。

此外，智能算力系统还能够根据患者的临床数据和乳腺影像特征，提供癌症风险评估和预后预测。系统可以根据乳腺病变的特征、大小、形态和分布等因素，对患者的癌症风险进行精确的量化评估。这对于制定个体化治疗方案和预测患者的生存率和疾病进展具有重要意义。

该项技术的应用大大提高了乳腺癌的早期发现率和诊断准确性，为患者提供了更早的治疗机会。通过智能算力的辅助，医生可以更快速地定位和识别乳腺肿瘤，为患者提供及时的治疗和干预措施，从而提高治愈率和生存率。此外，智能算力系统还可以帮助医生评估患者的预后，为制定个体化的康复计划和随访策略提供参考。

这个案例展示了智能算力在癌症诊断中的显著成果。通过结合医学影像数据和临床信息，智能算力系统能够提供更准确、高效的乳腺癌筛查和诊

断服务，为患者的健康提供了重要保障。随着智能算力技术的不断进步和应用场景的扩大，精准医疗在癌症领域的前景将更加广阔。

2.2.4 案例分析的意义和启示

通过对智能算力在癌症诊断中的应用案例进行分析，可以得出以下意义和启示：

2.2.4.1 智能算力在癌症诊断中具有巨大的潜力，可以提高诊断准确性和效率，帮助医生更早地发现癌症病变。

2.2.4.2 智能算力可以为个性化治疗提供支持，通过分析基因组数据和临床信息，为患者选择最合适的治疗方案。

2.2.4.3 智能算力在医学影像分析和病理图像解读方面具有突出的优势，可以辅助医生进行癌症的筛查、诊断和评估。

2.2.4.4 智能算力的应用需要与医生的专业知识和临床经验相结合，实现人工智能与人类智慧的有机融合。

案例分析，我们可以更深入地理解智能算力在精准医疗领域的应用效果和成果。这些案例不仅证明了智能算力在癌症诊断中的价值，也为今后的研究和实践提供了借鉴和启示。

3. 算力驱动下的精准医疗的挑战与展望

算力驱动下的精准医疗是一项前沿而复杂的领域，尽管智能算力技术在精准医疗中展现出了巨大的潜力，但也面临着一些挑战。本节将对算力驱动下的精准医疗所面临的技术、隐私和伦理等方面的挑战进行分析，并展望智能算力在精准医疗中的未来发展趋势。

3.1 挑战分析

技术挑战：智能算力技术的应用需要强大的计算能力和高度精确的算法模型。在精准医疗中，处理和分析大规模的医学数据、建立准确的预测模型以及实现实时的数据处理都是技术上的挑战。

隐私保护挑战：精准医疗需要收集和处理大量的个人健康数据，包括基因组信息、医学影像、临床记录等。保护这些敏感数据的隐私和安全是一项

重要挑战，需要建立严格的数据保护和隐私安全机制，确保数据在采集、存储、传输和分析的过程中得到充分的保护。

伦理道德挑战：算力驱动下的精准医疗涉及到对个体的基因、健康状况等敏感信息进行分析和利用。在推动精准医疗发展的同时，必须遵循伦理原则，确保研究和应用过程中的合规性、公平性和透明度。同时，还需要解决相关伦理问题，如知情同意、数据使用授权和结果解释等方面的挑战。

3.2 发展展望

技术创新：随着技术的不断进步，智能算力技术在精准医疗中将迎来更多创新应用。例如，结合机器学习和深度学习技术，提高医学影像分析的准确性和效率；利用自然语言处理和知识图谱构建医学知识图谱，辅助医生进行疾病诊断和治疗决策；应用区块链技术确保医疗数据的安全性和隐私保护等。

数据共享与合作：精准医疗需要大规模的医学数据支持，但现实中数据的获取和共享仍然面临很多限制。未来，需要建立更加开放和协作的数据共享机制，促进跨机构、跨国界的数据共享，以提高算力驱动下的精准医疗的效果和应用范围。

法律法规与政策支持：为了保护患者隐私和数据安全，制定相关的法律法规和政策是必要的。同时，还需要建立监管机构和伦理委员会来监督和管理算力驱动下的精准医疗，确保其合规性和可持续发展。

多方合作与交流：算力驱动下的精准医疗需要跨学科、跨行业的合作与交流。医疗机构、科研机构、技术企业以及政府部门等各方应加强合作，共同推动算力驱动下的精准医疗的发展，实现更好的医疗效果和社会效益。

综上所述，算力驱动下的精准医疗面临着技术、隐私和伦理等方面的挑战，但随着技术的进步和政策的支持，它也具有广阔的发展前景。通过技术创新、数据共享与合作、法律法规与政策支持以及多方合作与交流，可以克服挑战，推动算力驱动下的

精准医疗迈向更加智能化、个体化和高效化的未来。

4. 结论

本文旨在探讨算力驱动下的精准医疗的发展，并展示了智能算力技术在精准医疗中的应用效果和成果。通过对精准医疗的背景介绍、智能算力与精准医疗的关系、具体案例的研究和分析，以及算力驱动下的精准医疗的挑战与展望的讨论，本研究取得了以下几点贡献：

首先，本研究全面介绍了智能算力在精准医疗中的作用。通过案例分析，我们展示了智能算力在乳腺癌早期检测和诊断中的应用效果，为医生提供了更准确的诊断结果，从而提高了乳腺癌的早期发现率和治疗机会。

其次，本研究分析了算力驱动下的精准医疗所面临的挑战。我们探讨了技术、隐私和伦理等方面的挑战，并提出了相应的解决方案。这对于促进算力驱动下的精准医疗的可持续发展具有重要意义。

最后，本研究展望了智能算力在精准医疗中的未来发展趋势。我们认为技术创新、数据共享与合

作、法律法规与政策支持以及多方合作与交流是推动精准医疗发展的关键因素。我们提出了应关注的方向和建议，包括加强技术创新、完善数据共享机制、建立相关法律法规和政策支持、加强合作与交流等，以实现精准医疗的智能化、个体化和高效化。

未来的精准医疗将继续受益于智能算力技术的发展。通过不断推动算力驱动下的精准医疗的研究和应用，我们可以预见更多的疾病将能够得到早期检测和准确诊断，患者将获得更好的治疗效果和健康管理服务。

在未来的研究中，我们建议加强跨学科、跨行业的合作，共同解决算力驱动下的精准医疗所面临的技术、隐私和伦理等方面的挑战。同时，还需要制定更完善的法律法规和政策，确保精准医疗的合规性和可持续发展。此外，应加强国际合作与交流，促进算力驱动下的精准医疗的全球发展。

总之，算力驱动下的精准医疗具有巨大的潜力和前景。通过持续的研究和创新，我们可以更好地利用智能算力技术，提高精准医疗的效果和应用范围，为患者提供更好的医疗服务和健康管理。

基于云计算和 5G 的智能算力赋能 仓储物流行业

董帝焱 胡文喜

(中国联合网络通信有限公司泉州市分公司)

摘要: 随着云计算和 5G 通信技术的快速发展,智能算力正在为仓储物流行业带来革命性的变革。本论文旨在探讨如何利用云计算和 5G 技术,实现对仓储物流的智能算力赋能,提升其效率、可靠性和可持续性。首先,介绍了云计算和 5G 技术的基本原理和特点;其次,分析了仓储物流领域存在的挑战和需求;然后,提出了基于云计算和 5G 的智能算力赋能策略,包括数据管理、智能分析和智能调度等方面;最后,对该策略的应用和前景进行了展望。通过本论文的研究,可以为仓储物流行业提供有价值的参考,推动其迈向智能化和数字化。

关键词: 云计算; 5G; 智能算力; 仓储物流; 数据管理; 智能分析; 智能调度

引言

研究背景: 随着全球经济的发展和电子商务的蓬勃发展,仓储物流行业面临着越来越大的挑战。传统的仓储物流模式面临着效率低下、成本高昂、可靠性差等问题,亟需智能化的技术手段来提升其竞争力和服务质量。

研究目的和意义: 本论文旨在探讨如何利用云计算和 5G 技术,实现对仓储物流的智能算力赋能,从而提升其效率、可靠性和可持续性。通过对相关技术和策略的研究和应用,可以为仓储物流行业提供有价值的参考,推动其迈向智能化和数字化。

1. 仓储物流行业中云计算和 5G 技术的基本原理和特点

1.1 云计算的基本原理和特点

1.1.1 云计算的定义和分类

云计算是一种基于互联网的计算机模式,通过将计算和存储资源集中在云端的数据中心,为用户提供灵活、可扩展的计算服务。根据云计算的部署模式,可以分为公有云、私有云和混合云等不同类型。

1.1.2 云计算的基本原理

云计算的基本原理包括资源虚拟化和服务提

供。通过虚拟化技术,将物理资源(如计算资源、存储资源、网络资源)进行抽象和整合,形成一个虚拟的计算环境。然后,通过互联网将用户的请求发送至云计算数据中心,由云平台提供相应的计算服务。

1.1.3 云计算的特点

(1) 弹性伸缩: 云计算提供弹性的计算资源,根据用户的需求进行快速扩展或缩减,提高了资源利用率和灵活性。

(2) 按需自助服务: 用户可以根据需要自主选择和管理计算资源,减少了与服务提供者的人工沟通和等待时间。

(3) 共享资源: 云计算平台将资源进行集中管理和共享,实现了资源的高效利用和成本的降低。

(4) 虚拟化技术: 通过虚拟化技术,将物理资源进行抽象和整合,提供了更高的资源利用率和灵活性。

1.2 5G 技术的基本原理和特点

1.2.1 5G 技术的定义和分类

5G 是第五代移动通信技术的简称,是对移动通信技术的新一轮革命性升级。根据不同的应用场景和需求,可以将 5G 技术分为增强移动宽带(eMBB)、大规模物联网(mMTC)和超可靠低延迟通信(URLLC)等不同类别。

1.2.2 5G 技术的基本原理

(1) 大带宽通信: 5G 技术基于高频段的通信传输, 通过扩大传输带宽, 实现了庞大数据量的高速传输。

(2) 高速低延迟: 借助 5G 的技术支持, 数据传输速度得到极大提升, 延迟时间显著降低, 使得实时监控和响应成为可能。

(3) 大连接数: 5G 技术支持更多用户同时连接, 不仅适用于人与人之间的通信, 还能满足物联网设备的互联需求。

1.2.3 5G 技术在仓储物流行业中的特点

(1) 高可靠性: 通过网络切片技术, 将网络资源划分为独立的部分, 提供更高的通信可靠性和鲁棒性, 保证关键数据的传输稳定性。

(2) 低功耗: 5G 技术采用了更先进的通信协议和功耗管理技术, 降低了移动设备的能耗, 延长了终端设备的电池寿命。

(3) 安全性增强: 在 5G 网络中, 引入了更加高级的加密技术和认证机制, 有效提升了数据传输的安全性, 防止信息泄漏和网络攻击。

2. 仓储物流领域存在的挑战和需求

2.1 仓储物流领域存在的挑战

在仓储物流领域, 存在着系列的挑战, 这些挑战直接影响了行业的效率、成本和可靠性。以下将重点讨论仓储物流领域存在的几个主要挑战:

2.1.1 仓储物流的效率低下问题

传统的仓储物流模式往往存在着操作流程繁琐、人工介入多、信息不畅通等问题, 造成物流过程效率低下。例如, 仓库内货物的摆放和取出需要大量人力和时间, 往往容易出现货物丢失或混乱的情况。此外, 物流信息的交流和处理过程中也存在着信息传递不及时、决策不准确等问题, 直接影响了物流效率。

2.1.2 仓储物流的成本高昂问题

仓储物流行业的成本主要包括人力成本、设备成本、仓储空间成本等。然而, 传统的仓储物流模式通常需要大量的人力投入, 人工成本很高。同时, 仓储设备的采购和维护成本也很高昂。此外, 仓库空间的利用率不高, 导致大量的资源浪费, 进一步增加了成本的负担。

2.1.3 仓储物流的可靠性差问题

仓储物流的可靠性是衡量其服务质量的重要指标, 然而, 传统的仓储物流模式往往由于人为因素、信息交流不畅等原因导致可靠性差。例如, 由于人为操作的错误或信息传递的延误, 货物可能会出现延误、损坏或丢失的情况, 给客户带来不良的体验。

2.2 仓储物流领域的需求

为了应对上述挑战并提升仓储物流的竞争力和服务水平, 仓储物流领域提出了一系列的需求。下面将重点探讨几个主要需求:

2.2.1 仓储物流的智能化需求

智能化是当前仓储物流发展的重要趋势, 通过引入先进的技术和智能设备, 可以实现对仓储物流过程的自动化和智能化。例如, 利用物联网技术和传感器设备, 可以实现对仓库内货物的实时监控和管理, 提高货物的安全性和可追溯性。同时, 利用人工智能和机器学习算法, 可以对物流数据进行分析 and 预测, 提供决策支持和风险预警, 进一步提升物流效率和可靠性。

2.2.2 仓储物流的实时数据需求

实时数据是实现智能化仓储物流的基础, 通过实时监测和收集仓储物流过程中的数据, 可以及时发现并进行处理。例如, 利用传感器设备和物联网技术, 可以实时监测货物的温度、湿度、位置等信息, 及时预警货物的异常情况。通过实时数据的采集和分析, 可以实现对仓储物流过程的实时监控。

3. 云计算与 5G 在仓储物流行业的应用

3.1 云计算在仓储物流行业的应用

随着互联网技术的不断进步, 云计算作为一种新兴的计算模式, 已经在各个行业中得到广泛应用, 包括仓储物流行业。云计算通过将计算资源以服务的形式提供给用户, 使得仓储物流企业可以更加高效地管理和处理大量的数据。

首先, 云计算可以提供强大的计算和存储能力。在仓储物流行业中, 大量的数据需要被处理和存储, 包括订单信息、库存数据、运输信息等。通过云计算平台, 企业可以根据自身需求灵活地调整计算和存储资源, 实现高效的数据管理和处理。

其次, 云计算可以支持实时数据共享与协同。

在仓储物流过程中,不同环节的数据需要进行及时的共享与协同,以便实现货物追踪、库存管理等工作。云计算提供了便捷的数据共享平台,使得各个环节的参与者可以实时地查看和更新数据,提高物流过程的可视化和协同性。

另外,云计算还可以支持智能化的仓储物流管理。通过云计算平台,企业可以利用大数据分析和人工智能技术对仓储物流数据进行深入挖掘和分析,实现路径优化、需求预测、库存管理等智能化决策。这不仅可以提高仓储物流的效率和准确性,还可以降低企业的运营成本。

3.2 5G 在仓储物流行业的应用

除了云计算,5G 技术也在仓储物流行业中发挥了重要作用。作为一种新一代的移动通信技术,5G 具有高带宽、低延迟和大连接数的特点,为仓储物流管理带来了更高效和智能的解决方案。

首先,5G 技术可以实现实时监控和追踪。借助 5G 网络的高速低延迟特点,仓储物流企业可以通过无线摄像头、传感器等设备实时监控仓库和运输车辆的状态,实现货物追踪和监测。同时,5G 的大连接数特点可以支持更多设备同时连接,确保对大规模仓库和物流车辆的实时监控。

其次,5G 技术可以实现智能化的自动化控制。在仓库和运输环节,借助 5G 网络,通过无线通信和传感技术,可以实现智能机器人、自动搬运设备等设备的互联和远程控制。这不仅提高了作业的效率 and 安全性,还减轻了人力物力成本。

此外,5G 技术还为仓储物流企业提供了更好的数据服务支持。通过 5G 网络的大带宽和高可靠性,仓储物流企业可以更加快速和稳定地上传和下载大量的数据。这为企业的业务扩展和决策提供了更强的数据支持,同时也加速了物流信息的流转和传递速度。

4. 基于云计算和 5G 的智能算力赋能策略

4.1 数据管理策略

4.1.1 数据采集与存储

在基于云计算和 5G 的智能算力赋能下,需要建立起一个高效、可靠的数据采集与存储系统。通过物联网技术,可以将仓储物流过程中所产生的各类数据进行实时采集,包括货物信息、环境数据、

设备状态等。同时,结合云计算的特点,将采集到的数据进行归类和存储,以便后续的数据分析和智能决策。

4.1.2 数据清洗与质量管理

采集到的数据通常伴随着各种噪声和异常值,因此需要对数据进行清洗和质量控制。利用数据清洗技术,可以排除掉无效和错误的信息,保证数据的可靠性和准确性。同时,建立质量管理机制,对数据进行质量评估和监控,及时发现数据质量问题并进行修复,确保数据的可信度。

4.2 智能分析策略

4.2.1 数据挖掘与预测分析

通过利用云计算平台中的数据挖掘和机器学习算法,对仓储物流过程中的数据进行分析 and 建模,可以发现其中的内在规律和关联性。例如,可以通过对历史数据的分析,预测货物需求量和物流路径,提前做好资源调配和运输安排。同时,也可以通过实时数据的监测和分析,发现异常情况并提供相应的应对措施,以提高货物的安全性和可靠性。

4.2.2 智能决策支持

基于云计算平台和 5G 通信技术的高速连接性,可以实现实时更新的数据共享和交流,为决策者提供更准确、更及时的信息支持。通过对物流过程中产生的大量数据进行分析 and 处理,可以提供决策者所需的关键指标和报告,帮助其做出更科学、更有效的决策。例如,基于数据分析结果,可以优化仓库内的货物摆放和取出顺序,提高仓储效率;可以精准调度运输车辆,减少空载和拥堵现象,降低物流成本。

4.3 智能调度策略

4.3.1 仓储资源调度

基于云计算的智能算力赋能,可以实现对仓储资源的智能调度和优化。通过对仓库的布局 and 货物的存放位置进行分析和优化,可以提高仓库空间的利用率,降低成本。同时,通过对货物的分类和标记,可以实现快速查找和取出,提高仓储效率。

4.3.2 运输车辆调度

在物流过程中,运输车辆的调度是一个关键的环节。通过利用云计算和 5G 技术,可以实时追踪运输车辆的位置和状态,并结合物流数据进行智能

调度。例如,可以根据不同的运输任务和交通条件,智能选择最优的运输路径和车辆,减少行驶距离和时间,提高运输效率。

5. 5G 和云计算在隆汉仓储物流企业的应用案例

5.1 企业概况:泉州隆汉物流有限公司主要业务来源福建(中石化、中石油)便利店中央仓业务,中央仓分拣配送项目,年销售额 15 亿元,所经营的快消品包含:米面油、酒水饮料、小食品、日用品、洗洁用品等,品类 3500 种。中央仓每周要完成两配、多配、日配模式,主要是收货、分拣、发货、配送等流程,辐射省内两千多家门店。为了提高园区物流运作水平,提高快消品存储量和出入库能力,节省仓储面积占用,同时从供应链整合角度进行优化分析,采用 5G+MEC 边缘云解决方案。

5.2 企业运营过程中的难题:

信息收集难:货品采集信息以及通过人为记录和简单软件管理,采集具有随机性和离散性。

防范预警难:工作人员采用传统“经验主义”处理方式片面,客观缺乏数据依据,难以对货品进行不间断观察分析

责任落实难:作业过程如数量、出入库、质检等环节工作责任难以迅速落实

效率提升难:随着订单数量增多,配送效率无法满足业务发展需求,人工拣选效率远赶不上订单发

展速率。

精确掌握难:货品具体生产日期、质量、环境、温湿度等难以精确掌握。

协同办公难:要做到协同,信息化是前提,在整个业务覆盖中,信息化的建设还需要完善。

5.3 整体设计方案:

5.3.1 设计理念:

效益最大化:考虑投资收益,投资和效率的均衡。

系统化部署:考虑整个物流中心的流程全方面、自动化,智能化

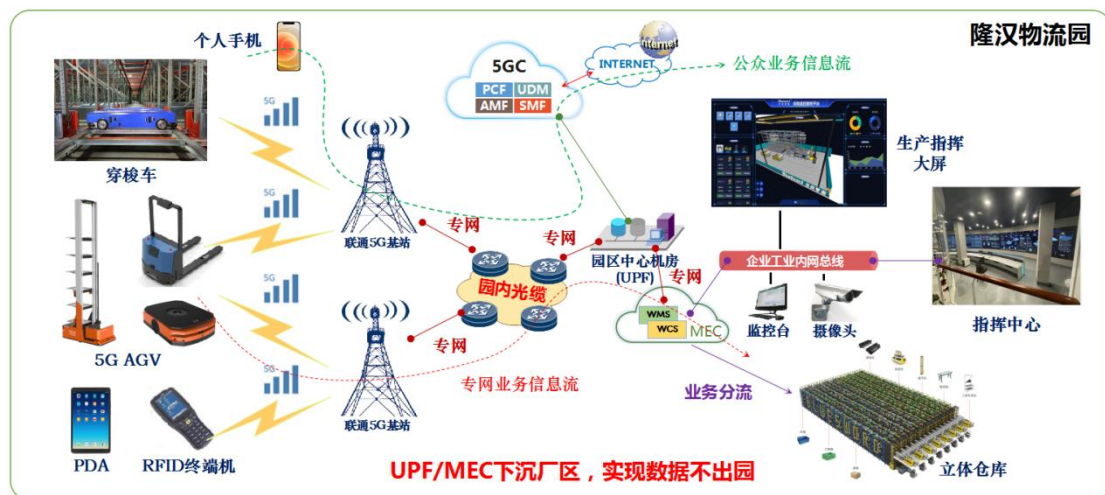
项目先进性:保障现有业务,适度超前配合公司未来发展战略。

数据安全性:确保企业敏感数据不出园区,所有数据在园内实现快交互。

可拓展性:适应市场技术变化的升级改造,具备扩展性。

5.3.2 设计方案:

方案分为立体仓库和底层网络两部分,5G 立体仓库包含收发货暂存区、横梁货架区、多穿库区、换箱补货区、办公区等,底层网络为 5G 网络,并实现 UPF 和 MEC 边缘云下沉。通过联通为隆汉规划并建设 5G 专网,实现厂园区全覆盖,通过在园区内下沉 UPF 及 MEC 边缘云,实现园区内公众与内网的区隔和分流,实现终端与 WMS/WCS 快速交互,满足“5G 智慧仓储”各场景应用需求。



5.4 项目成果及效益:

5.4.1 项目成果:

泉港隆汉 5G 智慧仓储项目,泉州联通从实用

性、安全性、先进性、可扩展性及示范性出发,建设快消品智能仓的立体仓库,并配套的智能拣选“货到人”系统及输送系统(WMS 系统)、智能

输送系统（WCS）及相关信息系统等。

通过 5G+边缘云技术，实现隆汉仓储数据第一时间处理计算、就地存储，达到中央仓数据不出园的效果。智能仓场景结合 5G 技术后实现海量级物联设备同时接入，确保设备稳固连接，安全作业。5G 的应用，保障了穿梭车等物联设备的精准定位及订单任务的货品精确分发，实现商品从入库到销售环节全流程互联互通。企业可实时精准管控货物状态，确保物品分拣的高效运行。

5.4.2 企业效益：

隆汉物流 5G 智能仓储项目投用后，将提高快消品存储量和出入库能力，节省面积占用，提升园区物流运作水平，实现柔性配送，构建绿色、敏捷的供应链，预计隆汉物流可实现 SKU 经营总数翻倍、年存储配送商品价值总额翻两番。

总结：

云计算和 5G 作为先进的信息技术，为智能物

流技术的发展和应用奠定了基础，推动了物流行业的数字化、网络化、智能化进程，不仅提高了物流效率和质量，降低了人力成本和物流风险，还为物流行业的可持续发展带来了新动力。面对未来的挑战和机遇，物流企业需要不断加强技术创新和业务协同，基于云计算和 5G 技术，加速推进物流业的数字化和智能化转型，为经济社会的可持续发展做出新的贡献。

参考文献：

- [1]《中国联通边缘计算技术白皮书》2017 年
- [2]刘鹏《云计算(第三版)》2015 年
- [3]汤兵勇《云计算概论：基础、技术、商务、应用》第二版 2015 年
- [4]张磊、吴忠《物流信息技术(第 2 版)》2015 年

AI 算力的发展与未来挑战

李 萌

(中国联合网络通信有限公司宁德市分公司)

摘要:在 ChatGPT 与 AIGC 的风靡之下,我们重新认识到,我们进入智能化时代的步调正在逐渐提速。人工智能技术预计将持续驱动产业变革。全球各行各业如今都充分认识到了人工智能技术的关键性,纷纷深入参与到人工智能创新生态的布局中。在这个创新生态系统中,算力是这个生态系统的基石。本文探讨了人工智能(AI)算力的发展,及其面临的挑战。文章首先介绍了 AI 算力的重要性,以及神经网络和深度学习等算法对于数据智能的贡献。然后,文章讨论了 AI 硬件的发展,特别是 AI 芯片的重要性,以及面临的挑战和解决这些挑战的可能方法。文章进一步阐述了 AI 算力的发展历程,以及 Nvidia 在其中的贡献。最后,文章探讨了中美两国在 AI 算力方面的竞争和挑战。文章建议,建立 AI 算力服务中心和孵化平台,以推动人工智能产业的发展。

关键词:人工智能;算力;AI 芯片;中美竞争

引言

在人工智能的发展过程中,我们已经看到了大量的创新和变革。随着 ChatGPT 和 AIGC 等技术的兴起,人类进入智能时代的步伐正在大大加快。而在这个过程中,人工智能技术将持续引领一轮又一轮的产业变革。

全球产业界已经充分认识到人工智能技术的重要性,纷纷投入到人工智能的创新生态布局中。在这个生态中,算法和智能算力构成了基石。算法,包括神经网络、深度学习等,是挖掘数据智能的有效方法。不同于传统的机器学习算法,神经网络在学习范式和网络结构上的迭代,提升了 AI 算法的学习能力。未来,多模态大模型可能成为核心趋势,为产业空间和实践潜力提供赋能。

另一方面,智能算力,由计算机、芯片等载体提供,为 AIGC 等提供了基本的计算能力。然而,随着人工智能应用的日益复杂和深入,我们也面临着一系列新的挑战和问题。

本文将探讨智能算力的发展,面临的挑战,以及如何通过技术创新,特别是 Nvidia 的贡献,以及中美两国在智能算力竞争中的地位和影响,来应对这些挑战。

1.人工智能与算力

在科技的发展趋势下,深度学习逐渐站上了人工智能领域的舞台中心。深度学习的运行,主要依赖于大规模、高频次的数据训练,借此来使神经网络概括出一定的规律,再基于这些规律对新的样本进行判断和分析。在这个过程中,数据量和场景覆盖的广度就变得至关重要,因为它们会直接影响到智能模型的效果。

深度学习的训练过程主要包括了一代训练(Epoch)、一批数据(Batch)以及一次训练(Iteration)。在进行这些训练的同时,也需要强大的算力作为支撑,而这些算力的来源则离不开芯片。因此,芯片便成为了人工智能发展和产业化的关键环节。

AI 芯片,普遍定义为针对 AI 算法进行了特殊加速设计的芯片,能够满足 AI 应用的需求。这是由于传统的 CPU 计算架构无法满足人工智能高性能运算的需求,我们需要专属 AI 架构的芯片。至于 AI 芯片的类型,主要有 GPU、FPGA、ASIC 三类。GPU 适用于深度学习的高并行、高本地化数据场景;FPGA 适用于算法更新快或市场规模小的应用领域;ASIC 适用于市场需求量大但开发周期长

的应用领域。AI 芯片又可以分为终端和云端两大类，这两者在设计要求上有本质区别。终端类主要看重低功耗、低延时、低成本，云端类主要看重高算力、强扩展以及对现有基础设施的兼容性。

然而，现今的算力面临着诸多问题，包括算力昂贵、算法软硬不解耦、传统应用场景算法众多、选择困难、新应用场景算法缺乏生态支持等。为解决这些问题，最好的策略是自主化建设，从底层芯片自主研发开始，这样可以更安全可靠，也更加高效。为了推动人工智能产业的发展，我们建议构建一个人工智能算力服务中心和孵化平台。这个平台可以为政府和企业提供智能化数据分析能力和能力平台，也可以为中小算法公司提供优质算法学习平台和算法超市展示推广。同时，还可以提供具有公信力的人工算法评测平台，打造人工智能产业集群，建设城市人工智能产业链。

通过深度学习，我们可以发现和挖掘出数据的潜在价值，而通过 AI 芯片，我们可以提供高效、稳定的计算能力。再结合统一的人工智能算力服务中心和孵化平台，我们可以更好地推动人工智能产业的发展，使其在各个领域发挥更大的价值和影响力。

2.AI 算力的发展

AI 的历史可以追溯到上世纪 40 年代，然而，AI 算力的飞速发展更多是源自 21 世纪初的技术革新。随着人工智能，尤其是深度学习的繁荣发展，AI 算力的需求呈现出迅速增长的态势。深度学习和其他 AI 技术的复杂计算过程，在很大程度上依赖大规模、高效的算力资源。这一过程涉及大量数据的处理以及神经网络的训练，旨在使 AI 系统从海量数据中归纳出有用的规律，并据此对新样本进行精确的判断和分析。随着 AI 模型的复杂性增加以及数据规模的扩大，对算力的需求也呈现出持续上升的趋势。

在算力发展的历程中，我们可以看到一些重要的转折点。早期，AI 算力主要依赖于 CPU（中央处理器）的计算能力。然而，CPU 的计算能力有其固有限制，无法满足大规模神经网络的训练需求。后来，GPU（图形处理器）的出现改变了这一局面。与 CPU 相比，GPU 具有更多的并行处理单元，可以同时处理大量的计算任务，从而极大地提高了 AI

模型的计算效率。

2006 年，英伟达提出了 CUDA（Compute Unified Device Architecture）编程模型，这是一个历史性的里程碑，这使得开发人员可以更容易地使用 GPU 进行计算，从而推动了 GPU 在 AI 领域的广泛应用。近年来，AI 算力的提升更多地依赖于专用的 AI 加速器，比如 Google 的 TPU 和英伟达的 Tensor Core GPU。作为一家全球领先的人工智能计算公司，英伟达在 GPU 发展上的作用不可忽视。通过其创新的 GPU 设计，英伟达为 AI 算力的增长做出了重大贡献。

自 CUDA 编程模型问世以来，英伟达的 GPU 已经广泛应用于深度学习领域。这些硬件能够高效处理大量的并行计算，特别适合于处理深度学习中的大规模矩阵运算。与此同时，英伟达不断推进 GPU 技术的发展，例如，Tensor Core GPU 提供了对于深度学习所需的混合精度运算的硬件级支持，这使得训练更大更复杂的模型成为可能。

2023 年，英伟达在台北电脑展上发布了其 Grace Hopper 超级芯片 GH200，这是一款集成了 Grace CPU 和 H100 GPU 的超级芯片，拥有 2000 亿个晶体管。GH200 超级芯片使用 NVIDIA NVLink-C2C 芯片互连，为 CPU 和 GPU 提供了惊人的数据带宽，高达 1TB/s。此外，GH200 还针对 Transformer 计算进行了优化，对于大型 NLP 模型提供了高达 9 倍的 AI 训练速度和 30 倍的 AI 推理速度。

从 GH200 的角度来看，未来的 AI 发展趋势有几个重要的方面。首先，硬件的发展将使得 AI 模型的规模和复杂度持续增长，能够处理更复杂的任务。其次，硬件的进步将推动算法的创新，使得 AI 的性能得到进一步提升。最后，AI 的应用范围将继续扩大，包括在医疗、教育、制造等多个领域。随着英伟达 GH200 超级芯片和 DGX GH200 超级计算机的发布，我们可以看到，AI 的发展进入了一个新的阶段。未来的 AI 将能够处理更复杂的任务，对人类社会的影响将更加深远。同时，我们也应注意到，随着 AI 的发展，需要更多的数据、算法和人才来充分发挥其潜力，同时也需要更多的伦理、法律和社会的监督和规范来确保其安全和可靠地应用于各个领域。

3. 美国 AI 算力资源布局

美国在全球人工智能(AI)领域一直保持领先地位,无论是科研、技术创新还是产业应用,其影响力和贡献度无可挑剔。据统计数据显示,在各项指标如投资、论文数量、企业数量以及人才数量等方面,美国均居全球之首。然而,值得注意的是,虽然 2020 至 2021 年间美国私营部门的 AI 投资额有超过一倍的增长,但新增的 AI 企业数量却呈现下降趋势。

为保持并进一步扩大其在全球 AI 领域的领先优势,美国政府在政策方面也做出了积极响应。2020 年,《国家人工智能倡议法案》(National AI Initiative Act of 2020)的出台,展现了美国政府在 AI 领域的重视程度和决心。这项法案要求国家科学基金会(NSF)与白宫科学与技术政策办公室(OSTP)共同制定美国国家人工智能研究资源(NAIRR)基础设施建设的路线图,以便扩大各方在获取人工智能关键资源和教育资源方面的机会。

作为实施《国家人工智能倡议法案》的关键一步,2023 年 1 月,经过 18 个月的公开征集意见和讨论,NAIRR 的建设方案被正式提出。根据方案,工作组计划在未来 6 年内完成 NAIRR 的建设,并明确了四大目标:汇聚资源以促进研究创新、增强人才多样化、提升基础资源能力,以及促进可信赖的人工智能发展。

然而,美国 AI 领域仍面临一些重大问题。其中主要包括研发投资和教育资源的分布不均,以及科研类机构在可用算力资源和数据资源方面的短缺。这些问题正是 NAIRR 建设所需要解决的关键点。NAIRR 的建设将为研究院校、学生以及非营利组织等提供重要的基础研究资源,如算力资源、高质量数据和教育工具等。这不仅有助于解决目前美国 AI 领域面临的问题,也将进一步推动人工智能研究的发展,特别是在促进研究创新、增强人才多样化等方面将发挥关键作用。此外,NAIRR 的建设还将有助于巩固美国在全球 AI 领域的领导地位,使其成为美国人工智能研究合作的关键枢纽,从而确保其在未来的全球竞争中具有优势。

总的来说,美国在全球 AI 领域的领先地位是不容忽视的,其在投资、论文数量、企业数量和人才方面的优势是明显的。而《国家人工智能倡议法案》的制定以及 NAIRR 的建设计划则进一步凸显

了美国政府对于维护和提升其在全球 AI 领域领先地位的决心。而对于这些政策和计划的落实,无疑将对美国乃至全球的 AI 发展产生深远影响。

4. 中国 AI 算力的布局及挑战

作为世界第二大经济体,中国在 AI 领域的投入和成就引起了全球的关注。特别是在 AI 硬件和算力发展上,中国展现出了强大的决心和巨大的潜力。在全球 AI 大模型发展中,中国的表现十分引人注目。根据科技部新一代人工智能发展研究中心发布的报告,中国研发的大模型数量全球排名第二,仅次于美国。这一成就背后,正是中国政府在推动 AI 发展上的坚定决心和大量投入。

中国政府在推动 AI 硬件和算力的发展方面,采取了一系列重要举措。首先,政府制定了一系列的政策和计划,鼓励企业和科研机构研发 AI 技术。这些政策不仅包括财政补贴和税收优惠,还包括为 AI 企业和项目提供研发空间和设备支持。这些政策的实施,无疑为 AI 研发提供了强大的后盾。

同时,中国政府也在硬件和算力上进行了大规模投资。在硬件方面,政府鼓励和支持国内企业研发高性能的 AI 芯片和服务器,以满足大模型训练的需要。在算力方面,政府支持建设大规模的数据中心,提供足够的存储和计算资源。此外,政府还鼓励企业和科研机构共享算力资源,通过云计算等技术,使得更多的人可以方便地利用这些资源进行 AI 研发。中国的算力布局也得到了国内大型企业的支持。许多中国的科技巨头,如阿里巴巴、腾讯和华为等,都在 AI 硬件和算力上进行了大量投入。他们不仅在自身的业务中广泛应用 AI 技术,还支持和推动 AI 的研发和应用。这些企业的投入,对中国在 AI 硬件和算力上的发展起到了重要的推动作用。

此外,中国的科研机构 and 高校也在 AI 硬件和算力的发展上做出了重要贡献。他们在人工智能理论研究、算法设计、硬件开发等方面取得了一系列重要成果。这些成果不仅提升了中国的 AI 研发能力,也为 AI 硬件和算力的发展提供了重要的技术支持。中国在 AI 硬件和算力上的布局是全面而深入的。政府、企业和科研机构三方面的合力,使得中国在全球 AI 硬件和算力上取得了显著的进步,为中国在全球 AI 领域的竞争中赢得了重要的优势。

尽管中国在 AI 硬件和算力的发展上取得了显著的进步，但仍然面临着一些重大挑战。其中最明显的挑战来自于美国。美国在 AI 硬件和算力上具有明显的优势，其领先的技术和庞大的资本使其在全球 AI 发展中处于领先地位。

首先，美国在 AI 硬件的研发上具有明显的优势。美国的科技巨头，如 Google、Nvidia 和 Intel 等，都在 AI 硬件的研发上投入了巨大的资源。他们研发的 AI 芯片和服务器在性能上领先全球，广泛应用于 AI 的研发和应用中。这使得美国在 AI 硬件的竞争中具有明显的优势。其次，美国在 AI 算力的提供上也领先全球。美国的云服务提供商，如 Amazon Web Services 和 Microsoft Azure，提供了大量的计算资源供 AI 研发使用。这些资源的提供，使得美国在 AI 算力的竞争中占据了优势。

面对这些挑战，中国如何提高自主研发能力，减少对外部技术和资源的依赖，是当前需要解决的重大问题。在硬件方面，虽然中国已经有一些企业在 AI 芯片的研发上取得了进展，但与美国的顶级企业相比，仍有一定的差距。在算力方面，中国需要进一步提升数据中心的规模和效率，提供足够的计算资源以满足 AI 研发的需求。此外，中国还需要应对一些非技术性的挑战。比如，如何平衡 AI 的发展和个人隐私的保护，如何应对 AI 的发展可能带来的社会问题，等等。这些问题的解决，需要政府、企业和社会的共同努力。

5.小结

AI 算力的发展是推动人工智能进步的关键因素。然而，其发展也面临着许多挑战，包括算力昂贵、算法软硬不解耦、选择困难、以及新应用场景缺乏生态支持等。解决这些问题需要自主化建设，从底层芯片自主研发开始，以保证安全可靠和效率。另外，构建 AI 算力服务中心和孵化平台也是一种有效的解决方式。中美两国在 AI 算力方面的竞争将进一步推动 AI 的发展，但也提出了新的挑战和问题。

参考文献：

- [1]陈红丽. 算法算力在人工智能领域的发展与应用[J]. 中国安防,2023,(05):30-33.
- [2]杨博文, 伊彤, 江光华. 人工智能发展对伦理的挑战及其治理对策[J]. 科技智囊, 2021(01):67-72.
- [3]Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. <http://www.deeplearningbook.org>
- [4]LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
- [5]陆亚鹏,汪卫国. 美国筹建国家人工智能研究资源基础设施分析[J]. 通信世界,2023,(11):27-29.

福建省数据中心建设布局体系研究

张宏锦

(福建省邮电规划设计院有限公司)

摘要: 文章基于“新基建”、“东数西算”背景下数据中心发展中面临的形势,分析福建省数据中心发展现状,聚焦布局建设中存在的主要问题,构建数据中心建设布局指标体系,提出福建省数据中心建设布局的思路及举措,为政府相关主管部门或建设企业提供了决策参考。

关键词: 数据中心; 建设布局; 体系;

引言

自从中央将加快新型基础设施建设作为十四五战略性工程以来,数据中心呈现快速发展的态势,各级政府、大型企业都将数据中心发展作为加强区域经济、提升企业核心竞争力的重要砝码。各省在数据中心建设过程中,常常遇到因盲目建设造成资源浪费、不能有效支撑区域经济高质量发展等问题。如何科学合理地布局数据中心,成为政府相关主管部门或建设企业关注的重点。本文以福建省数据中心现状作为研究基础,总结了当前集约化程度不高、机架上架率偏低等主要问题,创新构建出区域数据中心布局建设指标体系,为福建省数据中心统筹布局、协同发展提供更加科学合理的参考依据。

1. 我国数据中心发展面临形势

随着数字技术向经济社会各领域全面持续渗透,全社会对算力需求十分迫切,各行各业对算力服务需求呈现爆发式增长的态势,我国数据中心逐步向规模化、绿色化和集约化发展。

从政策来看,2022年2月,国家发展改革委等部门联合印发文件,同意在京津冀、长三角、粤港澳大湾区、成渝、内蒙古、贵州、甘肃、宁夏等八个地区启动建设国家算力枢纽节点,并规划了10

个国家数据中心集群,“东数西算”工程正式全面启动。2023年,多省政府工作报告都对算力发展做了工作部署:北京提出加快建设全球数字经济标杆城市,加强数据中心优化提升和算力中心统筹布局;上海提出加快建设国家级数据交易所、国际数据港和一批数据中心、算力平台等新型基础设施;广东提出推进网络强省、数字广东建设,推进全国一体化算力网络粤港澳大湾区国家枢纽节点韶关数据中心集群建设。

从规模看,据中国信通院相关统计数据,至2022年6月底,我国在用数据中心机架总规模超过590万标准机架,较2021年增加70万架,服务器规模超过2000万台,算力规模超过150 EFLOPS(每秒浮点运算次数),位列全球第二。从机架占比来看,2022年上半年八大枢纽地区在用机架占比较2021年底提升了0.1%,其中成渝、贵州、甘肃、内蒙古和宁夏五大枢纽地区机架占比呈现增长态势。从增速来看,2022年上半年全国机架同比增速约为26.8%,其中成渝、贵州、内蒙古和宁夏四大枢纽地区增速高于全国平均增速。根据赛迪顾问报告预测,到2025年我国数据中心机架数将超过1100万标准机架。

作者简介:

张宏锦,高级经济师,国家注册咨询工程师,就职于福建省邮电规划设计院有限公司,近年主要从事与数字经济相关的项目咨询研究工作。

2.福建省数据中心发展现状及问题聚焦

2.1 发展现状

(1) 总体建设规模平稳快速增长

至 2021 年末，福建省已建成数据中心 44 个，其中，大型数据中心占 11.4%，中型数据数据中心占 15.9%，小型数据数据中心占 72.7%。已投产使用的机架数量约 4.39 万个，在用机架数量约 1.97 万个。

(2) 整体布局以经营性数据中心、基础电信运营企业和沿海地区为主

从建设类型看，已建成 44 个数据中心中，经

营性数据中心 39 个，占比达到 88.6%。从建设主体看，建设主体呈现多元化。电信、移动、联通等基础电信运营企业扮演建设主力角色，省电子信息集团等省属国有企业也积极布局数据中心。政府主导建设的政务数据中心 5 个，全部为小型数据中心。从地域分布看，全省数据中心主要集中在福州、泉州和厦门等沿海地区，占全省数据中心总量 65.9%。

(3) 整体上架率稳步提升

全省已建数据中心机架数上架率约为 44.8%，较 2018 年提高 10 个百分点。上架率最高的地市是莆田(78.7%)，其次是漳州(78.2%)、厦门(59.8%)，最低的是龙岩(24.6%)。

表 1 全省各地市数据中心机架数和上架率

地市	机架数		上架率	
	数量(个)	排名	百分比(%)	排名
福州	25513	1	47.4	6
厦门	4709	3	61.6	3
宁德	851	4	59.8	4
莆田	733	5	78.7	1
泉州	11200	2	28.9	7
漳州	66	9	78.2	2
龙岩	228	8	24.6	9
南平	395	6	54.4	5
三明	249	7	24.9	8
全省	43944	/	44.8	/

2.2 数据中心布局建设中的主要问题

(1) 集约化程度不高，分布零散

当前福建省数据中心总体上仍处于粗放建设阶段，集约化程度不高。一方面，全省已建成数据中心主要集中在福州、泉州和厦门等沿海地区，占总量 65.9%，其余数据中心零散分布在各个地市。另一方面，数据中心多为中小型数据中心，多数电信运营企业的数据中心是由原有老旧机房改造，而一些大型企业尤其是大型国有企业通常倾向于自建数据中心，较少利用互联网数据中心提供的服务，这也导致了中小型数据中心占比高。

(2) 整体上架率低于全国水平，各地区上架率不均衡

虽然福建省数据中心整体上架率近年来有了一定提升，但 44.8% 的上架率与全国平均水平(超过 55%)以及广东等发达省份(61.8%)相比仍然较低。此外，部分地区机架数和上架率增长尚不均衡，泉州、龙岩和三明等地区上架率远低于全省平均水平。

(3) 需求与供给脱节，资源利用不充分

福建由于区域位置受限，经营性数据中心对长三角、粤港澳大湾区等地区的互联网企业吸引力较弱，而多数企业在数据中心建设前也未根据客户需求，进行科学合理地规划，多数地市数据中心业务承接不充分，导致资源浪费，全省数据中心整体处于供大于求的状态。

2.3 总结

在当前全国一体化大数据中心体系布局的背景下，各省在数据中心布局建设时均遵循统筹规划、科学布局、集约建设等基本原则，力求实现数据中心的良性发展。福建省区域面积小、地市级城市数量少，虽然数字经济占比在全国保持领先，但是整体规模较小。福建省在数据中心布局建设中应考虑集中资源，将数据中心重点布局在 2-3 个地级市，对不同业务进行协同部署，支撑福建省数字经济高质量发展。因此，如何构建一套数据中心建设空间布局指标体系、数据中心应集约建设在哪些城市，成为一直困惑相关主管部门或建设企业的难题。

3. 数据中心建设空间布局指标体系建立

3.1 参考依据

- 《全国数据中心应用发展指引（2020）》
- 《关于数据中心建设布局的指导意见（工信部

联通〔2013〕13号）》

《数据中心设计规范（GB50174-2017）》

《新型数据中心发展三年行动计划（2021-2023年）》

《关于加快构建全国一体化大数据中心协同创新体系的指导意见》等

3.2 指标体系构建

数据中心建设空间布局指标体系应重点聚焦经济性、可靠性和安全性三个方面，分析产业环境、价格成本、气候条件、网络资源、交通配套、人才供给、能源配套、地质条件、容灾备份等条件影响。本文建立了一套包含 3 个一级指标、9 个二级指标、13 个三级指标的数据中心建设空间布局指标体系（图 1），测算各维度及综合评估指数。其中，一级指标包括数据中心经济性指标、可靠性指标、安全性指标，权重各占 45%、40%、15%。

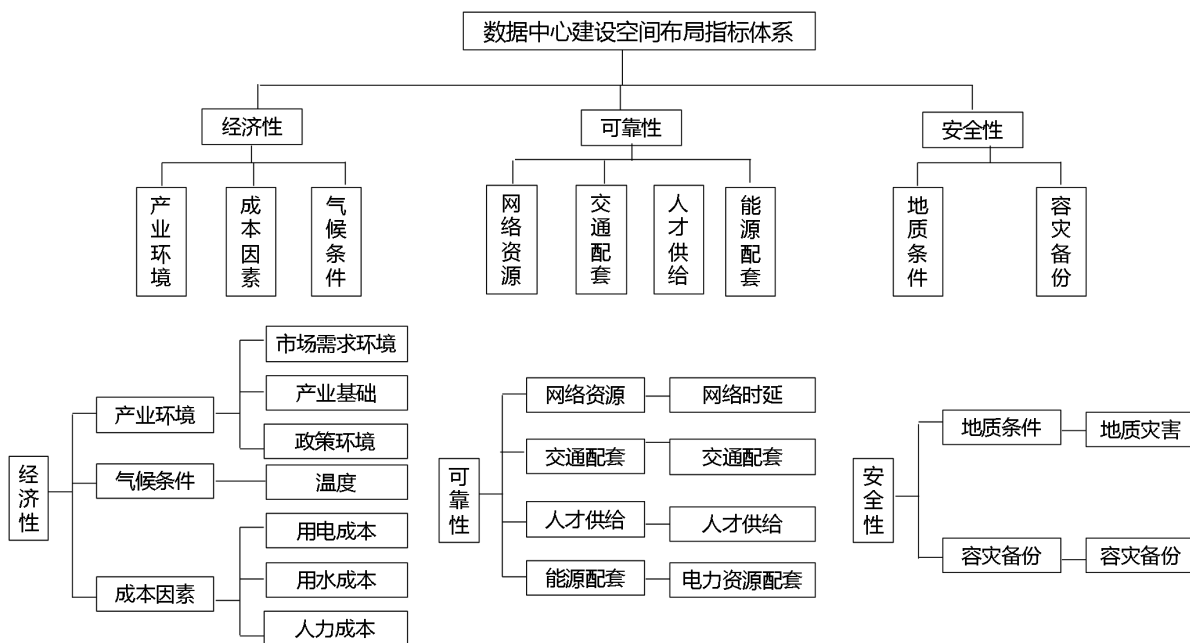


图 1 数据中心建设空间布局指标体系

表 2 数据中心建设空间布局指标

一级指标	二级指标	三级指标	三级权重	指标说明	统计口径
经济性（45）	产业环境	市场需求	8%	反映市场需求情况	数字经济发展指数
		产业发展	7%	反映支撑市场需求的经济 and 产业发展情况	地区生产总值指数

		政策环境	5%	反映地区政策是否有利于数据中心集聚并能提高收益	已经发布或实施的数据中心支持政策的数量
	气候条件	温度	10%	反映地区气候条件是否有利于通过节能提高经济性	平均温度
	成本因素	用电成本	10%	反映地区的用电比较优势	电价
		用水成本	3%	反映地区的经济支持力度	水价
		人力成本	2%	反映人力资源成本	年平均工资
可靠性 (40)	网络资源	网络时延	20%	反映地区网络资源配套可靠性	距离省级节点距离
	交通配套	交通配套	3%	反映地区的发展阶段和潜力	机场、高铁站的数量
	人才供给	人才供给	2%	反映获得人力资源的难度和趋势	不同档次大学的数量
	能源配套	电力资源配套	15%	反映地区对用电保障能力和力度	富余电量=发电量-全社会用电量
安全性 (15)	地质条件	地质灾害	8%	反映区域地质灾害易发程度	地质灾害重点防范区县数量
	容灾备份	容灾备份	7%	反映同城容灾备份的可能性	本地中型及以上数据中心在建及建成数据中心数量

(1) 经济性

经济性指标能够体现区域布局数据中心时的经济效益情况，包括当地的产业环境、气候条件、成本因素等。

产业环境重点突出市场需求、产业发展和政策环境三个影响因素。首先，数字经济时代，数据中心的**市场需求与数字经济发展之间密不可分的关系，选取数字经济发展指数作为评判指标，指数越高则得分越高。其次，产业发展重点体现了支撑市场需求的经济 and 产业发展情况，因此选取地区生产总值指数作为评判指标，指数越高得分越高。最后，政策环境反映地区政策是否有利于数据中心集聚并能提高收益，体现了当地政府对建设数据中心的**支持程度，可以用当地已经发布或实施的数据中心支持政策的数量作为评判指标，政策数量越多得分越高。

气候条件能够体现数据中心在运营阶段的优势，反映地区气候条件是否有利于通过节能提高经济性。选取当地平均气温作为评判指标，平均气温 = (日均最高气温 + 日均最低气温) / 2，平均气温越低得分越高。

成本因素包括用电成本、用水成本、人力成本，分别反映了地区的用电比较优势、经济支持力度以及人力资源成本，可选取当地电价、水价、年平均工资作为评判指标，当地电价、年平均工资越低则得分越高，由于多数地区对水价实行地区分类征收水价的政策，因此水价可按照地区分类进行评分。

(2) 可靠性

可靠性能够体现出数据中心在当地布局是否可靠，反映出当地在网络资源、交通配套及能源配套等硬件资源以及以人才供给为代表的软件资源给予数据中心建设的支持力度。

网络资源是数据中心的基础，数据中心拥有丰富的网络资源才能确保其能够提供可靠的、不间断的服务。网络时延反映网络资源配套可靠性，可以用当地距离省级节点距离作为评判指标，距离省级节点越近得分越高

其次是能源配套。能耗始终是数据中心发展过程中面临的严峻问题和挑战。2021 年中国数据中心耗电量达到 2166 亿千瓦时，占全社会用电量 2.6%。因此，当地对数据中心用电保障能力和力度是数据中心建设中需要考虑的重要因素，选取富余电量作

为评判指标。富余电量=当年发电量-全社会用电量，富余电量越高得分越高。

交通配套反映地区的发展阶段和潜力，可以用机场、高铁站的数量作为评判指标，当地高铁站和飞机场合计数量越多则得分越高。

人才供给反映获得人力资源的难度和趋势，可用当地不同档次大学的数量作为评判指标，当地各类大学合计数量越多则得分越高。

(3) 安全性

安全性对数据中心布局建设来说也是一个至关重要的影响因素，安全性指标包括地质条件和容灾备份。

一方面是地质条件，作为投资额亿甚至数十亿、承担国家经济发展重担的新型基础设施，数据中心在选址时应尽可能考虑地质灾害易发程度较低的区域，可用当地地质灾害重点防范区县数量作为评判指标，数量越少则得分越高。

另一方面是容灾备份，参考我国运营商数据中心建设经验，采取同城作为数据中心容灾备份建设，当地中大型数据中心数量越多得分越高。

4.福建省数据中心布局建设指标体系应用分析

本文以福建省数据中心建设布局作为实例分析，选取福建省九个地级市的三级指标进行测算，10分最高，0分最低。指标体系测算过程中，搜集整理的统计指标均为公开信息渠道发布的最新数据，力求做到指标测算的公平性和公正性，为福建省数据中心建设布局提供了实证依据和理论基石。下表为福州指标评分情况，通过加权计算，福州数据中心建设空间布局指标评分为 8.7 分。

表 3 数据中心建设空间布局指标详细评分情况（福州）

一级指标	二级指标	三级指标	三级权重	统计口径	单位	数量/分类	得分	加权得分
经济性 (45)	产业环境	市场需求环境	8%	数字经济发展指数(2020年)	/	88.49	10	0.8
		产业基础	7%	地区生产总值指数(2020年)	/	105.1	7	0.49
		政策环境	5%	已经发布或实施的数据中心支持政策的数量	项	4	10	0.5
	气候条件	温湿度	10%	年平均温度	度	22	9	0.9
	成本因素	用电成本	10%	电价	元/千瓦时	0.6254	4	0.4
		用水成本	3%	水价	元/吨	二类地区	5	0.15
		人力成本	2%	年平均工资	元/年	96478	3	0.06
可靠性 (40)	网络资源	网络时延	20%	距离省级节点(福州)距离	公里	0	10	2
	交通配套	交通配套	3%	机场、高铁站的数量	个	6	8	0.24
	人才供给	人才供给	2%	不同档次大学的数量	个	35	10	0.2
	能源配套	电力资源配套	15%	富余电量	亿千瓦时	415.6	10	1.5
安全性 (15)	地质条件	地质灾害	8%	地质灾害重点防范区县数量	个	0	10	0.8
	容灾备份	容灾备份	7%	中型及以上数据中心在建及建成数据中心数量	个	5	10	0.7
合计	/	/	100%	/	/	/	/	8.7

通过对全省九个地级市的三级指标进行评分汇总后,各地市得分情况见表4。其中,福州市得分最高,厦门市得分排名第二。结合经济发展、资源配套等实际情况,福建省可将福州、厦门作为全省数据中心的“两核”,统筹数据中心规划布局。

立足福州和厦门,主要依托数字福建(长乐)产业园、厦门软件园,打造高速互联、数据流通、优势互补的闽东北和闽西南数据中心集群,加快建设中国电信(长乐)东南信息园数据中心、中国移动(福州、厦门)数据中心等区域性数据中心,打造支撑国家数字经济创新发展试验区的新型一体化大数据中心,推动全省大型和超大型数据中心园区化、规模化发展。

围绕福州大数据产业发展定位,发挥数字福建云计算中心(政务云、商务云)优势,打造国家东南区域大数据中心。围绕厦门金砖国家新工业革命伙伴关系创新基地战略实施,打造面向先进制造业的行业数据中心,有效弥补国内外工业超算领域短板。引导福厦数据中心集群重点满足金融证券、灾害预警、网络游戏和远程医疗等高频低时延类业务需求,深度对接先进制造业、现代服务业、特色旅游和现代农业等上云需求。

表4 福建省数据中心建设空间布局指标总体评分情况

序号	地市	得分
1	福州	8.7
2	厦门	8.3
3	宁德	7.7
4	莆田	7.4
5	泉州	6.6
6	龙岩	6.1
7	三明	5.7

8	漳州	5.5
9	南平	4.9

5.结束语

本文以东数西算工程为背景,针对当前全国一体化大数据中心体系布局中存在的集约化程度低、区域布局建设与实际情况联系不紧密等问题,虽然提出通过构建数据中心建设指标体系,根据指标评分结果选取数据中心建设区域的优化思路,但在实际运用过程中还存在很多需要完善的地方,因此仍需要在实践中不断探索,持续优化,提高数据中心布局建设的科学性和有效性。

参考文献

- [1]广东省工业和信息化厅关于印发广东省5G基站和数据中心总体布局规划(2021-2025年)的通知【Z/OL】.【2020.7】. http://gdii.gd.gov.cn/gkmlpt/content/3/3026/post_3026281.html#906
- [2]中国信通院.数据中心白皮书(2022年)【R/OL】.【2022.4】. <http://www.caict.ac.cn/kxyj/qwfb/bps/202204/P020220422707354529853.pdf>
- [3]中国信通院.中国算力发展指数白皮书(2022年)【R/OL】.【2021.11】. <http://www.caict.ac.cn/kxyj/qwfb/bps/202211/P020221105727522653499.pdf>
- [4]艾瑞咨询.2020年中国数据中心行业研究报告【R/OL】.【2021.4.30】. https://report.iresearch.cn/report_pdf.aspx?id=3699.
- [5]张瀚.对新基建数据中心建设布局的有关思考【J】.改革与开放,2020(13):29-33.
- [6]高鹏,李玮,唐利莉,李俊晓.新基建风潮下的全国数据中心算力布局规划研究【J】.电信工程技术与标准化,2021(8):2-5.

基于算网协同的智慧社区居家养老应用方案研究

林启笔 谢丹卉 张浩宇 骆晓霞 林泽剑 游思萍 詹丰源

(中国移动通信集团设计院有限公司福建分公司)

摘要: 针对居家养老安全监护存在的风险隐患多、干预不及时、人工看护难等问题,研究院自主研发居家安全健康监测系统,结合人工智能、物联网等新兴技术识别老人室内日常行为和异常事件,实现对居家老人安全智能监护,及时发现危险事件和健康问题。在算网建立移动云应用新生态,通过系统上云,存储上云,应用上云,AI 赋能释放过亿家用 STB 终端存量端算力,打造“端+云+业务+内容”四位一体的业务解决方案,赋能居家养老业务场景,实现家庭云时代。

关键词: 智慧社区; 居家养老; 移动云; 端算力; 家庭云。

引言

我国老年人大多数选择居家和社区养老,形成“9073”的养老格局,以家庭为核心,以社区为依托,以专业服务为依靠。在居家养老模式下,安全健康是最基本且最重要的照护需求,跌倒、坠床、烫伤等事故处理不及时威胁居家老人生命安全。据统计,65岁及以上老年人群体日均收视时长达到276分钟,电视成为老年人居家娱乐休闲的主要方式。电视大屏的核心是机顶盒,目前电视端开展的业务类型相对匮乏,并未涉及智慧养老项目,因此在不改变家用机顶盒(STB)终端本身应用方式的前提下,通过系统上云,存储上云,应用上云,AI赋能,打造“端+云+业务+内容”四位一体的业务解决方案开展智慧养老业务具备可实施性。以此实现家庭云时代,丰富居家智慧养老新生态,缓解居家养老所带来的各类风险和威胁。

同时,居家养老目前存在基础设施不全,低意愿和监管困难的问题,养老服务缺乏有效的监管机制,导致出现部分老年人信息泄露、健康数据泄漏等问题的出现。引入区块链不可篡改的特性,与社区的康养小屋、独居老人家中的智能健康设备联动,解决传统养老的难题,实现养老生态数字化。

1. 智慧社区居家养老业务前景及应用场景分析

1.1 业务前景分析

未进一步推动智慧健康养老产业发展,工业和信息化部、民政部、国家卫生健康委共同制定了《智慧健康养老产业发展行动计划(2021-2025年)》,其中明确提出加强跨学科、跨领域合作,推动物联网、大数据、云计算、人工智能、区块链、超高清视频、虚拟现实等新一代信息技术在健康及养老领域的集成创新和融合应用,提升健康养老产品及服务的智慧化水平。

第七次全国人口普查数据显示从人口结构看,近10年间,中国已跨过了第一个快速人口老龄化期,我们很快还需应对一个更快速的人口老龄化期。2020年,大陆地区60岁及以上的老年人口总量为2.64亿人,已占到总人口的18.7%。自2000年步入老龄化社会以来的20年间,老年人口比例增长了8.4个百分点,中国的人口老龄化水平从最近几年短暂的相对缓速的演进状态扭转至增长的“快车道”,缓解养老压力迫在眉睫。

1.2 应用场景分析

场景1: 算网云边端协同,实现家用机顶盒(STB)及摄像头等终端计算卸载,开展居家养老业务

随着家庭业务应用越来越丰富,对终端产品的软件硬件性能要求也越来越高。现有家用机顶盒(STB)及摄像头算力有限,无法对采集到的视频进行智能提取和分析,如果新购置更高性能的摄像

头，价格也相对昂贵。为盘活大量现有家用机顶盒（STB）及摄像头等终端卸载的算力，发挥端侧算力对居家安全健康监测数据采集、计算、分析等应用的本地数据处理、近端低时延优势，开展居家安全健康监测等新业务，有必要在不改变应用本身展现方式的前提下，据业务的特点，通过多级 POP 点部署，将家庭终端的算力转移到到算网的最佳的位置，通过系统上云、存储上云、应用上云，腾出终端算力，通过云边端协同，打造“端+云+业务+内容”四位一体的业务解决方案，赋能移动云应用

新生态。

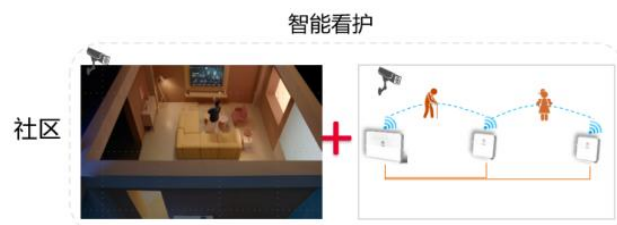


图 1 多热点网络覆盖，构建电子围栏的防护监测

应用	场景（范围：社区内）
老人看护	重点场景：老人跌倒监测，发现老人跌倒后，迅速上报该事件给子女，并截取事件前后关键视频发送到子女手机。 其他场景：活动范围提醒；异常行为监测；久坐或久卧提醒；半夜醉酒提醒。
儿童监护	儿童离家、回家通知，摄像头语音互动

场景 2：利用视频+AI 的数据信息，结合区块链技术，构建居家养老信息管理数据库

利用视频+AI 的数据信息，实现对于社区老人的健康档案整理；结合区块链不可篡改性、去中心化以及分布式记账等技术特点技术，对老年人身体数据进行安全加密，并全程记录养老服务流程，融合社区卫生院与养老机构，在现有的基础上完善数据统计，建立老年人健康信息管理数据库，为老人提供平台+硬件+机构上门服务和远程诊疗的居家养老服务，有效解决传统养老服务难以监管的问题。

2.算网智慧社区居家养老应用方案研究

本文以国内某运营商算力网络为背景进行智慧社区居家养老应用方案研究。

2.1 技术方案说明

以“端+边+中心”算网协同为主线，端侧的摄

像头对视频作初步的识别、分析和处理，但部分摄像头本身的算力不足导致无法支撑视频的处理或者不支持复杂的视频处理，这种情况下部署在边缘节点就可以实现对算力做补充，并实现云资源统一管理、网络 VPC 互通、资源弹性调度等功能。此外，在合营云上部署健康养老统一门户，实现统一账号和服务入口、统一资源视图和运维功能。“端+边+中心”算力协同极大地缓解了端侧算力不足、或者算力分配不均的问题，并结合区块链技术将老年人身体数据（如身高体重、血压、心率、血液浓度等信息）上链存证，保证数据真实不可被篡改；同时，并将老人养老服务流程的各关键环节（如基本身体数据、体验报告、医院数据、社区相关数据等）记录到区块链上，从而实现养老服务流程追溯管理，有效解决传统养老服务难以监管的问题，让居家养老更完善，实现居家养老的高效运行。

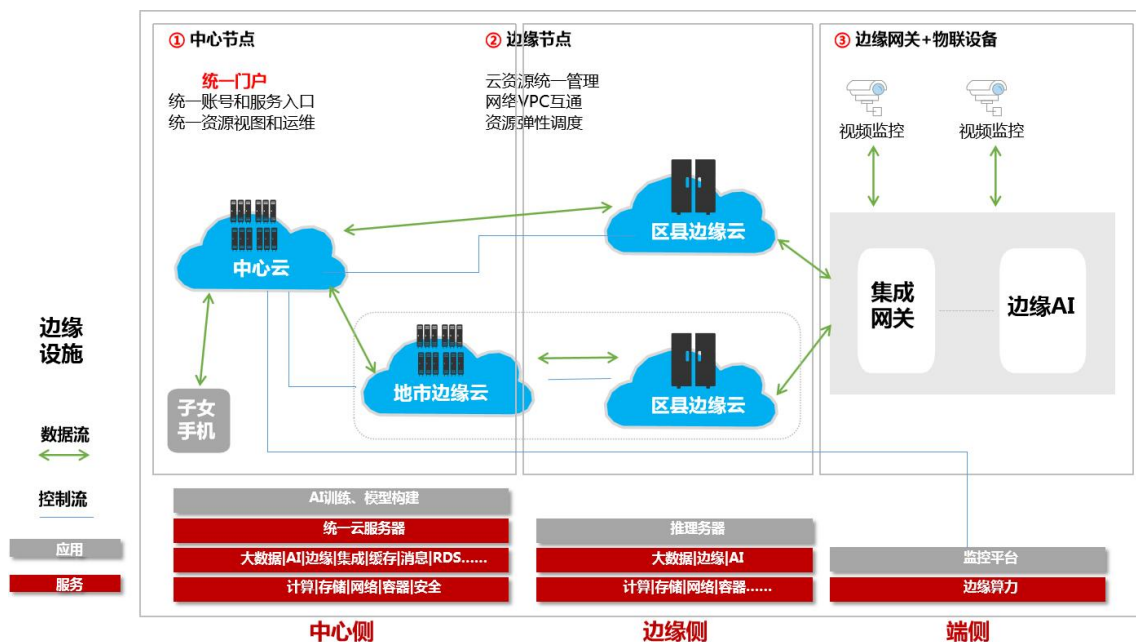


图2 技术方案设计图

端侧：通过边缘网关+物联网设备+摄像头等，实现端侧数据收集和初步处理。同时可部署监控平台、集成网关等边缘设备。

边缘侧：根据业务分布情况，在区县或地市区域，部署计算、存储、网络等资源，形成边缘算力，进行云资源统一管理、网络 VPC 互通、资源弹性调度等功能。

中心侧：中心云上部署统一门户，接管全省所有边缘云业务，形成统一账号和服务入口、统一资源视图和运维功能，面向全省用户提供健康养老服务。

2.2 功能架构设计

功能架构主要分为硬件层、存储层、服务器及应用层。

硬件层：多类型的监测设备采集视频，汇聚于手机顶盒作初步的分析。

存储层：依托区块链产品，构建去中心化、分布式、可信数据存储能力，构建老年人健康信息数据库，为上层应用提供支撑。

服务层：端算力不足由云端云服务平台作算力补充和算力资源的灵活调度，通过容器化的部署方式可以实现按需创建和资源的弹性伸缩。再由先进的算法进行分析并传递结果。

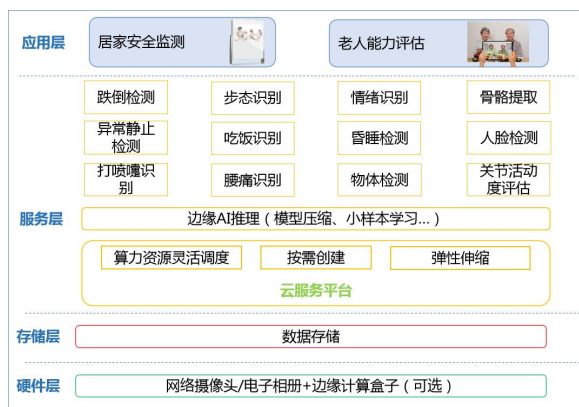


图3 功能架构图

应用层：各类应用根据算法分析得出的结果给出相应的反馈，并用于指导用户行为。

2.3 组网方案设计

云平台按照 3 级架构设置，功能分别如下：

- 省级中心云：省级集中设置在福州，负责视频 AI 业务的数据训练、模型下发边缘云等，初期也可兼顾地市历史视频的回看业务的存储。

- 地市边缘云：以地市为单位集中设置，结合业务量规模按需设置，负责地市历史视频的回看业务。

- 区县边缘云：以区县为单位分布式设置，

实现实时推理,面向边缘应用实现超强计算性能、环境适应性强、易于维护和支持云边协同能力。

匹配上述平台架构,基于 PON+CMNET+省云专网实现云、边、端三级接入和协同。

(1) 摄像头终端通过 PON 进行接入;

(2) 边缘云下沉至区县 BRAS 机房,摄像头基于传统 VLAN 或 QinQ 方案接入,经 BRAS 终结后接入区县边缘云;

(3) 地市边缘云部署在地市云网 PE 机房,摄像头数据经 OLT 汇聚上行至 BRAS, BRAS 部署 SA (业务感知) 插板并使能 SRv6 Policy, 精准识别出视频 AI 业务后,经 SRv6 “一跳入云”接入地市边缘云;

(4) 经 SRv6 “智能路径规划”按需入地市云网 PE 后接入省级中心云;

(5) 通过智能云图算法调度接入邻近地市负载更优的地市边缘云,实现地市云池间业务的负载均衡;

(6) 地市边缘云与省级中心云经省云专网疏导互通,端到端使能 SRv6;

(6) 省级中心云间建立 SRv6 预连接,实现数据备份或同步。

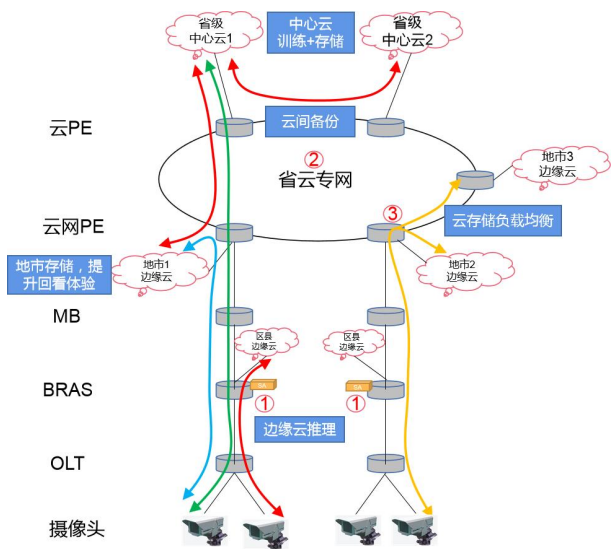


图 4 组网方案图

上述方案中使用到的关键网络和技术如下:

(1) 算力泛在协同

云边 VPC 互通,可使用云端服务一致体验,

并支持本地安全部署;

轻量起步,可弹性调度云端资源,应对业务洪峰;

公有云远程运维,降低客户运维成本及复杂度,可跟随公有云持续升级;

本地化部署,公网短暂中断,不影响业务连续性。

(2) 网络按需调度能力

快速开通,体验保障,视频 AI 业务经专用隧道智能引导至省云专网承载,业务快速部署,安全隔离有保障;

预连省内多云(省级+地市+边缘),多云间使能智能规划调度,业务间硬隔离,租户级业务 SLA 可视,网络资源按需弹性调整;

智能云图算法实现摄像头和省级中心云、地市边缘云之间的灵活调度,实现跨地市云池存储效率负载均衡,为企业上云提供性价比最优的路径,为运营商提供云资源智能均衡调度。

(3) 区块链技术

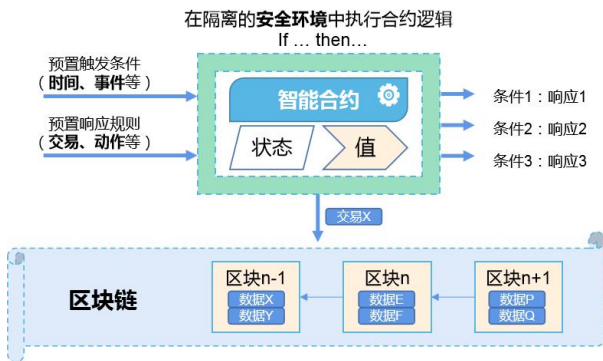


图 5 区块链技术示意图

基于区块链技术打造一站式、高可用、高性能的区块链管理、开发、服务、运维等核心能力,支持存证、溯源、数字身份、可信对账等通用应用。通过业务网关,实现业务快速上链;通过区块链管理平台,实现对链、联盟的全方位管理;通过独立的运维监控工具,实现对系统和业务的全方位监控告警。

3.社会效益分析

积极履行企业在健康养老领域的社会责任与义务,通过响应智慧健康养老产业发展行动计划

(2021-2025 年)等文件要求,协助政府部门着力破解养老服务发展障碍,打通“堵点”,消除“痛点”,增加优质养老服务供给,扎实推进养老服务高质量发展。

(1) 协助民政开展居家养老床位试点工作。

积极响应省民政厅居家养老床位试点工作,通过与民政系统对接,提升了运营商在养老解决方案上的社会形象。

(2) 协助残联打造助残视频客服

协助残联,基于运营商服务热线打造“自助服务+在线客服+视频客服”的助残视频客服能力。

(3) 协助养老机构进行公益照护培训

协助民政、金太阳及老年大学等养老机构,开设互联网电视专区。通过合作录制教学内容、引入公益内容的形式,让老年人在家就可享受优质的公益课程。

(4) 可与传统养老机构开展线下合作

与传统社区养老服务中心开展线下合作,提供健康养老信息化服务。

4. 结语

在人口老龄化不断加快的今天,基于智慧社区的居家养老服务模式,是在当前建设智慧城市的大背景下,探索出的一种以老年人需求为导向的新型居家养老服务模式。此种模式是将智慧社区建设与社区居家养老服务相结合,将老人需求与市场供给相结合,构建一种便利高效的社区居家养老服务体系,是未来社区建设和居家养老服务模式的必然发展方向。通过 AI 赋能、引入区块链不可篡改的特性,与社区的康养小屋、独居老人家中的智能健康设备联动,解决传统养老的难题,实现养老生态数字化,打造数字福建“区块链+智慧社区”新标杆。

参考文献

[1] 中国工业和信息化部,民政部,国家卫生健康委. 智慧健康养老产业发展行动计划(2021-2025 年)[R]. 北京,工业和信息化部,民政部,国家卫生健康委,2021.

面向 5G 智能网络的实时车辆检测方法

曾泽鹏

(福建省邮电规划设计院有限公司)

摘要: 当前,国内外人均车辆拥有率比以往更高,而车辆的复杂性成为管理交通干道和停车场等区域的一大难题。此外,异常驾驶行为可能导致交通事故,对公共区域和行人安全构成待解决的问题。因此,有必要对公共区域上行驶的车辆进行实时检测和跟踪,监测每辆车的实际驾驶状态,并记录其基本通行信息。这对于公共交通安全和公共区域管理效率具有重要意义。本文主要基于具备多路视频功能的 Jetson AGX 检测系统,利用其 Deep Stream 分级模型对不同照度下场景中的目标进行检测、识别和反馈。该系统能够无差别地识别行人、传统车辆或新能源汽车,并在高效率、低时延的 5G 网络下持续处理和反馈检测目标。本文所提出的算法系统具有实时性强、高精度的特点,可以应用于交通干道、停车场等公共场所进行车牌识别、位置分析等检测任务,其真实检测率高达 98.9%。

关键词: 智能算力;边缘计算;5G 网络;车辆检测;实时视频流

引言

随着社会的不断发展和科技的进步,车辆作为我们日常生活中不可或缺的交通工具,其数量和种类也日益增多。然而,随之而来的是车辆事故和交通拥堵等问题的不断增加。为了提高交通安全性和效率,车辆检测技术成为了当前研究的热点之一^[1]。车辆检测是计算机视觉领域的重要研究方向,旨在通过利用图像或视频数据来自动检测和识别道路上的车辆。这项技术具有广泛的应用前景,如交通管理、智能驾驶、安全监控等。它可以帮助交通部门实时监测交通流量、改善交通信号控制系统,为驾驶员提供驾驶辅助,甚至在自动驾驶技术中发挥关键作用^[2]。

然而,由于车辆检测任务的复杂性和挑战性,例如车辆外观的多样性、光照条件的变化、遮挡和视角变化等,车辆检测仍然是一个具有挑战性的问题。在车辆的检测和识别技术领域,不断涌

现出创新的研究。例如,有学者提出基于 Canny 算子来进行车牌的边缘检测,然后通过 Hough 变换找到车牌的水平 and 垂直边线^[3]。这种算法速度很快,但在模糊图像的情况下,检测结果可能不够理想。另一种基于颜色的车牌检测方法利用 HSV 颜色空间模型和直方图来得出候选车牌矩形框^[4]。然而,这种方法在车牌颜色与车身颜色相近时效果较差。有研究团队还提到了一种基于 Haar 特征的车牌检测算法,它是一种基于纹理的车牌检测方法的应用^[5]。这种算法对车牌的亮度、颜色和大小的变换具有鲁棒性,但相比前两种方法需要更复杂的计算过程。近年来,深度学习技术的发展和超大数据集的可用性使得车辆检测领域取得了显著进展。相比传统算法,深度学习方法具有自动学习特征的能力,避免了手工特征提取的错误。这些方法利用深度神经网络从图像中提取更具代表性的特征,提高了车辆检测的准确性和鲁棒性。

作者简介:

曾泽鹏, 本科毕业于福建理工大学, 现任职于福建省邮电规划设计院有限公司, 主要从事集客类项目, 负责信息化、智能化项目的全过程支撑及项目管理。多次发表 5G 在行业中应用期刊, 并完成公司 5G+行业应用白皮书。

研究关注深度神经网络在车辆检测中的应用,特别是基于 Faster R-CNN 和 YOLO 网络的方法^[8]。在 Faster R-CNN 方面,有学者通过修改模型参数,如训练规模、测试规模和建议值数量,来评估其在 KITTI 数据集上进行车辆检测时的性能效果^[9]。另外, Espinosa 等人比较了 AlexNet 和 Faster R-CNN 在城市区域视频中进行移动车辆检测的性能,并得出了 Faster R-CNN 具有更好 F1 分数的结论^[10]。Mo 等人^[11]设计了基于 Faster R-CNN 的深度神经网络,用于航空图像中的车辆检测。他们采用了 ResNet101^[12]作为特征提取器,并通过双线性插值放大特征图以减少信息损失。此外,他们还设计了一个联合损失函数,以便同时检测水平和定向的车辆。在基于 YOLO 网络的方法中, Xu 等人^[13]使用经过修改的 YOLO3 网络进行航空图像中的车辆检测。Ghoreyshi 等人^[14]设计了两个不同的车辆检测网络,采用相同架构用于检测伊朗网站图像中的车辆。Zhou 等人^[15]采用改进版的 YOLO 版本 3 网络进行卫星图像中的车辆检测。Doan 等人^[16]使用 YOLO 版本 4 网络^[17]和 DeepSORT 网络^[18]进行车辆检测和计数。YOLO 版本 4 网络用于预测边界框的坐标、检测到的对象的类别和对象的置信度分数,DeepSORT 网络用于跟踪检测到的物体。

综上所述,现有的研究在车辆检测领域应用了深度学习技术,克服了传统模型在照明条件差、复杂天气等场景下的检测困难。然而,在真实的 5G 网络背景下,对实时视频流中的车辆进行检测的问题仍然具有挑战性。为了解决这个问题,本文提出了一种基于 Jetson AGX 的车辆检测识别模型,旨在实现对实时视频流中车辆的实时检测,并实现用户端与检测端信息的实时交互。本文的主要贡献如下:

(1) 该模型能够对传统车辆和新能源车辆进行准确的检测和识别,不存在偏差。此外,它还能进一步检测车辆的属性,例如车身颜色和汽车品牌,并最终输出每个检测到的车辆的属性类别标签和边界框坐标。

(2) 在不同照度检测场景下,能够保持较高的检测准确度,同时能够对行人和车辆进行差别标注,检测来往车辆的异常驾驶情况。

(3) 多路视频采集与检测工作多路并行,实

现 5G 云端与客户端的低时延信息交互。

1. 实验原理及方案

1.1 实验原理及平台介绍

实验项目使用 jetson AGX + NVIDIA tlt 包,并且用 jetson AGX 设备的视频分析套件实现车牌识别与车辆检测功能,项目的执行原理主要分为三个部分:

(1) 在 jetson AGX 上使用 Deep Stream 识别中文车牌。

(2) 用 NVIDIA TLT 训练 LPD (License Plate Detection) 模型,负责获取车牌位置。

(3) 用 NVIDIA TLT 训练 LPR (License Plate Recognition) 模型,负责识别车牌内文字。

Deep Stream 应用程序:该应用程序旨在将深度神经网络和其他复杂处理任务应用于流处理管道,以实现视频和其他传感器数据的实时分析。它采用了一组模块化插件,这些插件相互连接形成处理管道。每个插件代表一个功能块,例如使用 TensorRT 进行推理或进行多流解码。硬件加速的插件与底层硬件进行交互,以提供最佳性能。例如,解码插件与 NVDEC 进行交互,推理插件与 GPU 或 DLA 进行交互。

Deep Stream SDK:通用流分析架构定义了可扩展的视频处理管道,用于执行推理、对象跟踪和报告等任务。在 DeepStream 应用程序中,每个视频帧都会经过管道的分析,并通过插件提取信息,将其存储为级联元数据记录的一部分,以保持记录与源帧的关联。管道末端的完整元数据集包含了深度学习模型和其他分析插件从框架中提取的全部信息。这些完整的元数据可以用于在 DeepStream 应用程序中进行显示,也可以作为消息的一部分传输给外部系统,以进行进一步的分析或长期归档。每个插件可以根据需要在管道中实例化多次,以实现灵活的功能扩展。DeepStream 还支持“分级模型”的功能,即能够组合检测和分类的任务。这一特性使得车牌识别等应用可以被拆分成三个主要部分,如图 1 所示。

一级 PGIE:这是 Deep Stream 的主模型,以车为检测 (detection) 目标。

二级 SGIE:这里以“车牌位置检测 (LPD)”

的模型，在 PGIE 所找到的“Car”范围内，定位出“车牌”的标框位置。

三级 SGIE：从二级 SGIE 得到的车牌图形中，执行“车牌内容识别（LPR）”任务，并将识别的文字回传。

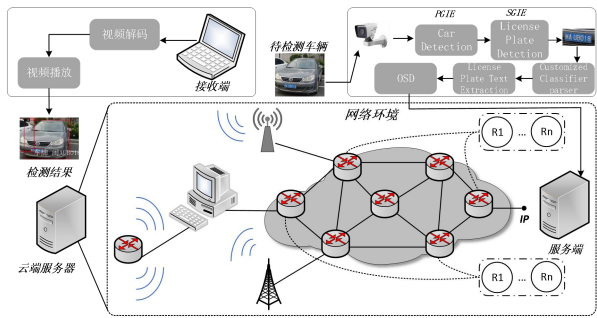


图 1 基于 jetson AGX 系统车辆检测识别软件流程图

LPR: 该车牌识别网络旨在从裁剪的 RGB 车牌图像中识别车牌中的字符。它提供了两个预训练的 LPRNet 模型，一个在美国车牌数据集上进行训练，另一个在中国车牌数据集上进行训练。这个模型使用了 ResNet 主干的序列分类模型，可以将图像作为输入并生成序列输出。主要应用是从裁剪的 RGB 车牌图像中识别车牌。训练优化算法主要使用连接主义时间分类（CTC）损失来最小化真实字符序列和预测字符序列之间的差异。然后，通过最佳路径解码方法（贪婪解码）

从模型的序列输出中提取车牌。模型的输入是尺寸为 3*48*96（CHW）的 RGB 图像，输出是通过 Deep Stream 后处理插件获得的包含车牌字符的序列。

LPD: 该模型基于 TAO Toolkit 中的 Detectnet_v2 网络。训练算法旨在通过最小化对象的定位和置信度损失来优化网络。训练过程主要分为两个阶段。在第一阶段，使用正则化来训练网络以促进剪枝。在第一阶段结束后，移除内核规范低于修剪阈值的通道。在第二阶段，对修剪后的网络进行重新训练（不包括正则化）。该模型使用数据集调整图像/标签的大小，并通过下载训练规范配置训练参数。规范文件中的参数为 LPD 指定 NGC 预训练模型。预训练模型经过在特定数据集上的训练和微调。

实验方案

根据图 2 所示，在 Jetson Nano 视频检测系统的基础上，本文进一步应用具备多路功能的 Jetson AGX 检测系统。首先，该系统通过实时捕获摄像头获取车辆信息，并对所获取的车辆视频信息进行处理。然后将计算任务从云端下沉到位于网络边缘的终端设备和边缘服务器上。通过客户端，可以对检测的目标、实时位置、车辆颜色和车牌识别分析进行异常跟踪。为了方便实验，本文采用 4 路摄像头进行实时视频采集工作。该系统提供实时性强、高精度的目标检测技术支持，适用于交通干道、停车场等场景，用于车牌识别、位置分析等。

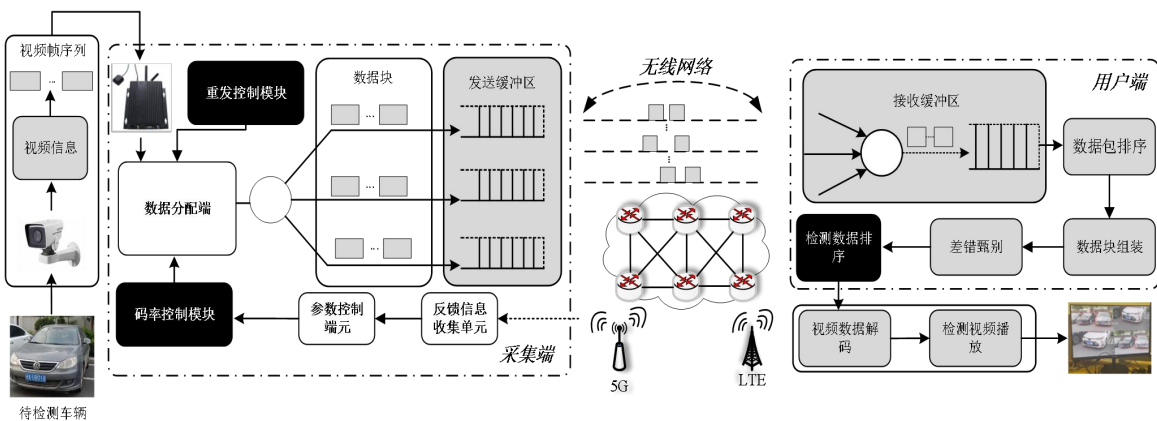


图 2 基于 jetson AGX 系统车辆检测识别平台图

该模型基于 jetson nano 系统扩展到多路采集、并行缓存、多路实时播放的 jetson AGX 系统，该系统与其他系统的对比结果如下表 1 所示。它同时能够保持视频采集、识别处理、结果反馈能够持续保持高速率、低延时的工作状态。

表 1 系统对比结果

设备	视频数量	Batch size	NvDCF	IOU
AGX Xavier	5	5	115	153
Xavier NX/FP16	3	3	63	81
Nano 4GB	1	1	9	13
Nano 2GB	1	1	5	7

2. 实验

2.1 实验相关设置

实验选择永嘉天地停车场作为测试地点。采用多路摄像头同时捕获车辆的视频信息。经过计算处理，生成计算任务，并通过云端下沉至网络边缘的终端设备和边缘服务器。设计好的车辆检测系统将对从云端下沉的计算数据进行处理。将待检测序列统一打包，并传送至检测系统的输入。经过训练的车牌检测 (LPD) 和车牌识别 (LPR) 模块各自负责其职能。DeepStream 主模型明确后续模块的工作重点，它能够检测出目标车辆的位置，并用红色方框标记出检测区域，以清晰展示。同时，它还能准确定位车牌的位置，并将其传递给 LPR 作为输入进行识别任务。LPR 根据已标识的车牌位置执行识别，并将识别结果返回。检测结果将生成 TXT 文件，以供客户直接查看系统的检测准确度。同时，通过 5G 网络的高速率和低延迟特性，客户终端的显示器可以播放检测画面和检测结果。

测试环境搭建：具体测试环境搭建由上图 2 车辆检测的硬件框图所示。

测试设备：摄像头、jetson AGX、显示器、无线路由。

2.2 实验结果分析

根据图 3 所示，该模型具备车牌识别、人物识别和计数功能。车辆检测模型 (PGIE) 可以从图像中检测一个或多个物体，并在每个物体周围返回一个红色方框和类别标签。该模型检测的两类对象是车辆和行人，输出包括每个检测到的对象的类别标签和边界框坐标。每一帧图像都可以识别多个对象。当检测到的对象属于车辆或人类，并且大小超过相应类别的最小边界框限制时，将使用相应的类别标签对这些对象进行标记，如图 4 和图 5 所示。例如，汽车将被标记为 "car"，行人将被标记为 "person"。对于部分被遮挡的对象，如果可见部分占总体的约 60%，仍会显示边界框；但如果可见性低于 60%，则不会添加边界框。



图 3 车辆检测识别



图 4 电瓶车与行人的检测



图 5 道路行人的检测

车牌检测模型 (LPD) 基于车辆检测模型 (PGIE), 在 "Car" 类别范围内定位车牌的位置。它能够从汽车图像中检测一个或多个车牌对象, 并返回每个对象的边界框和标签。输出包括图像中每个检测到的车牌的类别标签和边界框坐标。然而, 当汽车被遮挡或截断过多时, 该模型可能无法检测到车牌。

车牌识别模型 (LPR) 用于执行车牌内容识别任务, 并将识别的文本返回。该模型专门用于识别车牌中的字符, 并通过 DeepStream 后处理插件获得最终的车牌结果。然而, 模型可能无法很好地处理被截断或字符形状不完整的车牌。关键的性能指标是车牌识别的准确性。中国 LPRNet 在中国城市停车数据集 (CCPD) 的验证分割上进行评估, 该数据集包含 99996 张图像, 识别准确性高达 99.67%。因此, 该模型在车牌字符识别的准确性方面有实验证据。根据实验的检测结果, 未发现错误的标识符。识别的信息将输出到如图 3 所示的 TXT 文本中, 并与实时视频流一起发送到云端并传送至接收端。接收端的显示结果如图 6 所示。



图 6 接收端显示画面

3. 总结

本文所设计的 Jetson AGX 检测系统实现了多路实时摄像头的并行采集工作, 并通过 5G 云

端与用户端低时延信息反馈。该系统能够在不同场景下对不同目标进行检测, 并实时跟踪目标的位置、车辆颜色和车牌识别分析, 以进行异常跟踪。在保证广泛适用于多元场景的同时, 该系统始终保持较高的检测精准度, 能够提高交通、物业管理等领域的效率, 并实现管理与用户的双重受益。目前, 该系统针对车辆检测场景的研究受限于测试场景。未来, 我们将考虑在港口、闹市等复杂区域进行进一步的研究, 以拓展系统的应用范围。

参考文献:

- [1] 林宇. (2012). 车牌识别在智能交通中的应用与发展趋势. *中国安防*, (11), 63-65.
- [2] 段红燕, 邵豪, 张淑珍, 张晓宇, & 王小宏. (2016). 一种基于 Canny 算子的图像边缘检测改进算法. *上海交通大学学报*, 50(12), 1861-1865.
- [3] 王能, 赵红立, 琚生根, & 杜洪. (2014). 一种改进的 canny 边缘检测自适应算法. *四川大学学报: 自然科学版*, 51(3), 4.
- [4] 谭司庭, & 胡志坤. (2011). 基于 hsv 色空间的 车牌定位综合方法. *计算机与应用化学*, 28(7), 4.
- [5] Deb, K., & Jo, K. H. . (2008). HSI color based vehicle license plate detection. *International Conference on Control, Automation & Systems* (pp.687-691). IEEE.
- [6] 朱晓红, 张永来, & 刘冬林. (2009). 非正交 haar 小波变换的车牌定位技术研究. *计算机工程与应用*.
- [7] 颜子凤. (2022). 基于改进 faster r-cnn 及 yolo 的大角度车牌识别算法. *工业控制计算机*, 35(2), 4.

基于互联网 GIS 的无线网络优化可视化分析

董帝焯 黄炳塔 杜丕加 吴晶晶 何晋毅
(中国联合网络通信有限公司泉州市分公司)

摘要: 为了更好地推进网络优化工作向互联网及智能化转型,介绍了利用互联网爬虫技术实现网络优化地理可视化的思路,研究了互联网爬虫技术获取互联网地理空间数据,利用自研算法进行建筑物和网络小区的关联,实现了高铁高速等线路、建筑物的覆盖可视化,网络性能感知指标的可视化,以及用户轨迹可视化的应用。

关键词: 互联网, GIS, 爬虫, 网络优化, 大数据, 可视化

引言

近几年,互联网、大数据、人工智能的技术正在越来越多地应用到网络优化中,成为提升网络优化效率,降低网络运营成本的新手段,网络优化思路正在从之前依靠网优人员的技能经验向大数据关联分析优化转变,网优手段从以人工分析手段向智能化、自动化的优化手段转变。网络优化是面向移动网络开展网络优化工作,提升网络性能和用户感知,操作的对象是移动网络中的小区。在网络优化效果评估中,以小区为单位,对于结果的展现缺乏更合理、直观的可视化展现,在此背景下,利用互联网爬虫技术,结合网络的数据和性能,能更好的呈现出无线环境覆盖情况及网络性能指标。

1. 分析思路

在现在大数据时代,互联网上充斥着各种大量的数据,这些数据中隐藏着巨大的价值。特别是在地理化可视化方面,互联网上的地理空间数据不仅具有普通数据的属性值,还包含有时间和空间信息,具有空间、时间、属性三维特征,蕴藏着地表地物和经济社会活动的时间变化和空间分布规律,从互联网中挖掘出相关的地理信息成为了地理信息来源的一个新渠道。通过互联网数据的挖掘,可以从互联网海量数据中发现网页的读取模式、互联网结构特点和互联网内容描述所存在的规律性和动态性,可以有效地提取有用

的信息。基于互联网技术的高德矢量边界轮廓数据获取,得到高价值重点场景轮廓数据,结合 MR 大数据进行分析,利用 SEQ 进行精细化优化,提高用户感知。利用爬取的数据与 MR、网络性能指标等关联分析,评估网络性能,识别高价值区域。

在这种互联网爬虫技术手段下,利用数据挖掘技术,通过自主开发工具,实现从数据到价值的转化,改变原有的模式。自主开发图层工具及利用 GIS 应用,三维建模,精确至楼层基本的可视化展示,360 度展示无线环境情况及数据分析可视化展示。一方面可以从基于网络、小区开展的网络优化转变成基于建筑物、立体化的网络优化,另一方面从基于网络指标的网络优化转变成基于用户感知、体验效果的网络优化,在不断探索与实践中,以打造低成本、高质量的移动网络为目标,推动网优智能化发展。

2. 实现方法的研究

2.1 互联网 POI 数据爬取

利用 python 编写数据挖掘脚本,从互联网获取包含场景轮廓数据、腾讯位置大数据等地理信息,通过高德地图,百度地图等获取 POI 信息。POI 信息有两部分,一部分 POI 的 info 信息部分,一部分是 POI 的场景边框数据部分,两种数据均需要用于后续的场景数据处理。

第一步,分析地图 URL,利用 python 构建

下载规则脚本，进行多线程下载地理数据，达到

速度快，信息全，效率高。

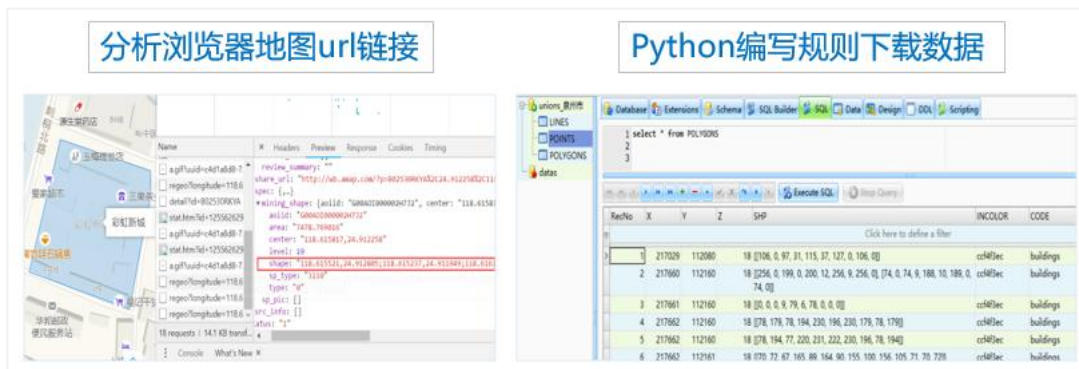


图 1：互联网 GIS 数据获取

第二步，解析包含楼层高度的建筑物轮廓加密数据，为住宅区的覆盖评估提供数据基础，数据中高德道路和建筑物数据为加密信息，需要进行解密解析。通过墨卡托坐标与瓦片编号的计算解析位置信息，多个步骤流程最终得到包含楼层高度的点线面数据。

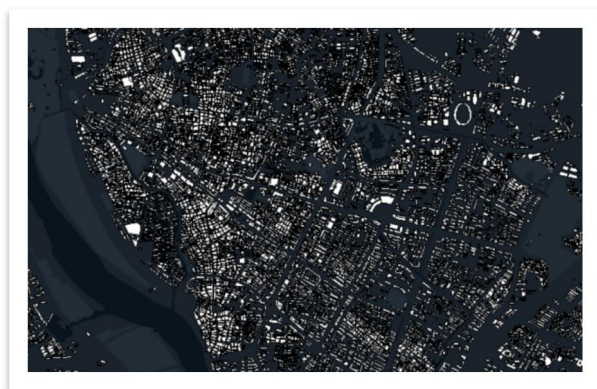


图 3：互联网 GIS 数据场景信息图层

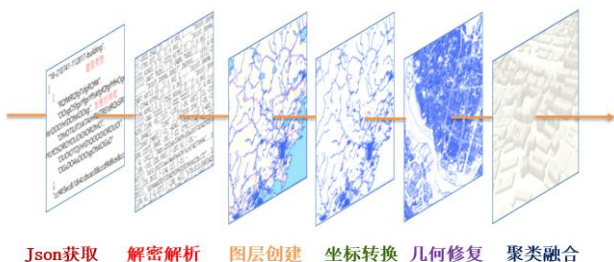


图 2：互联网 GIS 数据解析

第三步，通过自主开发图层软件创建图层实现数据地理化。具有自主知识产权的图层工具，支持 3 种图层格式（Mapinfo_Tab, ESRI_Shp, GoogleEarth_Kml），自由定义生成规则条件（半径，高度，扇区波瓣角，填充颜色、线条颜色，标注等），国内 3 种坐标转换(大地坐标、火星坐标、百度坐标)，利用 KDTree 算法，计算最短距离，区别以往的遍历循环，速度比原来快百倍，并支持 K 近邻设置。

第四步，获取最终包含场景信息的数据图层。获取的数据全，速度快，包含高德地图上所有点、线、面数据，所见即所得。

2.2 POI 与网络小区数据关联

现网 LTE 网络主要覆盖采用 1800 MHz 频段，该频段适用于 COST231 Hata 模型，基本公式如下：

$$\begin{aligned}
 \text{Path_Loss} = & 46.3 + 33.9 \times \log(f) - 13.82 \\
 & \times \log(h_b) - \alpha(h_m) \\
 & + [44.9 - 6.55 \times \log(h_b)] \times \log(d) \\
 & + C_m
 \end{aligned}$$

公式中 f 表示工作频率， h_b 表示基站天线的有效高度， h_m 表示手机终端天线的有效高度， $\alpha(h_m)$ 表示终端天线高度修正因子， d 表示基站与手机终端的距离。 C_m 表示城市修正因子，中等城市及郊区取值为 0dB，大城市取值为 3dB。

现网中 RS 参考信号的功率设置一般为 15.2dBm，通过链路预算的公式，为了方便计算，在暂时不考虑天线增益、馈线损耗、穿透损耗、阴影效应等因素的情况下，可以计算出 UE 的接收电平 $\text{RSRP} = \text{RS}_{\text{power}} - \text{Path_Loss}$ 。

用 KD-Tree 方法或者 GIS 软件功能计算与建筑物最近距离的 3 个小区，通过链路预算模型，可以计算互联网爬取的 POI 建筑物点的 UE 接收电平 RSRP，如果 3 个小区 RSRP 相差 3dB 以上，则取 RSRP 最大值为主覆盖小区，否则 3 个小区均作为主覆盖小区。计算出每个互联网 POI 建筑物点的主覆盖小区，就能得到 POI 与网络小区的关联关系。

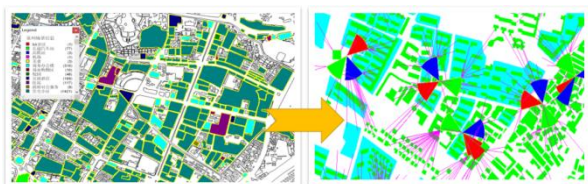


图 4：建筑物与小区关联分析

通过链路预算模型计算不同 UE 高度的主覆盖小区的 UE 接收电平 RSRP，高层超出波瓣角度的覆盖，加入波瓣角度超出 3dB 范围后信号衰减计算评估。通过链路覆盖模型，计算不同楼层理论接收电平，利用 6 个楼层模型的不同覆盖电平进行渲染，完成 3D 覆盖模拟的建模。

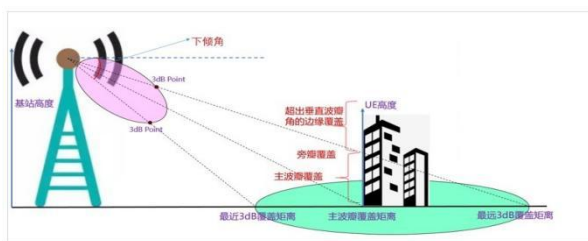


图 5:3D 覆盖模拟

3.网络优化可视化应用

3.1 无线环境及性能指标可视化

基于爬取的建筑轮廓矢量图层，自主开发的图层工具和 GIS 软件应用，通过 K 近邻算法及链路预算模型，建立建筑物 ID 与主覆盖小区，进行空间分析及可视化呈现，直观展现无线环境及网络情况。

3.1.1 高速高铁线路可视化应用

高速高铁爬取的路线以缩放级别 ZOOM=18 的地图瓦片 256px × 256px 大小分段，1 个瓦片大小编号 1 个 id；高速高铁线路有专属的覆盖小区，

因此只需计算最短距离的小区即是其主覆盖小区，引入主覆盖小区指标或叠加 MR 数据进行覆盖和网络指标可视化。

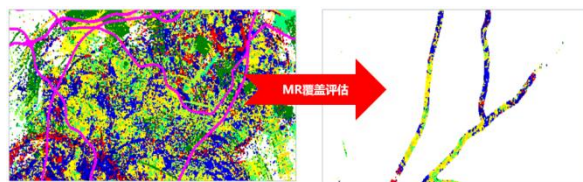


图 6：MR 覆盖评估

3.1.2 城区道路可视化应用

城区道路的覆盖小区关联可与高速高铁关联方法类似，利用 KDTree 算法计算 K 近邻覆盖小区，利用链路预算模型得出道路主覆盖小区，形成 1 小段路 1 个 id 编号关联小区覆盖，加入网络指标数据进行可视化渲染，指标统计的数字宜使用比例的方式考察评估。

3.1.3 基于建筑物级别的可视化应用

基于城区基站较为密集的特性，直接计算建筑物与小区最短距离即为覆盖小区，可通过 GIS 软件功能计算对象距离，如果通过 KDTree 算法计算因同站的不同小区经纬度都相同，需要将经纬度往方向角方向偏移 100 米左右进行计算，这样能较为准确的区分同站不同小区。

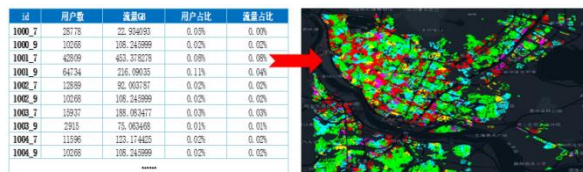


图 7：城区基站用户热力图

3.1.4 基于建筑级别的 3D 建模可视化

利用爬取的 Floor 字段属性作为楼高，在 GIS 软件拔高楼宇，创建专题地图，以关联的指标数据作为颜色规则渲染地图，得到直观的三维建筑模型，这种地图可以 360 度旋转，自由的垂直视角调整，特别适合于特定场景的分析、规划，实现覆盖的精准定位，并助力规划建设。



图 8：城市 3D 建筑网络流量热力图

3.1.5 基于建筑楼层级别的 3D 建模可视化

对城市建筑的批量建模，了解无线环境，以上帝视角观察城市每一个角落，可以自由编写规则渲染覆盖、指标性能等，助力优化和规划建设，直观的地图呈现在汇报工作中让领导一目了然，提高决策精准度，提升投资效益比。



图 9：建筑楼层级别 3D 建模

id	name	dt_type	r_typecode	address
BOFFH06A2	厦门鑫利隆网络科技有限公司企业公司	公司	170200	集美中融景苑大厦5楼仁室
BOFFH0R59	泉州丰泽福汇通餐饮公司企业公司	公司	170200	金街万达广场(福源酒店)一楼
BOFFH03153	香格里拉酒店(宝龙)购物服务购物相关场所购物相关		500000	世纪大道宝龙城市广场5024号店
BOFFH0AZJ5	泉州91商务酒店	住宿服务住宿服务相关住宿服务	1000000	龙河锦隆中庭路中小区15号楼
BO25304JCS	拉卡萨巴西烤肉自助餐服务外带餐厅外带餐		50215	泉州城东段大自融酒店2层
BO25302A9K	泉州崇德审计师事务所生活服务事务所审计事务所		70704	龙溪路番塔34号
BOFFH07R7E	泰山山花苑-南区	商务住宅住宅区住宅区	120300	安吉南路与南营街交叉口西150米
BOFFF2Y571	鲤城实验小学城市号	科教文化服务文化文化室	140800	兴才路2号
BOFFH0K2JG	饮水处(泉州站)	室内设施室内设施室内设施	970000	东西大道泉州站B2层
BO2530C08Y	鼎盛汽车城	汽车销售汽车销售汽车销售	20000	迎宾大道附近
BOFFH0W7M2	鼎盛汽贸有限公司	汽车销售汽车销售汽车销售	20000	站前大道福源小区对面多种中心门口
BOFFH419VQ	泉州开市二手车	汽车销售汽车销售汽车销售	20000	站前大道与南营路交叉口东南100米
BO253034CQ	亿门摩配城	摩托车服务摩托车服务相关摩托	40000	城华南路50米
BO25304F8C	龙门摩配商行	摩托车服务摩托车服务相关摩托	40000	浮桥街道东兴楼附近(浮桥商会大楼北)
BOFFGA839E	阿来摩托车电动车	摩托车服务摩托车维修摩托车	40200	博东路与玉书路交叉口东100米
BO2530C0SV	万通车行	摩托车服务摩托车维修摩托车	40200	浮桥强火地工业区康安街
BOFFH0K0PQ	小南摩托车电动车	摩托车服务摩托车维修摩托车	40200	前峰路与博东路交叉口东50米



图 11：用户轨迹

按照经常性、清晰性和全面性原则对 POI 的 22 大类、261 小类社区公共服务设施类型进行筛选：①经常性：包括居民日常生活所需的公共服务设施；②清晰性：即具有类别描述清晰、指向明确的公共服务设施；③全面性：能够广泛地囊括某种类型的公共服务设施。依据上述原则，研究筛选出 8 大类、24 小类公共服务设施，整理形成 5645 个 POI 数据。

大类	小类	数量	大类	小类	数量	
A 餐饮	A01 茶艺馆	88	C 体育	C04 健身房	60	
	A02 糕饼店	258		D 生活	D01 美容美发店	1323
	A03 咖啡厅	119	D02 摄影冲印		139	
	A04 快餐厅	521	E 医疗		E01 药房	249
	A05 冷饮店	266			E02 诊所	184
	B 购物	A06 甜品店	33	F 教育	F01 小学	35
B01 便民商店		1018	F02 图书馆		22	
B02 超级市场		140	F03 幼儿园		134	
C 体育	B03 综合市场	878	G 休闲	G01 网吧	81	
	C01 篮球场馆	3		G02 棋牌室	36	
	C02 乒乓球馆	6		G03 公园广场	34	
C03 羽毛球馆	4	G04 疗养院		14		

3.1.6 性能感知指标可视化应用

通过链路预算模型关联建筑物的主覆盖小区 ID，结合小区 SEQ 感知指标（终端 RTT，流量等）实现基于建筑物的感知评估。

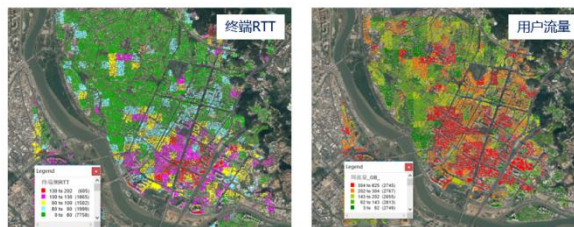
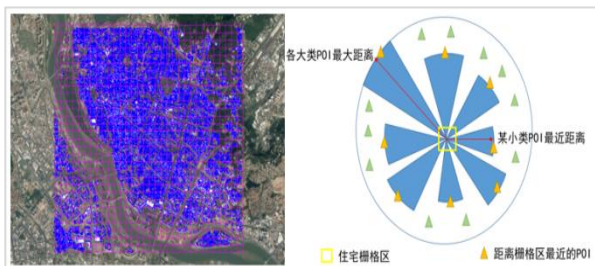


图 10：性能指标 3D 建筑可视化

3.2 用户轨迹可视化

个体一般会选择距离最近的公共服务设施开展活动。因此，从一个居住地出发，到能满足日常生活需求的公共服务设施的最短距离，可以认为是模拟现状出行轨迹，即生活圈的空间范围，反映了人口密集及高价值区域。通过爬取的海量 POI 整理并进行数据清洗。

对泉州中心市区进行评估,取正方形约 37.71 平方公里,以 200 米大小栅格化划分,以每个栅格中心为出发点,利用缓冲区分析及空间统计等方法,测度每个栅格到 POI 点的最近距离,将包含 8 大类的 POI 最近距离中的最大值作为生活圈的半径,其中半径按成人步行 15 分钟约 1 公里为限进行划分:



经过统计分析,各类占比如下:

生活圈	大类POI占比	建筑占比
1类(1公里)	3.4%	4.9%
2类(2公里)	14.3%	18.4%
3类(3公里)	15.8%	17.4%
4类(4公里)	13.5%	13.9%
5类(5公里)	13.8%	11.2%
6类(5公里以外)	39.2%	34.2%

通过研究分析 POI, 构建了反映居民日常行为的生活圈, 统计范围内的公共服务设施数量, 一定程度上反映了生活圈内人口的密集程度及高价值区域。

对栅格内的建筑小区按生活圈类型渲染, 1 类生活圈为人口密度最大, 5 类人口密度最小, 6 类处于评估截取地界边缘, 误差会较大, 可按实际情况考虑是否列入评估结果:



图 13: 3D 建筑用户热力图可视化

4. 总结

在网络优化工作中引入互联网爬虫和大数据分析技术, 开始网络优化的互联网转型, 通过互联网+大数据实践应用提升网络优化质量并助力市场决策, 实现精细化营销策略, 提升企业的核心竞争力。利用爬取的数据与 MR、网络性能指标等关联分析, 评估网络性能, 识别高价值区域, 指导精准规划建设, 提升投资效益。

参考文献

[1] 埃里克·马瑟斯. Python 编程从入门到实践[M]. 人民邮电出版社.2016.07

[2] 黑马程序员. Python 应用编程丛书:解析 Python 网络爬虫:核心技术、Scrapy 框架、分布式爬虫[M]. 中国铁道出版社.2018.08

[3] 卫斯理. Python 核心编程[M]. 人民邮电出版社.2016.05

[4] Chris Garrard. Python 地理数据处理 [M]. 北京:人民邮电出版社, 2017.

[5] Joel Lawhead. Python 地理空间分析指南 (第 2 版) [M]. 北京:人民邮电出版社, 2017.

[6] 朱秀丽, 周治武, 李静, 等. 网络矢量地图瓦片技术研究 [J]. 测绘通报, 2016(11)

[7] 章谦骅, 章坚武. 一种新的射线追踪方法-测试镜像跟踪法. 微波学报 2014 年 6 月

智能算力驱动下的数字经济产业融合： 机遇与挑战

李 晨

(中国电信股份有限公司福建分公司)

摘要：随着数字经济的快速发展，智能算力作为一种重要的驱动力量，在数字经济产业融合中扮演着关键的角色。本论文旨在探讨智能算力驱动下的数字经济产业融合所带来的机遇和面临的挑战。首先，论文将回顾智能算力的定义和发展趋势，深入探讨其在数字经济中的重要性和应用领域。其次，通过案例研究和分析，论文将探讨智能算力在不同行业（如互联网、制造业、金融等）中的具体应用，并阐述其对产业融合发展的促进作用。进一步，论文将分析智能算力驱动下的数字经济产业融合所面临的挑战，包括数据隐私与安全、人才需求、技术标准等方面，并提出相应的解决方案。最后，论文将总结智能算力驱动下的数字经济产业融合的机遇与挑战，并探讨未来发展的趋势和前景。本论文旨在为相关领域的研究者、政策制定者和企业提供参考，推动智能算力的应用与发展，促进数字经济产业融合的持续推进。

关键词：智能算力、数字经济、产业融合、数字化转型

引言

背景介绍：随着信息技术的飞速发展和数字经济的兴起，福建省积极推动数字经济发展，将其作为实现高质量发展的重要引擎。福建省作为中国东南沿海经济发达地区之一，具有丰富的自然资源和优越的地理位置优势。近年来，福建省加大了数字化转型的力度，不断推进产业融合发展，将数字经济与传统实体经济相结合，实现协同发展和创新驱动。在这个背景下，智能算力作为数字经济发展的关键支撑和驱动力量，正逐渐发挥着关键作用。

福建省数字经济发展现状和重要性：福建省在数字经济发展方面取得了显著的成就。目前，福建省已形成了以电子商务、云计算、大数据、人工智能等为核心的数字经济产业体系。福建省数字经济产业增速持续较快，对地方经济的拉动效应显著。数字经济产业融合的深入发展，为福建省的经济结构优化、创新能力提升、产业升级和就业增长提供了新的动力。

研究目的与意义：本研究的目的是探讨智能算力在福建省数字经济产业融合中的关键作用，并分析其对福建省经济发展的重要性。通过深入研究福建省数字经济的现状和智能算力的发展趋势，将揭示智能算力在推动福建省数字经济产业融合中的关键作用和挑战，并提出相应的对策和建议，为福建省的数字经济发展和产业升级提供决策参考。

1. 福建省数字经济产业融合与智能算力

1.1 福建省数字经济发展概况

福建省作为中国沿海经济带的核心省份，在数字经济发展方面取得了显著成就。福建省拥有丰富的信息技术人才和优越的地理位置，以及强大的制造业基础和活跃的互联网应用场景。数字经济在福建省的发展涉及多个领域，包括电子商务、云计算、大数据、人工智能等。福建省数字经济产值不断增长，对推动福建省经济增长、提升产业竞争力起到了重要的推动作用。

根据福建省经济信息中心发布的《福建省数字经济发展指数评价报告(2022 年)》，在数字发展基础方面，福州和厦门优势明显，特别是在传统基础设施和新型基础设施方面表现出色。数字技术创新方面，福州、厦门和泉州位列前列，其中福州在数字创新项目和数字技术应用方面得分突出。厦门、福州在数字社会应用方面位居前列，福州的数字民生应用得分全省最高。在数字治理水平方面，厦门、福州领先，尤其福州在数字治理手段指标上全省领先。福建省数字经济发展的成功得益于政府政策支持和企业投资，数字经济产业结构不断优化，新动能不断增强。然而，数字经济发展仍面临数据隐私与安全、人才需求与培养、技术标准与互操作性等挑战，需要加强相关对策。展望未来，福建省将进一步推动数字经济发展，加强跨界合作与开放创新，努力实现经济高质量增长和产业转型升级。

1.2 福建省数字经济产业融合的现状与挑战

根据福建省经济信息中心 2023 年初发布的数据，福建省的数字经济发展取得了显著成就。数字经济规模超过 2.3 万亿元，占全省 GDP 比重超过 47%，彰显了数字经济在福建省经济中的重要地位。

福建省数字经济发展的现状显示出以下特点和挑战：

首先，福建省在数字经济发展基础方面具备优势。特别是福州和厦门在数字发展基础方面位居前列，包括传统基础设施和新型基础设施的建设，为数字经济的发展提供了坚实支撑。

其次，福建省在数字技术创新方面取得了显著成果。福州、厦门和泉州位列数字技术创新方面的前列，特别是福州在数字创新项目和数字技术应用方面得分突出。这反映了福建省在数字技术研发和应用方面的积极探索和创新能力。

第三，福建省在数字社会应用方面取得了显著进展。厦门和福州在数字社会应用方面位居前列，其中福州的数字民生应用得分全省最高。这表明福建省在数字技术应用于民生服务和社会治理方面取得了显著成效。

然而，福建省数字经济产业融合发展仍面临一些挑战。其中，数字经济顶层设计统筹协调不

足是一个突出问题。虽然各地纷纷制定数字经济发展规划和措施，但缺乏整体统筹和差异化发展，导致同质化现象明显，特色化和个性化不足。此外，数字经济营商环境还需要进一步优化，以促进数字经济发展的活力和创新能力。同时，数字经济新动能的培育也面临巨大挑战，包括国际环境的不确定性风险和国内经济受疫情重创等问题。

为了应对这些挑战，福建省可以采取以下措施：

首先，加强数字经济发展的顶层设计和统筹规划，避免同质化现象，注重特色化和个性化发展。各地可以结合自身优势和特点，明确发展方向和重点领域，推动数字经济产业融合发展的差异化发展。

其次，进一步优化数字经济营商环境，提供有利于数字经济发展的政策和制度支持。加快构建符合国情的数字经济发展长效机制，减少市场准入壁垒，提高数字素养，鼓励创新创业，打造有吸引力的数字经济创业环境。

此外，积极推进数字经济发展手段的创新，包括不同领域的数字基础设施建设和监管政策的差异化，结合技术趋势加强数字技术在不同领域中的创新应用。

福建省在数字经济产业融合发展方面取得了显著成就，但仍面临顶层设计统筹协调不足、营商环境优化和新动能培育等挑战。通过加强统筹规划、优化营商环境和创新手段，福建省可以进一步推动数字经济产业融合发展，为经济高质量增长和产业转型升级注入新动力。

1.3 智能算力在福建数字经济产业融合中的关键作用

智能算力在福建数字经济产业融合中发挥着关键作用。近年来，随着新一代信息技术的迅猛发展，特别是人工智能、云计算和 5G 等核心科技的崛起，智能算力成为推动产业升级和经济高质量增长的重要引擎。

首先，智能算力为数字经济提供了强大的技术支撑。在数字经济时代，海量数据和丰富的应用场景要求高效的数据处理和分析能力。智能算力通过强大的计算和存储能力，使得数据的处理和分析变得更加高效和精确。它为各行各业的数

数字化转型提供了技术保障，推动了实体经济与数字技术的深度融合。

其次，智能算力加速了人工智能技术的应用和发展。人工智能作为新一代关键性技术能力，正在成为推动国家数字化经济发展的核心推动力。智能算力为人工智能算法和模型的训练、推理和优化提供了强大的计算能力，使得人工智能在各个领域得以广泛应用。例如，在制造业中，智能算力支持智能质检和生产过程的优化；在能源领域，智能算力提供智能巡检和能源管理的解决方案；在交通行业，智能算力推动智慧交通和智能运输的发展。智能算力的发展为人工智能产业带来了新的增长点，促进了各行业的数字化升级和创新。

此外，智能算力在数字经济产业融合中还具有更多潜力。随着 5G 网络的普及和应用，边缘计算和物联网等技术的发展，智能算力的边缘部署和分布式计算能力将更加重要。它可以支持智能设备和传感器的实时数据处理和决策，推动工业互联网和智能城市等领域的发展。同时，智能算力还有助于构建开放的数字平台和生态系统，促进不同行业和企业之间的协同合作和创新共享。

总之，智能算力在福建数字经济产业融合中发挥着关键作用。它提供了强大的技术支撑，加速了人工智能技术的应用和发展，并为数字经济产业融合带来更多的潜力。福建省可以利用智能算力的优势，加强技术创新和产业合作，推动数字经济的健康发展，实现经济高质量增长的目标。

2. 福建省智能算力发展现状与趋势

2.1 福建省智能算力资源布局与发展态势

福建省智能算力的发展得益于政府的支持和产业布局。福建省在智能算力资源的布局上积极推动建设智能算力中心和数据中心，建立了一批具备较高算力和数据存储能力的基础设施。福建省还吸引了一批国内外知名的云计算和人工智能企业进驻，为智能算力的发展提供了强有力的支撑。

福建省智能算力的发展态势十分乐观。随着福建省数字经济的快速发展和产业融合的深入推

进，对智能算力的需求日益增长。福建省在智能算力资源布局上不断完善，加大投入力度，提升算力技术和服务质量。福建省的智能算力产业正呈现出蓬勃发展的态势，成为推动福建省数字经济发展的引擎。

2.2 福建省智能算力技术创新与应用

福建省在智能算力技术创新和应用方面取得了显著的成就。福建省的高校、科研机构和企业积极推动智能算力相关技术的研发与创新，涵盖了云计算、大数据分析、人工智能等多个领域。

在云计算方面，福建省建设了一批具有较强算力和服务能力的云平台，为企业提供弹性的计算资源和高效的数据存储服务。福建省的云计算企业也积极开展技术创新，推出了一系列创新产品和解决方案，为福建省数字经济的发展提供了支持。

在大数据分析方面，福建省探索了多种数据挖掘和分析方法，应用于不同行业的数据处理和决策支持中。福建省的企业利用智能算力技术，开展了精准营销、智能制造等方面的创新应用，提高了企业的竞争力和运营效率。

在人工智能方面，福建省加强了人工智能技术的研发和应用。福建省的企业在人脸识别、语音识别、自然语言处理等领域取得了重要突破，为福建省的安防、智慧交通、智慧医疗等领域的应用提供了支持。

2.3 福建省智能算力发展的未来趋势

福建省智能算力的未来发展呈现出一些明显趋势。首先，福建省将继续加大对智能算力资源的投入和布局，进一步提升算力技术和服务能力。福建省将继续打造智能算力中心和数据中心，推动云计算和人工智能技术的创新和应用。

其次，福建省将加强智能算力与其他新兴技术的融合，如 5G、边缘计算等。通过将智能算力与其他技术结合，福建省将推动更多领域的数字化转型和创新。

此外，福建省还将注重智能算力人才培养和引进，加强与高校和科研机构的合作，培养更多具备智能算力相关技术的人才队伍，推动智能算力技术的创新和应用。

综上所述,福建省智能算力的发展现状良好,未来趋势十分乐观。福建省将继续加大对智能算力的支持力度,推动其在数字经济产业融合中的关键作用,为福建省的经济发展注入新的活力。

3.福建省智能算力在关键产业的应用与融合

3.1 福建省互联网行业中的智能算力应用与融合发展

福建省的互联网行业在数字经济产业融合中发挥着重要作用,而智能算力的应用和融合更是推动福建省互联网行业发展的关键驱动力。

首先,在电子商务领域,福建省的互联网企业通过智能算力技术实现了更高效的供应链管理和订单处理,提供了个性化的推荐和购物体验。智能算力技术可以分析用户行为和偏好,精准推荐商品,提高用户满意度和购买转化率。

其次,在在线教育领域,福建省的互联网平台利用智能算力技术实现了智能化的教学和评估。通过对学生学习数据的分析和挖掘,智能算力可以为学生提供个性化的学习推荐和辅导,提高学习效果和教学质量。

此外,在智慧城市建设中,福建省的互联网行业利用智能算力技术实现了智能交通、智慧物流等应用。通过智能算力的支持,福建省的交通管理可以实现实时交通监控和优化调度,提高交通效率和安全性;智能物流可以实现快递配送的路径规划和优化,提高物流效率和准确性。

总之,福建省互联网行业中智能算力的应用和融合发展为数字经济产业融合提供了强有力的支持,推动了福建省互联网行业的创新和升级。

3.2 福建省制造业中的智能算力应用与融合发展

福建省的互联网行业在数字经济产业融合中发挥着重要作用,而智能算力的应用和融合更是推动福建省互联网行业发展的关键驱动力。

首先,在电子商务领域,福建省的互联网企业通过智能算力技术实现了更高效的供应链管理和订单处理,提供了个性化的推荐和购物体验。智能算力技术可以分析用户行为和偏好,精准推荐商品,提高用户满意度和购买转化率。

其次,在在线教育领域,福建省的互联网平

台利用智能算力技术实现了智能化的教学和评估。通过对学生学习数据的分析和挖掘,智能算力可以为学生提供个性化的学习推荐和辅导,提高学习效果和教学质量。

此外,在智慧城市建设中,福建省的互联网行业利用智能算力技术实现了智能交通、智慧物流等应用。通过智能算力的支持,福建省的交通管理可以实现实时交通监控和优化调度,提高交通效率和安全性;智能物流可以实现快递配送的路径规划和优化,提高物流效率和准确性。

总之,福建省互联网行业中智能算力的应用和融合发展为数字经济产业融合提供了强有力的支持,推动了福建省互联网行业的创新和升级。

3.3 福建省金融业中的智能算力应用与融合发展

福建省的金融业在数字经济产业融合中发挥着重要作用,而智能算力的应用和融合为福建省金融业的创新和发展提供了关键支撑。

首先,在风险管理方面,福建省的金融机构利用智能算力技术实现了智能风险评估和风控管理。通过智能算力的支持,福建省的金融机构可以对客户的信用风险进行预测和评估,提供个性化的风控方案,降低风险损失。

其次,在智能客服方面,福建省的金融机构利用智能算力技术实现了智能化的客户服务和咨询。通过智能算力的支持,福建省的金融机构可以实现智能语音识别和自然语言处理,提供个性化的金融服务,提高客户满意度和服务效率。

此外,在智能投资方面,福建省的金融机构利用智能算力技术实现了智能化的投资决策和资产配置。通过智能算力的支持,福建省的金融机构可以对市场数据进行分析和挖掘,提供智能化的投资建议,提高投资效果和风险控制能力。

3.4 其他福建省重要行业中的智能算力应用与融合发展案例

除了互联网行业、制造业和金融业,福建省的其他重要行业也积极应用智能算力技术,取得了一些实实在在的发展成果。

在农业领域,福建省农业企业借助智能算力技术实现了精准农业的发展。通过智能算力的支

持,福建省的农业企业可以对土壤、气象、水资源等农业生产数据进行精准分析和预测,制定科学的农业生产方案和管理策略,提高农产品的产量和质量。

在旅游行业中,福建省的旅游企业利用智能算力技术提升了旅游服务的品质和体验。通过智能算力的支持,福建省的旅游企业可以为游客提供个性化的旅游线路规划、景点推荐和导游服务,提升游客的旅游体验和满意度。

在文化创意产业中,福建省的文化企业运用智能算力技术推动了创意产品的设计和开发。通过智能算力的支持,福建省的文化企业可以进行大规模的创意设计和艺术创作,提高创意产品的研发效率和市场竞争力。

这些实际案例展示了福建省智能算力在农业、旅游和文化创意等重要行业中的应用与融合发展。通过智能算力技术的运用,福建省不仅提升了行业的效率和竞争力,还为福建省经济的转型升级注入了新动力。

4.福建省智能算力驱动下的数字经济产业融合挑战与对策

4.1 数据隐私与安全挑战及解决对策

在数字经济产业融合中,数据的隐私和安全是一个重要挑战。随着智能算力的应用和数据的大规模收集和处理,个人隐私泄露的风险增加,网络安全威胁也日益严峻。

为应对这一挑战,福建省可以采取一系列对策。首先,建立完善的数据隐私保护法规和政策,规范数据的收集、使用和共享,保护个人隐私不受侵犯。其次,加强数据安全技术的研发和应用,包括数据加密、身份认证和访问控制等,确保数据在传输和存储过程中的安全性。此外,加强对数据泄露和网络攻击的监测和应急响应能力,及时发现和应对安全威胁。

4.2 人才需求与培养挑战及解决对策

数字经济产业融合对人才的需求提出了新的挑战。福建省需要培养一批既懂技术又懂产业的复合型人才,能够熟练运用智能算力技术推动各行业的融合发展。

为解决人才需求与培养的挑战,福建省可以

采取多种对策。首先,加强智能算力相关专业的人才培养,包括开设相关专业课程、组织培训和实践项目,提高学生在智能算力领域的专业素养和实践能力。其次,与高校、科研机构和企业建立紧密的合作关系,共同培养智能算力领域的人才,促进科技创新和产业协同发展。此外,加强人才引进和人才流动,吸引具有智能算力背景和经验的专业人才来到福建省,为产业融合发展注入新的活力。

4.3 技术标准与互操作性挑战及解决对策

在数字经济产业融合中,不同行业、不同企业之间存在着技术标准的差异,互操作性成为一个挑战。福建省需要解决技术标准和互操作性的问题,促进各个行业之间的协同发展。

为解决这些问题,福建省可以推动制定智能算力相关的技术标准,促进行业间的互通互联。同时,加强与相关行业协会、标准化组织的合作,参与国内外标准制定和制定过程,推动福建省的技术标准与国际接轨,提高福建省产业的竞争力和影响力。

4.4 法律法规与政策支持挑战及解决对策

在数字经济产业融合中,法律法规和政策的支持是保障产业发展的重要保障。福建省需要建立健全相应的法律法规体系,明确智能算力的应用范围和规则,促进产业的规范发展。

为解决法律法规与政策支持的挑战,福建省可以加强对相关法律法规的研究和制定,及时跟踪国家和国际的发展趋势,根据福建省的实际情况进行适应性的调整和完善。同时,提供政策支持,包括财政资金、税收减免和创新创业支持等,鼓励企业加大对智能算力的投入,推动产业的创新和升级。

5.福建省智能算力发展的政策与支持

5.1 福建省智能算力产业政策概述

福建省积极出台政策和措施,推动智能算力产业的发展。其中,福建省制定了一系列智能算力产业发展规划和政策文件,包括鼓励企业投资智能算力技术研发、推动智能算力技术与行业融合发展、支持人才培养和创新创业等方面。

此外，福建省还设立了智能算力产业发展专项资金，用于支持智能算力技术的研发与应用，鼓励企业加大投入，推动产业的快速发展。同时，福建省还积极引导和支持智能算力企业参与国家和省级的科技创新项目，提供更多的资金和政策支持。

5.2 政府与企业合作促进智能算力发展的成功案例

福建省政府与企业之间的合作对于智能算力发展起到了积极的推动作用。政府提供政策支持和资金扶持，企业则积极投入研发和创新，取得了一些成功的案例。

一个成功的案例是福建省与某互联网巨头合作推动智能算力在物流领域的应用。政府提供政策支持和资金补贴，鼓励企业投资智能算力技术的研发和应用。企业通过智能算力技术实现了物流数据的实时监控和智能调度，提高了物流运输的效率和准确性，降低了物流成本。这一合作案例推动了福建省物流行业的升级和数字化转型。

另一个成功的案例是福建省与某创新企业合作推动智能算力在农业领域的应用。政府鼓励企业投资智能算力技术的研发和应用，提供资金和政策支持。企业利用智能算力技术实现了农业数据的精准分析和农业生产的智能化管理，帮助农民提高农产品的产量和质量，实现农业的可持续发展。这一合作案例推动了福建省农业领域的现代化和智能化。

5.3 加强跨界合作与开放创新的建议

为进一步推动福建省智能算力的发展，建议加强跨界合作与开放创新。福建省可以鼓励不同行业、不同领域的企业共享智能算力技术和资源，推动产业的融合发展。同时，积极引进国内外的先进技术和经验，加强与其他地区和国家的合作

交流，促进智能算力产业的国际化发展。

6. 结论与展望

6.1 总结论文的主要观点和研究发现

本论文主要围绕福建省智能算力在数字经济产业融合发展中的应用与挑战展开研究。通过对福建省数字经济产业融合现状的分析，我们发现智能算力在推动产业融合发展中发挥了重要作用。福建省在互联网、制造业、金融业等关键行业中成功应用智能算力技术，取得了一些实际案例。

同时，我们也发现在智能算力发展过程中存在一些挑战，包括数据隐私与安全、人才需求与培养、技术标准与互操作性、法律法规与政策支持等方面。针对这些挑战，我们提出了一系列对策和建议，包括加强数据隐私保护和安全技术的研究应用、加大人才培养力度、推动技术标准的制定和互通互联、建立健全的法律法规体系和提供政策支持等。

6.2 对福建省智能算力在数字经济产业融合中未来发展的展望

展望未来，福建省智能算力在数字经济产业融合中具有广阔的发展前景。随着技术的不断进步和应用的深入，福建省可以进一步推动智能算力在各个行业中的应用和融合，促进产业的创新和升级，为福建省数字经济发展注入新动能。

6.3 研究的局限性与进一步研究的建议

本研究虽然对福建省智能算力在数字经济产业融合发展中的发展进行了探讨，但仍存在一定的局限性。进一步研究可以拓展以下方面：深入研究福建省智能算力在其他行业中的应用案例、分析福建省与其他地区的智能算力发展对比、探索智能算力与人工智能、区块链等前沿技术的融合等。

算云融合赋能千行百业

夏少良

(中电福富信息科技有限公司)

摘要: 随着数字化时代的到来,计算需求呈现爆发式增长,从科学研究到商业应用,从工业制造到金融服务,各行各业对于高性能、灵活可扩展的计算能力的需求日益迫切。云计算的发展促进了计算资源的共享与集中,提供了强大的计算能力和数据处理能力。算力和经济增长的模式正相关,一个国家或者区域想要快速发展,算力达不到便无法实现。在此背景下,算力与云计算的融合成为实现智能化发展和推动数字经济产业发展的重要途径。

然而,算力与云计算融合也面临着一些挑战,如数据安全与隐私保护、网络带宽与延迟的限制等。为了实现智能算力赋能千行百业的目标,需要进一步研究和解决这些挑战,推动算力与云计算融合的发展。

关键词: 算力、云计算、融合

引言

算力与云计算的背景:在过去的几十年里,信息技术的快速发展以及互联网的普及使得数据量和计算需求呈指数级增长。传统的计算资源往往受限于本地硬件设备的性能和规模,无法满足大规模、复杂计算任务的需求。同时,许多企业和组织需要灵活的计算能力来应对不断变化的业务需求。云计算的出现为解决这些挑战提供了一种新的解决方案。

云计算是一种基于互联网的计算模式,通过提供按需分配和释放计算资源的能力,以弹性和灵活的方式满足用户的计算需求。用户可以根据实际需求,动态地调整计算资源的规模和配置,从而提高计算效率和资源利用率。算力是指计算能力,它代表着处理计算任务的能力。传统的算力技术包括多核处理器、加速器(如GPU和FPGA)等。

随着数据量和计算需求的进一步增长,传统的云计算模式面临一些挑战。大规模数据处理、高性能计算和实时计算等应用场景对计算能力的要求越来越高。为了满足这些需求,将传统的算力技术与云计算相结合成为一种趋势。

算云融合的目的与意义:随着信息技术的迅猛发展和互联网的广泛应用,计算需求呈现指数级增长,对于高性能、灵活可扩展的计算能力的需求日益迫切。一些研究数据已经表明,算力和经济增长的模式正相关。算力每增加一个点,经济可能也会增加近一个点,在整个经济活动以及高质量发展中,算力是非常关键的领域。[1]在这一背景下,算力与云计算的融合成为实现智能化发展和推动数字经济产业融合发展的重要途径。通过本论文的研究,我们将揭示算力与云计算融合的实际应用价值,为智能算力赋能千行百业的目标提供新思路 and 见解,为各行各业的数字化转型和智能化发展注入新动力。

1. 算力与云计算概述

1.1 算力的定义与发展

算力是计算机或计算系统在一定时间内执行计算任务的能力。它通常用于衡量计算机处理速度、吞吐量和执行复杂计算任务的能力。算力的衡量指标可以包括浮点运算能力(FLOPS,每秒浮点运算次数)、整数运算能力(IOPS,每秒整数运算次数)、内存访问速度等。算力的提升意

意味着计算机可以更快速、高效地完成复杂的计算任务，为科学研究、工程设计、数据分析和人工智能等领域提供强大的支持。算力在经济、科技、社会三方面具有重要价值，《2021-2022 全球算力指数评估报告》显示一个国家或地区的算力指数平均每提高 1 个点数字经济和 GDP 将分别增长 3.5%和 1.8%:[2]

1.2 云计算的基本概念与特点

云计算是一种通过互联网提供计算资源和服务的模式。它将计算能力、存储资源、数据库、网络服务等以服务的形式提供给用户，使用户可以根据需要按需获取和使用这些资源，而无需关心资源的实际物理位置和管理细节。云计算基于虚拟化技术，将底层的硬件资源抽象成虚拟的计算资源池，以便更高效地共享和利用。国际标准化组织 ISO/IECJTC101 定义云计算是可以提取的、高级的、可升级的池，能够为终端用户提供主机应用和通过消费买单的管理计算基础设施”[3],具备弹性与灵活性、自助服务、资源共享、可伸缩性、可靠性与高可用性、按需付费、跨平台与跨设备访问等特点。

1.3 算力与云计算的融合方式

算力和云计算相结合，使得计算任务的规模和计算能力之间可以动态匹配，提高了计算资源的利用效率。可以实现弹性计算资源与高性能计算能力的整合，满足不同规模和复杂度的计算需求，推动各行各业的数字化转型和智能化发展，算力与云计算的融合可以通过以下方式实现：

- (1) 弹性计算资源与高性能计算能力的整合
- (2) 加速器技术的应用
- (3) 异构与超异构计算架构的利用
- (4) 边缘计算与云计算的协同

2.算力与云计算融合的优势与挑战

2.1 优势

算力与云计算的优势在于高性能计算能力、弹性计算资源、节约成本、大规模数据处理与分析以及支持多样化计算任务等方面，通过对算力与云计算技术相融合，不仅可以有效提升企业对数据信息的计算效率，同时更可以推动社会的经

济发展。[4]这些优势为企业、科研机构和个人提供了强大的计算支持，推动了各行各业的数字化转型和智能化发展。

2.2 挑战

云计算面临性能挑战、安全性挑战、成本问题三方面的挑战。算力面临算力产业基础依旧薄弱、算力技术创新不足、算力供需匹配失衡等挑战[5]，面对这些挑战，云计算平台提供商和用户都需要采取相应的策略和措施。云计算提供商应不断优化平台性能，增强安全性能，提供更具竞争力的计费模式，以满足用户不断增长的计算需求。同时，用户应该认真评估自身算力的需求，采取合理的算力资源规划和管理，以最大限度地发挥算力与云计算融合的优势，并克服挑战。通过克服这些挑战，算力与云计算融合将为各行各业提供更强大、高效的计算支持，推动数字化转型和智能化发展的进程。

3.算力与云计算融合的应用实践案例

3.1 东数西训的应用案例

3.1.1 背景

根据“十四五规划”要求，需“加快数字化发展，建设数字中国”，建设数字中国是数字时代推进中国式现代化的重要引擎，是构筑国家竞争新优势的有力支撑。随着 AI 应用场景的高速发展，赋智应用场景激增。模型训练，需要海量训练数据，和专用 GPU 算力资源，是典型的大计算量场景，对 AI 算力提出新需求。

3.1.2 目标

依托于算力调度引擎能力，服务使用方调用 AI 推理服务，算网调度引擎在接收到服务请求后，实时判断各个算力资源池忙闲，通过“路由调度策略”调度到恰当的算力池，分别实现“数据调度”和“训练调度”，完成数据的西迁、异地的模型训练。实现以下三大目标：

(1) 提能：提升算力发展水平，促进算力转化为更大生产力。

(2) 创新：牵引战略和业务转型促进技术创新和变革。

(3) 降本：算力资源最大化利用，实现国家、企业降本增效。

3.1.3 架构设计

面向多样化的异构算力，构建分布式云原生调度服务，实现在使用多样化算力时，感受不到地域、环境、技术栈、资源类型的差异，提升算力使用的便捷性及一致性。[6-9]具备以下四大特性：

- (1) 汇聚全局资源，可视可管：
不同区域存算力资源、不同类型的存算力资源、不同云存算力资源实现一点汇聚、一点看全看清。
- (2) 融合三力调度，最优部署：

对存算力进行温冷热分级管理，根据 SLA 要求，计算出最优的部署路径。

- (3) 统一多云操作，极致体验：
面向使用者提供一致的部署体验、一致的运维观测体验、一致的操作体验。
- (4) 智能运维，故障自愈：
综合应用故障感知、定界、决策和自愈 AIOps 场景能力，打造智能化故障治理体系。

总体设计如图 4-1-3-1 所示。

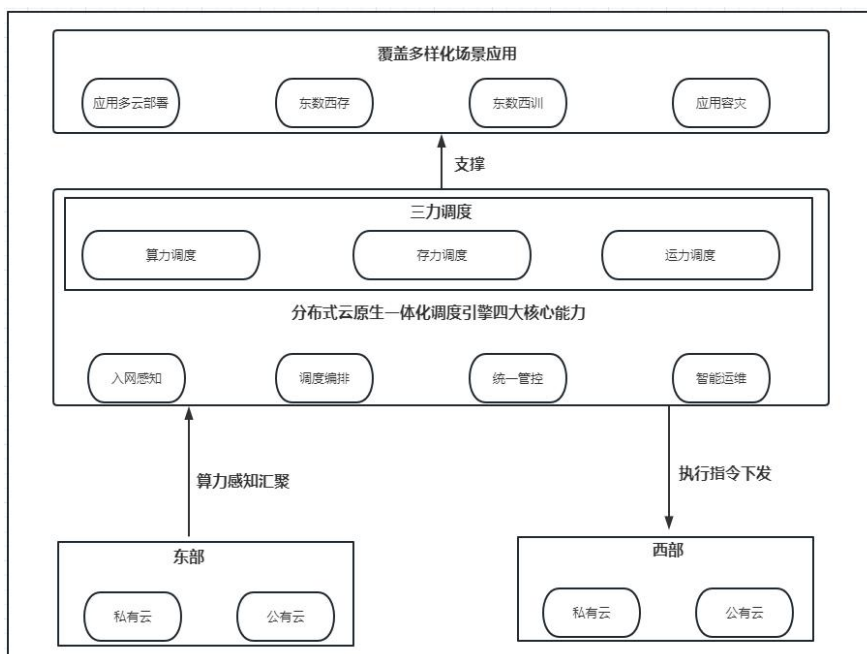


图 4-1-3-1 总体设计图

借助于算力调度平台的“算力分析”、“算力调度”以及“数据存储调度”的能力，完成模型训练场景的算力调度的工作，在西部算力资源充足的情况下，实现“西训”的目的。具备以下四大功能：

- (1) 任务集中编排：
训练/推理任务一处编排，多处执行。
- (2) 任务自动调度：
考虑任务资源需求，结合算力空闲及数

据位置，实现自动按需调度。

- (3) 任务集中监控管理：
跨地域多中心任务的全生命周期集中监控，了解任务执行情况。
- (4) 推理服务跨中心调度：
训练模型跨中心使用，推理服务的跨中心部署与调度。

场景架构设计如图 4-2-3-2 所示。

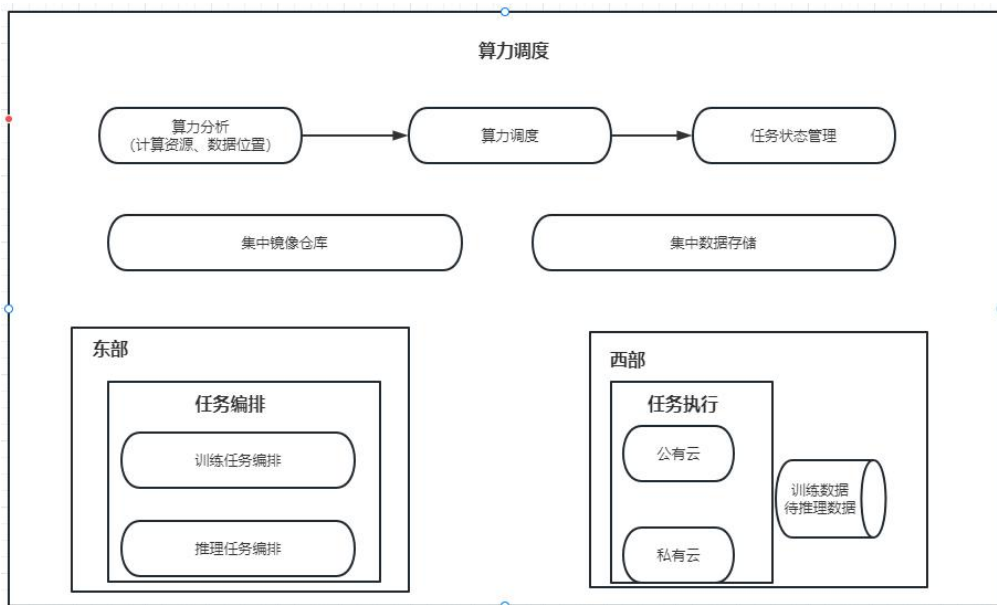


图 4-2-3-2 场景架构设计图

3.1.4 经验及应用推广

通过算力与云计算的融合，业务落地经验展现了三个关键方面：首先，任务一处编排多处执行的策略，实现训练和推理任务的灵活部署，同时跨中心调度资源，有效提升计算效率与资源利用率；其次，算力资源的自动分析与任务自动调度的经验，建立智能化调度系统，优化资源利用，提升任务执行效率，满足业务高速发展需求；最后，训练与推理流程跨地域多中心串接，实现计算任务的无缝衔接，为各个场景的智能化应用提供稳定可靠的计算支持。

在应用价值方面，融合带来的两个关键优势是：首先，算力供需互利，提升整体资源利用率。充分利用空闲资源池，实现 AI 场景的计算推理，保障 AI 业务稳定运营的同时，提高空闲资源中心的资源利用率，实现资源价值最大化。其次，降本增效，满足业务高速发展。充分利用西部资源中心低成本的优势，扩大基础资源池，为智能场景的高速发展提供最优的硬件投入，有效控制成本，提升业务效率。

场景推广方面，融合在智慧医疗 AI 诊片、智慧交通 AI 指挥和证券趋势分析等场景的应用具有广泛价值。智慧医疗 AI 诊片通过在西部资源池进行模型训练，实现样本数据的实时传输，满足 AI 诊片的实时性要求。智慧交通 AI 指挥则根据历史

路口交通数据，优化红绿灯间隔，有效缓解路口拥堵情况。证券趋势分析则基于历史数据进行离线趋势预测，无实时性要求，处理大量数据，提供有力的决策支持。

4. 算云融合赋能千行百业的实践意义

4.1 提升计算效率与灵活性

算力与云计算的融合为千行百业带来了显著的计算效率与灵活性提升。以下四点是其中的关键点：

(1) 高性能计算能力：云计算平台提供了强大的高性能计算资源，可以支持大规模、复杂的计算任务。用户无需投入大量资金来购买昂贵的计算设备，而是可以根据实际需求租用云计算资源，从而提高计算效率并节省成本。

(2) 弹性计算资源：云计算平台的弹性计算资源允许用户根据需要快速扩展或缩减计算规模。这种灵活性使得企业和组织可以根据业务需求动态调整计算资源，避免资源浪费，实现最优资源配置，从而提高计算的灵活性和适应性。

(3) 容器技术的应用：容器技术如 Docker 和 Kubernetes 等的应用，使得计算环境可以被打包成独立的容器，实现资源隔离和快速部署。通过容器化，用户可以更加方便地迁移和管理计算任务，减

少环境配置的复杂性,进一步提高计算效率和灵活性。

(4) 自动化与智能化:云计算平台提供自动化和智能化的服务接口,使得计算任务的部署和管理可以更加高效和智能。自动化的计算资源调度和任务管理减轻了人工干预的负担,降低了人为错误的风险,提升了计算效率和灵活性。

4.2 加速算力产业开放生态进程

算力与云计算的融合为算力产业开放生态进程提供了强大的加速效应,具体体现在以下四个方面:

(1) 云计算平台的强大计算能力使得算力产业的资源得到高效整合和共享。通过云计算的灵活性,算力产业不再局限于固定的硬件设备,而是可以根据需求实时调整计算资源规模,提高资源利用率,降低成本。

(2) 云计算平台为算力产业带来了全球范围内的资源共享和协作机会。不同地区的算力产业可以通过云计算平台共享技术和知识,加强国际合作,推动全球算力产业的发展和革新。

(3) 算力与云计算的融合也促进了算力产业的数字化转型和智能化升级。云计算平台支持数据驱动的决策和智能算法的应用,使得算力产业能够更好地挖掘数据价值,实现智能化的运营和管理。

(4) 计算平台为算力产业开放生态进程带来了更多的市场机遇。企业和开发者可以通过云计算平台构建自己的服务,将算力和应用无缝结合,开发出更具创新性和个性化的产品,进一步满足市场需求。

4.3 促进算力创新与技术交流

算力与云计算的融合能促进算力创新及加强技术交流,具体体现在以下两个方面:

(1) 算力与云计算的融合为算力产业带来了广阔的合作和创新机遇。云计算平台的全球性和开放性促进了不同地区的算力产业之间的交流与合作,加速技术的传播和共享。

(2) 算力与云计算的融合也促进了跨行业的技术融合和应用拓展。不同行业的企业可以通过云平台共享技术和经验,将算力应用于更广泛的领域,实现技术的互通有无。

5. 总结

本文围绕算力与云计算的融合展开讨论。首先,对算力与云计算的背景进行了介绍,解释了算力是指计算能力的强大和智能化。阐述了算力与云计算融合的目的与意义。在论文的主体部分,系统性地探讨了算力与云计算的概述,包括算力的定义与发展历程,云计算的基本概念与特点,以及算力与云计算的关系与融合方式。此外,还分析了算力与云计算的融合的优势及所面临的挑战,为算力与云计算的进一步发展提供了新的思路和方向。

在实际应用方面,本文以东数西训的应用案例,阐述了背景目标,分析架构设计分享经验,说明了算力与云计算的融合在提能、创新、降本等方面的巨大潜力。从整体角度审视算力与云计算的融合,本论文深入探讨了算云融合赋能千行百业的实践意义。通过提升计算效率与灵活性、加速算力产业开放生态进程、促进算力创新与技术交流,为各行各业带来了巨大的优势与机遇,推动数字经济产业高质量发展。

参考文献

- [1] 高文.高文院士:人工智能和智能算力助力数字经济高效发展[J].高科技与产业化,2023,29(04):12-13.
- [2] 许可,王筑,罗曼婷等.新算力业务动能,新算力经营模式[J].通信企业管理,2023(06):13-19.
- [3] 项华灵,马东伟,张鑫等.云计算标准化赋能数字经济产业发展研究[J].中国标准化,2023(01):53-59.
- [4] 周新华.人工智能、大数据与云计算的融合研究[J].电脑知识与技术,2022,18(09):52-53+62.DOI:10.14004/j.cnki.ckt.2022.0567.
- [5] 王骏成.我国算力产业发展面临三大挑战[J].中国信息界,2022(06):50-53.
- [6] 雷波,刘增义,王旭亮等.基于云、网、边融合的边缘计算新方案:算力网络[J].电信科学,2019,35(09):44-51.
- [7] 李少鹤,李泰新,周旭.算力网络:以网络为中心的融合资源供给[J].中兴通讯技术,2021,27(03):29-34.

5G 无线通信中的新型索引调制系统研究

范世文

(爱瑞通信系统(厦门)有限公司)

摘要:空间调制(Spatial modulation, SM)^[1-6]作为一种新型多输入多输出(multiple input multiple output, MIMO)技术,具有低射频开销、硬件实现简单和传输性能可靠等特点,因此成为第五代(Fifth generation, 5G)移动通信的物理层技术提案。为了兼容5G的多载波调制,空间调制的概念可以进一步扩展到时域和频域,形成新的空时频联合传输技术:广义空时频索引调制(Generalized Space-Time-Frequency Index Modulation, GSTFIM)^[7],其具有抗多径效应强、低峰均比和载波间干扰较小的优点。正交时频空(Orthogonal Time Frequency Space, OTFS)^[8-12]是一种针对5G高速移动场景多维调制方案,通过辛有限傅里叶逆变换(inverse Symplectic Finite Fourier Transform, ISFFT)和辛有限傅里叶变换(Symplectic Finite Fourier Transform, SFFT)将时频选择信道转换成一个可分离、静态和正交的时延-多普勒域信道,其具有所有的星座符号都经历相同的信道增益的特点,从而更适用5G的高速移动场景。

基于上述技术背景,我们将GSTFIM和OTFS结合,形成新型索引调制系统,并对系统的信号检测问题进行了探讨。具体来说,新系统GSTFIM-OTFS具有更好的系统性能,可以有效解决5G高速移动场景中的传输可靠性问题。

关键词:空间调制,正交时频空,检测,索引调制

引言

为了适应5G无线通信系统对于高速移动环境和传输性能的要求,本文将广义空时频索引调制(Generalized Space-Time-Frequency Index Modulation, GSTFIM)^{[7][22-25]}和正交时频空(Orthogonal Time Frequency Space, OTFS)^[8-12]结合,形成新型的广义空时频索引调制正交时频空(GSTFIM-OTFS)系统。GSTFIM-OTFS系统不仅具有索引调制载波间干扰小、低峰均比和抗多径能力强的优点,而且继承了OTFS系统对抗多普勒扩展的能力。

文献[7]提出了空间调制下的空频联合传输技术:广义空时频索引调制(Generalized Space-Time-Frequency Index Modulation, GSTFIM),来适应5G系统的多载波传输体制。作为一种多维调制系统,其比特信息不仅包含在星座符号上,还存在于空域、时域、频域的索引位置上,

从而增加了频谱效率。空时频索引调制和广义空时频索引调制的区别在于:空时频索引调制系统激活的天线、时隙和子载波数目为一,而广义空时频索引调制系统激活的天线、时隙和子载波数目高于一个,激活的天线、时隙和子载波数目可以根据通信场景的要求进行灵活调整。

另一方面,OTFS是一种能够对抗多普勒扩展的5G技术,它是在时延多普勒域实现星座符号复用的二维调制方案。文献[13]对OTFS中的PAPR问题进行了研究,文献[14-15]对OTFS中的信号检测问题进行了相关研究。文献[16]介绍了一种用于OTFS通信的系统和方法,该方法包括接收多个符号,并通过将多个符号中的每一个在时间和频率上展开,将包含多个符号的 $N \times M$ 阵列编码成调制符号的二维阵列,然后,使用 M 个频带内包含的 M 个相互正交的波形来发送调制符号的二维阵列。本方法兼容了OFDM和OTFS,解决了OTFS技术中的波形产生问题。专利[17]提出了利用OTFS导频符

号波形突发自动生成详细的信道状态 2D 模型的方法。这种 2D 信道状态可以用来优化数据传输。对于无线数据信道，在这个过程中，利用极化和多天线可以产生更详细的信道状态的二维模型。一旦 2D 信道状态已知，该系统就利用信道缺陷来提高数据传输速率，从而将不完善的数据信道从可靠性变为优势。这些方法可用于改进多种媒体中的传统数据传输模式，尤其适用于使用非传统 OTFS 数据传输方法产生新型的高容量无线通信。专利[18] 利用 OTFS 技术实现了固定无线接入系统。此专利为 OTFS 在 5G 系统中的应用提供了一定的参考性。OTFS 通过辛有限傅里叶逆变换 (inverse Symplectic Finite Fourier Transform, ISFFT) 和辛有限傅里叶变换 (Symplectic Finite Fourier Transform, SFFT)，将时频选择信道转换为一个可分离、静态和正交的信

道，全部的星座符号都经历相同的信道增益，具有稳定性，因此与 OFDM 相比，OTFS 更适用于 5G 的高速移动场景。

为了进一步提升系统性能，以便进一步适应未来 5G 对于高可靠性的要求，本文通过 GSTFIM 和 OTFS 的结合，形成新型的广义空时频索引调制正交时频空 (GSTFIM-OTFS) 系统来获得更好的性能。本文对于 GSTFIM-OTFS 系统的信号检测进行了研究。

本章结构如下:1.2 节介绍了在 GSTFIM-OTFS 的系统模型; 1.3 节描述了线性 MMSE 均衡算法; 1.4 节给出计算机仿真结果; 1.5 节总结本文。

1.广义空时频索引调制正交时频空系统模型

本小节介绍了广义空时频索引调制正交时频空的模型，具体内容如下。

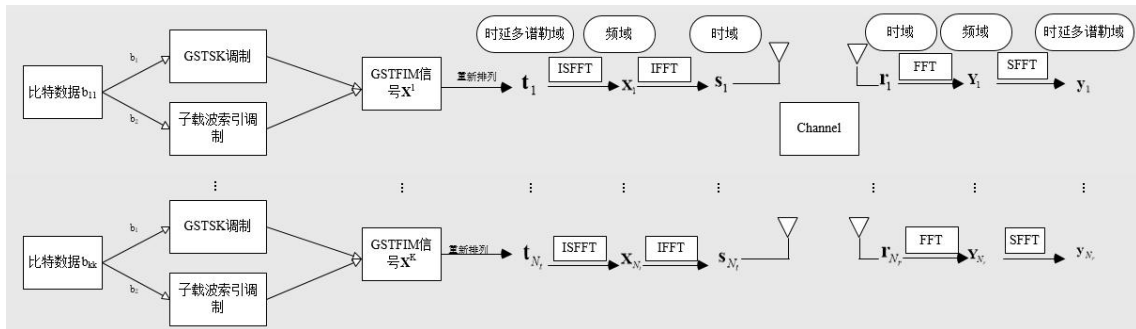


图 1 GSTFIM-OTFS 系统模型图

如上图所示，考虑参数为： N_t 根发射天线、 N_r 根接收天线、 M -PSK/QAM 调制、 T 个符号周期和 N 个子载波的 GSTFIM 系统。首先把 N 个子载波分成 K 个块，每个块有 N_b 个子载波。对于第 k 块子载波，信息比特 b_{kk} 分成两部分，其中 $b_{kk} = b_1 + b_2$ 和 $k = 1, \dots, K$ 。第一部分 b_1 个比特用来调制 N_a 个 GSTSK 符号^[19]。根据文献[19]，GSTSK 符号是通过激活 Q 个弥散矩阵 $A_1^k, A_2^k, \dots, A_Q^k$ 中的 L 个和映射 L 个 N -PSK/QAM 个星座符号来产生的，其中弥散矩阵的维度是 $N_t \times T$ 。这样映射一个 GSTSK 符号所需要的比特数为 $\lceil \log_2 C_Q^L \rceil + L \log_2 M$ ，因此 $b_1 = N_a \lceil \log_2 C_Q^L \rceil + N_a L \log_2 M$ 。第 i 个 GSTSK 符号可以表示为：

$$S_i^k = \sum_{l=1}^L s_l A_l^k \in \mathbb{C}^{N_t \times T} \quad (1-1)$$

其中 $i = 1, \dots, N_a$ 。其中 Q 和 L 的选择满足条件：

$$\lceil \log_2 C_Q^L \rceil = L \left(\lceil \log_2 C_{N_t}^{N_t'} \rceil + \lceil \log_2 C_T^{T'} \rceil \right)。N_t' 和 T' 分别表示激活的天线数和时隙数。$$

另一方面， b_{kk} 中的另外 b_2 个比特用来选择 N_b 个子载波中的 N_a 个激活子载波，定义这 N_a 个激活子载波索引为 $I^k = \{I_1, I_2, \dots, I_{N_a}\}$ ，此时可知 $b_2 = \lceil \log_2 C_{N_b}^{N_a} \rceil$ 。将 N_a 个 GSTSK 符号 $[S_1^k, S_2^k, \dots, S_{N_a}^k]^T$ 映射到 N_b 个子载波上，可得第 k 块子载波的发射信号为：

$$X^k = [X_1^k, X_2^k, \dots, X_{N_b}^k]^T \quad (1-2)$$

X^k 中的 N_a 个元素是 $[S_1^k, S_2^k, \dots, S_{N_a}^k]^T$ ，对应的索引是 I^k 。合并 K 块子载波的发射信号，得到时延多普勒域的 GSTFIM-OTFS 发射信号：

$$t(t, n) = [X^1, X^2, \dots, X^K] \in \mathbb{C}^{N N_t \times T} \quad (1-3)$$

为了简便表述,重新定义 GSTFIM-OTFS 信号为 $\mathbf{t}(t, n) = [t_1(t, n), t_2(t, n), \dots, t_{N_t}(t, n)]^T$ 。为了得到每根发射天线上的信号,将 GSTFIM-OTFS 符号 $\mathbf{t}(n, m)$ 的天线维度择出来,按横轴是子载波和纵轴是时隙的维度重新表示为:

$$\mathbf{t}_i = \begin{bmatrix} t_i(1,1) & t_i(1,2) & \dots & t_i(1,N) \\ t_i(2,1) & t_i(2,2) & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ t_i(T,1) & t_i(T,2) & \dots & t_i(T,N) \end{bmatrix} \quad (1-4)$$

其中 \mathbf{t}_i 表示第 i 根发射天线上的所有信号。在 GSTFIM-OTFS 系统的发射端,对时延多普勒域发射符号 \mathbf{t}_i 进行 ISFFT 操作。具体来说, \mathbf{t}_i 的第 t 个时隙第 n 个子载波上的 ISFFT 操作可以表示为:

$$\begin{aligned} X_i[t, n] &= \text{ISFFT}(\mathbf{t}_i[t, n]) \\ &= \frac{1}{MN} \sum_{l,k} t_i[k, l] e^{j2\pi[\frac{t(k-1)}{T} - \frac{n(l-1)}{N}]} \end{aligned} \quad (1-5)$$

其中 $l \in \{1, 2, \dots, N\}$ 和 $k \in \{1, 2, \dots, T\}$ 。因此频域的发射信号 \mathbf{X}_i 可以表示为:

$$\mathbf{X}_i = \begin{bmatrix} X_i(1,1) & X_i(1,2) & \dots & X_i(1,N) \\ X_i(2,1) & X_i(2,2) & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ X_i(T,1) & X_i(T,2) & \dots & X_i(T,N) \end{bmatrix} \quad (1-6)$$

其中 $i \in \{1, 2, \dots, N_t\}$ 。第 t 个时隙第 n 个子载波上的时域发射信号 $s_i(t, n)$ 可以通过逆傅利叶变换 (inverse fast fourier transform, IFFT) 表示为:

$$s_i(t, n) = \frac{1}{\sqrt{N}} \sum_{l=1}^N X_i(t, l) e^{j2\pi(n-1)(l-1)/N} \quad (1-7)$$

这样第 i 根发射天线上的时域信号 \mathbf{s}_i 可以表示为:

$$\mathbf{s}_i = \begin{bmatrix} s_i(1,1) & s_i(1,2) & \dots & s_i(1,N) \\ s_i(2,1) & s_i(2,2) & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ s_i(T,1) & s_i(T,2) & \dots & s_i(T,N) \end{bmatrix} \quad (1-8)$$

最终在发射端,对时域信号 \mathbf{s}_i 加 CP 然后通过发射天线传输。

假设在接收端,信道状态信息 (channel state information, CSI) 已知。为了便于表述,假设 $\mathbf{s}_i(t) = [s_i(t, 1) \ s_i(t, 2) \ \dots \ s_i(t, N)]^T$ 。经过去 CP 操作,第 t 个时隙第 j 根接收天线上的信号 $\mathbf{r}_j(t)$ 可以表示为:

$$\mathbf{r}_j(t) = \sum_{i=1}^{N_t} \mathbf{h}_{j,i}^t \mathbf{s}_i(t) + \mathbf{z}_j(t) \quad (1-9)$$

其中 $\mathbf{h}_{j,i}^t$ 表示第 j 根接收天线和第 i 根发射天线间的时域信道冲激响应,表示为

$$\mathbf{h}_{j,i}^t = \begin{bmatrix} h_{j,i}^t(1) & h_{j,i}^t(N) & \dots & h_{j,i}^t(2) \\ h_{j,i}^t(2) & h_{j,i}^t(1) & \dots & h_{j,i}^t(3) \\ \vdots & \vdots & \dots & \vdots \\ h_{j,i}^t(N-1) & h_{j,i}^t(N-2) & \dots & h_{j,i}^t(N) \\ h_{j,i}^t(N) & h_{j,i}^t(N-1) & \dots & h_{j,i}^t(1) \end{bmatrix} \quad (1-10)$$

另外 $\mathbf{z}_j(t)$ 表示噪声向量,它的元素服从均值为 0,方差为 σ^2 的复高斯分布。假设 $\mathbf{r}_j(t) = [r_j(t, 1) \ r_j(t, 2) \ \dots \ r_j(t, N)]^T$,那么第 j 根接收天线上的接收信号 \mathbf{r}_j 为:

$$\mathbf{r}_j = \begin{bmatrix} r_j(1,1) & r_j(1,2) & \dots & r_j(1,N) \\ r_j(2,1) & r_j(2,2) & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ r_j(T,1) & r_j(T,2) & \dots & r_j(T,N) \end{bmatrix} \quad (1-11)$$

其中 $j \in \{1, 2, \dots, N_r\}$ 。时域模型(1-9)的频域模型可通过对其进行 FFT 变换得到:

$$\mathbf{Y}_j(t) = \sum_{i=1}^{N_t} \mathbf{H}_{j,i}^t \mathbf{X}_i(t) + \mathbf{Z}_j(t) \quad (1-12)$$

其中, $\mathbf{H}_{j,i}^t = \mathbf{F} \mathbf{h}_{j,i}^t \mathbf{F}^H$ 表示频域冲激响应矩阵, \mathbf{F} 表示归一化的 FFT 矩阵。从 OFDM 结构可知,时域卷积相当于频域相乘,因此频域矩阵 $\mathbf{H}_{j,i}^t$ 是一个对角矩阵 $\mathbf{H}_{j,i}^t = \text{diag}[H_{j,i}^t(1) \ H_{j,i}^t(2) \ \dots \ H_{j,i}^t(N)]$,其中 $\text{diag}(\mathbf{A})$ 表示对角矩阵,它的元素是 \mathbf{A} 。 $\mathbf{X}_i(t) = \mathbf{F} \mathbf{s}_i(t)$ 是时域信号 $\mathbf{s}_i(t)$ 的频域形式。基于等式(1-7),可以得到 $\mathbf{s}_i(t, n)$ 的频域形式 $\mathbf{X}_i(t, n)$,因此 $\mathbf{X}_i(t)$ 可以描述为:

$$\mathbf{X}_i(t) = [X_i(t, 1) \ X_i(t, 2) \ \dots \ X_i(t, N)]^T \quad (1-13)$$

$\mathbf{Y}_j(t) = \mathbf{F} \mathbf{r}_j(t)$ 和 $\mathbf{Z}_j(t) = \mathbf{F} \mathbf{z}_j(t)$ 分别表示时域接收信号 $\mathbf{r}_j(t)$ 和时域噪声信号 $\mathbf{z}_j(t)$ 的频域形式,具体可表示为:

$$\mathbf{Y}_j(t) = [Y_j(t, 1) \ Y_j(t, 2) \ \dots \ Y_j(t, N)]^T \quad (1-14)$$

$$\mathbf{Z}_j(t) = [Z_j(t, 1) \ Z_j(t, 2) \ \dots \ Z_j(t, N)]^T \quad (1-15)$$

与等式(1-4)类似,按横轴是子载波和纵轴是时隙的维度表示,第 j 根接收天线上的频域接收信号 \mathbf{Y}_j 可以表示为:

$$\mathbf{Y}_j = \begin{bmatrix} Y_j(1,1) & Y_j(1,2) & \dots & Y_j(1,N) \\ Y_j(2,1) & Y_j(2,2) & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ Y_j(T,1) & Y_j(T,2) & \dots & Y_j(T,N) \end{bmatrix}$$

$$(1-16)$$

接着，对频域接收信号 \mathbf{Y}_j 进行 SFFT 操作，得到时延多普勒域接收信号 $y_j(k, l)$ ：

$$y_j(k, l) = SFFT(\mathbf{Y}_j, [k, l]) = \sum_{t=1}^T \sum_{n=1}^N Y_j(t, n) e^{-j2\pi[\frac{k(t-1)}{T} - \frac{l(n-1)}{N}]}, \quad (1-17)$$

其中 $l = \{1, 2, \dots, N\}$ 和 $k = \{1, 2, \dots, T\}$ 。

最后，第 j 根接收天线的时延多普勒域接收信号 \mathbf{y}_j 可以重新表述为：

$$\mathbf{y}_j = \begin{bmatrix} y_j(1,1) & y_j(1,2) & \dots & y_j(1,N) \\ y_j(2,1) & y_j(2,2) & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ y_j(T,1) & y_j(T,2) & \dots & y_j(T,N) \end{bmatrix} \quad (1-18)$$

2. 频域均衡算法

这里介绍频域均衡 (frequency domain equalizer, FDE) 算法。从频域模型(1-12) 可知，等式右边是标量乘法，因此第 n 个时隙第 m 个子载波上的频域模型可以表示为：

$$\underbrace{\begin{bmatrix} Y_1(t, n) \\ Y_2(t, n) \\ \vdots \\ Y_{N_r}(t, m) \end{bmatrix}}_{\mathbf{Y}(t, n)} = \underbrace{\begin{bmatrix} H_{1,1}^t(n) & H_{1,2}^t(n) & \dots & H_{1,N_t}^t(n) \\ H_{2,1}^t(n) & H_{1,1}^t(n) & \ddots & H_{1,N_t}^t(n) \\ \vdots & \ddots & \ddots & \vdots \\ H_{N_r,1}^t(n) & H_{N_r,1}^t(n) & \dots & H_{N_r,N_t}^t(n) \end{bmatrix}}_{\mathbf{H}^t(n)} \underbrace{\begin{bmatrix} X_1(t, n) \\ X_2(t, n) \\ \vdots \\ X_{N_t}(t, n) \end{bmatrix}}_{\mathbf{X}(t, n)} + \underbrace{\begin{bmatrix} Z_1(t, n) \\ Z_2(t, n) \\ \vdots \\ Z_{N_r}(t, n) \end{bmatrix}}_{\mathbf{Z}(t, n)}$$

基于 MMSE 准则[20]，第 t 个时隙第 n 个子载波上的线性 FDE 滤波系数计算如下：

$$\mathbf{W}_{MMSE}(t, n) = [\mathbf{H}^t(n)]^H (\mathbf{H}^t(n) [\mathbf{H}^t(n)]^H + \sigma^2 I_{N_r})^{-1} \quad (1-20)$$

因此 FDE 检测算法的输出结果为：

$$\hat{\mathbf{X}}(t, n) = \mathbf{W}_{MMSE}(t, n) \mathbf{Y}(t, n) \quad (1-21)$$

遍历所有的时隙和子载波，得到所有的频域估计信号 $\hat{\mathbf{X}}(t, n)$ ；然后通过式(1-17)的 SFFT 操作，将估计

信号 $\hat{\mathbf{X}}(t, n)$ 从频域转换到时延多普勒域 $\hat{\mathbf{t}}(t, n) = [\hat{t}_1(t, n), \hat{t}_2(t, n), \dots, \hat{t}_{N_t}(t, n)]^T$ ；最后对时延多普勒

域估计信号 $\hat{\mathbf{t}}(t, n)$ 按照 N_b 个子载波为一个子载波块划分为 K 块： $\hat{\mathbf{t}}(t, n) = [\hat{\mathbf{X}}^1, \hat{\mathbf{X}}^2, \dots, \hat{\mathbf{X}}^K] \in \mathbb{C}^{N N_t \times T}$ 。其中 $\hat{\mathbf{X}}^k$ 表示 N_b 个子载波上的时延多普勒域信号，它是由 N_a 个 GSTSK 符号 $[\mathbf{S}_1^k, \mathbf{S}_2^k, \dots, \mathbf{S}_{N_a}^k]^T$ 和 $(N_b - N_a)$ 个零矩阵组成的。通过遍历所有可能的激活子载波和星座符号的组合，选出其中范数最小的组合，可以认为是估计结果，进而可以转化为原始信息比特的估计，从而完成了解调。

上述 FDE 检测算法，其实现复杂度较低，适应于对时延敏感的 5G 场景。

举例说明 GSTFIM 的 N_b 个子载波上的时延多普勒域信号。考虑系数参数为： $N_t = 2, T = 4, L = 1, N_b = 2, N_a = 1, Q = 8, N'_t = 1$ 和 $T' = 1$ 。弥散矩阵 $\mathbf{A}_1^k, \mathbf{A}_2^k, \mathbf{A}_3^k, \mathbf{A}_4^k, \mathbf{A}_5^k, \mathbf{A}_6^k, \mathbf{A}_7^k, \mathbf{A}_8^k$ 表示为：

$$\begin{aligned} \mathbf{A}_1^k &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{A}_2^k = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{A}_3^k = \\ & \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{A}_4^k = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ \mathbf{A}_5^k &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \mathbf{A}_6^k = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \mathbf{A}_7^k = \\ & \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \mathbf{A}_8^k = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned} \quad (1-22)$$

那么 GSTSK 符号是 $\mathbf{S}_1^k = \begin{bmatrix} s_l & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ ，将 GSTSK 符号映射到 $N_b = 2$ 个子载波上去，得到第 k 块子载波的 GSTFIM 符号 \mathbf{X}^k ：

$$\mathbf{X}^k = \left[\begin{bmatrix} s_l & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right]^T \quad (1-23)$$

3. 仿真结果

本小节针对所提出的新型 GSTFIM-OTFS 系统，比较了它与传统的 GSTFIM 的性能。本文采用标准的采用扩展车辆 (Extended Vehicular A, EVA) 无线信道^[21]。系统的载波频率为 $4e9\text{Hz}$ ，采样周期 $T_s = 1/(15000M)$ ，GSTFIM-OTFS 的符号周期为 $T = 1.25NTT_s$ 。

为了简洁地表示各个算法，第 1.3 节介绍的频域均衡算法标注为“GSTFIM-OTFS-MMSE”，传

统的 GSTFIM 中的频域均衡算法标注为“GSTFIM-OFDM-MMSE”。

图 1-2 给出了传统的 GSTFIM-OFDM 系统和所提出的 GSTFIM-OTFS 在 EVA 信道下，基于 2 发 2 收配置的性能对比图。

图 1-2 的仿真参数具体如下：发射天线 $N_t = 2$ ，接收天线 $N_r = 2$ ，散矩阵个数 $Q = 8$ ，符号周期 $T = 4$ ，子载波间隔 15kHz，调制方式 QPSK，接收天线 $N_r = 2$ ，子载波 $N = 128$ ，子载波块 $K = 64$ ，每个块有 $N_b = 2$ 个子载波，每个块激活 $N_a = 1$ 个子载波，GSTSK 符号的弥散矩阵激活个数 $L = 1$ ，CP 长度是 32，车速设置为 120km/h。

GSTFIM-OFDM 和 GSTFIM-OTFS 系统均采用第 1.3 节的传统频域均衡算法。从图 1-2 的仿真可知，与 GSTFIM-OFDM 相比，所提出的 GSTFIM-OTFS 系统在 $BER=10^{-3}$ 时能够获得约 7 dB 的性能增益。

图 1-3 仿真了 GSTFIM-OTFS 在 EVA 信道下，基于 2 发 2 收 2 符号周期配置下，GSTFIM-OFDM 和 GSTFIM-OTFS 系统的性能。图 1-3 的仿真参数具体如下：发射天线 $N_t = 2$ ，接收天线 $N_r = 2$ ，散矩阵个数 $Q = 4$ ，符号周期 $T = 2$ ，子载波间隔 15kHz，调制方式 QPSK，接收天线 $N_r = 2$ ，子载波 $N = 128$ ，子载波块 $K = 64$ ，每个块有 $N_b = 2$ 个子载波，每个块激活 $N_a = 1$ 个子载波，GSTSK 符号的弥散矩阵激活个数 $L = 1$ ，CP 长度是 32，车速设置为 120km/h。

图 1-3 可以得到与图 1-2 类似的结论，所提出的 GSTFIM-OTFS 系统在 $BER=10^{-3}$ 时能够获得约 8 dB 的性能增益。

从上述两个系统仿真可以得到以下结论：对于 EVA 和随机独立多径信道，在采用相同检测算法的条件下，所提出的 GSTFIM-OTFS 系统可以凭借 OTFS 的高分集增益和对抗多普勒扩展的能力，获得更好的性能。

图 1-4 仿真了 GSTFIM-OTFS 在 EVA 信道下，基于 2 发 2 收 4 符号周期配置，不同车速下的性能图。可以看出，随着车速的增加，OTFS 系统的性能随之提高，这是由于更高的车速提供了更多的多普勒增益。

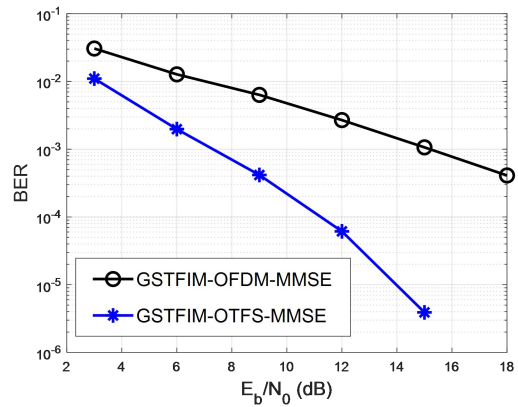


图 1-1 4 符号周期 EVA 信道下两种系统传统 MMSE 检测性能对比图

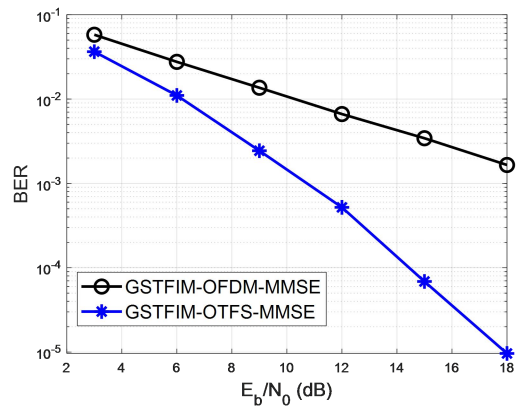


图 1-2 2 符号周期 EVA 信道下两种系统传统 MMSE 检测性能对比图

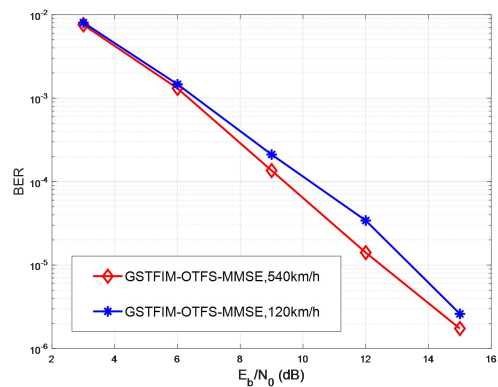


图 1-3 GSTFIM-OTFS 系统不同车速检测性能对比图

4. 本文小结

为了适应 5G 高速移动场景, 本文将索引调制和正交时频空技术相结合, 提出了广义空时频索引调制正交时频空系统。Matlab 仿真结果表明, 本文所提出的新型索引调制系统, 与传统的索引调制系统相比, 更适用于 5G 高速移动场景, 具有更好的传输性能。

参考文献

[1] Y. Xiao, L. Xiao, L. Dan, and X. Lei. Spatial modulation for 5G MIMO Communications[C]. In Proc. 19th Int. Conf. Digit. Signal Process., 2014, 847 - 851.

[2] M. Di Renzo, H. Haas, A. Ghayeb, et al. Spatial modulation for generalized MIMO: challenges, opportunities and implementation[J]. Proc. IEEE, 2014,

102(1): 56-103.

[3] P. Yang, M. Di Renzo, Y. Xiao, et al. Design guidelines for spatial modulation[J]. IEEE Commun. Surveys. Tuts., 2015, 17(1): 6-26.

[4] Y. A. Chau, S. H. Yu. Space modulation on wireless fading channels[C]. IEEE Veh. Technol. Conf. (VTC)-Fall, Atlantic City, 2001, 1668-1671.

[5] M. Di Renzo, H. Haas. Bit error probability of SM-MIMO over generalized fading channels[J]. IEEE Trans. Veh. Technol., 2012, 61(3): 1124-1144.

[6] S. Guo, H. Zhang, S. Jin, and P. Zhang. Spatial modulation via 3-D mapping[J]. IEEE Commun. Lett., 2016, 20(6): 1096-1099.

[7] M. I. Kadir. Generalized space-time-frequency index modulation[J]. IEEE Commun. Lett., 2019, 23(2): 250 - 253.

单相浸没式液冷 BBU 机柜的优化和模拟分析

张亚洲 沈 斌

(福州六尺智算科技有限公司)

摘要: 针对在用项目的单相浸没式液冷室内基带处理单元 (Building Baseband Unit, BBU) 机柜, 为了提升其热可靠性, 降低电源使用效率 (Power Usage Effectiveness, PUE), 进行了测试和数值模拟研究。通过冷却液和空气 2 种不同媒介在 BBU 内部的流场和温度场分析, 发现基于空气为冷却介质进行热设计的翅片散热器, 在液冷设备里并不适用; 基于某国产冷却液实际特性的 BBU 设备的仿真, 提出了去除原翅片散热器来优化 BBU 内部流道和设置不同盲板优化方案, 以确保冷液进入机柜后, 将冷却液的动量和流量, 优先给予高发热元件, 增强换热。

关键词: 单相浸没; 液冷; 室内基带处理单元 (BBU); 机柜

引言

目前 5G 网络建设采用集中无线接入网 (Centralized Radio Access Network, C-RAN) 模式, 集中在室内布置^[1]。通常, 4G 室内基带处理单元 (Building Baseband Unit, BBU) 的功耗为 150 ~ 300W, 而 5G BBU 达到 500 ~ 1200W^[2]。此外, 在同规格覆盖情况下, 5G 中频段基站数量将是 4G 的 1.5 倍左右。5G 风冷 BBU 单机柜可以布置 10 台左右设备, 但是超过 6kW 的情况下风冷散热已经达到瓶颈, 无法满配^[3]。为了节约场地和能源成本, 液冷方案在 BBU 的应用已经开始有代替风冷的趋势^[4]。5G BBU 采用液冷技术, 冷却介质直接或者间接与 BBU 内部接触, 利用冷却液相对于空气较高的热容和导热系数, 使得机柜可容纳 10 台以上 BBU, 并且实现 10kW 以上冷却能力, 这意味着即使 10 台 BBU 满载也无局部温度超标的问题。传统风冷 BBU 机柜即使有 20U 以上空位, 也只能配置 5~6 台 BBU 设备, 采用液冷方案可节省机柜数量, 从而使得运营商可以在有限的空间部署更多数量的设备^[5]。除此之外, 对于液冷的数据中心, 单相 (两相) 浸没式、喷淋式和冷板式都能获得比空气冷却优异的多的电源使用效率 (Power Usage Effectiveness, PUE) 值, 其中浸没式的理论 PUE 可以达到 1.1 以下, 满足政策层面上对于机房能耗的要求^[6]。

以一个在役的采用单相浸没式技术的液冷 BBU 机柜为研究对象, 通过现场测量和调取历史运行数据的方法, 建立了机柜级和设备级 2 种模型。对于机柜级模型, 研究关注于冷却液流入或流出 BBU 的流量和机柜内宏观空间的速度场和温度场。基于 BBU 设备其内部结构复杂的特点, 数亿级的网格数量耗费大量的计算机资源是不合理的, 而将 BBU 内部板卡和散热器组件简化为多孔介质模型, 分析了原机柜内部的流场, 并针对其存在的局部热点问题提出 2 种优化方案, 增设了盲板导流结构, 使整个机柜内部的冷却液的无组织流动变为有组织流动, 使得更多的冷却液扫掠板发热元器件表面^[7]。针对设备级模型, 建立了去除风扇模块和其他支撑件、保留了散热翅片模型的 BBU 模型, 分别使用空气和某国产冷却液 2 种不同工质, 对比分析其散热翅片热环境, 研究了 2 种冷却介质的对散热器设计的影响, 并提出了优化意见。

1. 问题概述

1.1 机柜概况

本文选取浙江地区某 5G 液冷 BBU 机柜为研究对象, 其 BBU 机柜平面布局如图 1 所示。机柜采用某国产单相冷却液, 使用浸没式冷却方案。冷却液由冷却液分配单元

(CoolantDistributionUnit, CDU)通过开孔管进入 Tank 底部一侧,进入其内部,冷却液流进 BBU 入口,将发热模块的热量带走,排出 BBU 出口。各 BBU 出口汇集经过机柜后部上液面附近的冷却液出口进入 CDU 进行换热。Tank 的内部尺寸为 $950\text{mm} \times 500\text{mm} \times 500\text{mm}$ 。一台直流配电单元 (DirectionCurrentDistributionUnit, DCDU) 和 10 台 BBU 并排放置在 Tank 内部, BBU 上表面浸入液面 3cm 以下。BBU 为华为某型号,采用左进右出气流组织方案。原设计采用风冷,进行液冷改造时去除了风扇模块,单台 BBU 的外形尺寸为 $441\text{mm} \times 89\text{mm} \times 350\text{mm}$ 。

此液冷机柜设计负荷 10kW , 额定进液温度 40°C , 额定出液温度 45°C , 配置 2 台液泵, 额定流量 5t/h , 变频控制。每一台 BBU 有 8 张板卡, 根据厂家电气规格书和历史测量数据, 电功率约为 1kW 。

1.2 问题描述和分析

项目于 2022 年 3 月投用, 瞬时最大 PUE 保持在 1.15 以下, 整体运行良好。据现场运维反馈, BBU 编号 1-4 在最热月份会出现超温报警, 开启了增强 BBU 内部扰动的射流模块后, 解决了这个问题。为了进一步降低 PUE, 减少射流模块的开启频率, 达到 1.1 以下的全年 PUE, 需要对原有液冷机柜进行流体组织分析。

一般而言, 在空调房间中, 风口的位置和射流参数是最主要的影响房间流体组织的因素^[8]。在液冷机柜中, 机柜的进液口和出液口的位置也是影响整个液柜速度场的最主要因素。此外, 进液机构采用孔管和孔板组合结构, 冷却液的物性参数也是影响其内部流动的重要因素, BBU 设备本身的流通通道也是换热的重要组成部分^[9]。通常, 华为 BBU 采用左进右出的流体组织方案, 而最优的方案, 是冷却液从 BBU 进口, 以垂直于进口截面的角度、充分发展流的流态进入其内部, 扫掠 BBU 发热部件和散热器将热量带走。但是原型机柜的出口位于液柜上部侧边一角位置, 导致整个汇流集中至液柜一角, 远离回流口的 BBU 获得相对少的冷却液流量, 这是影响 BBU 出现温度热点的最主要因素。如果将液柜出液口改为满布液

柜上部区域的长条形出液口, 会较大程度改善流量不均现象, 如图 1 所示。但是, BBU 设备插板和翅片散热器机构复杂, 冷却液又是高粘性流体, 在运行发热状况下, 其物理性质受到温度的影响产生密度差受到浮力影响, 属于强迫对流方式, 因此更精确的流量分布需要运用 CFD 技术进行热仿真分析。

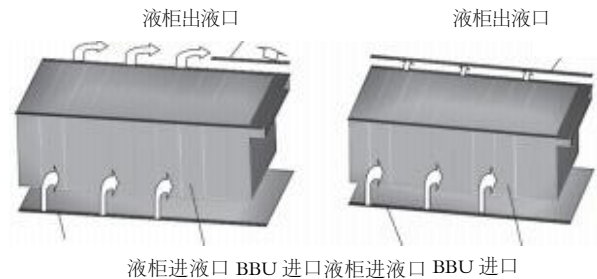


图 1 新旧液柜流动示意

2 模型与仿真

2.1 建立机柜级模型

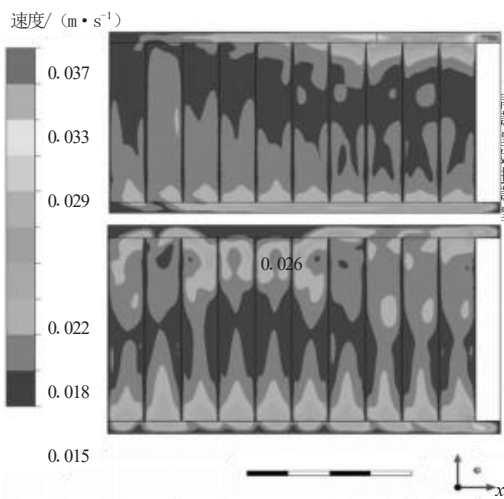
利用 ANSYSFLUENT 进行求解, 使用 RNGk- ϵ 湍流模型、SIMPLE 分离算法, 考虑浮力开启 FullBuoyancyEffect, 压力差值使用 BodyForceWeighted。冷却液进口设置质量进口 1.55kg/s , 冷却液进口设置压力出口。采用某国产冷却液, 其相关物性参数见表 1。BBU 设置进口和出口的内部面, BBU 内部结构设置 PorousZone, 设置内部热源, 并对网格进行独立性验证。

稳态情况下, 使用软件计算得到通过 BBU 的冷却液的流量, 得到通过 BBU 进口的流量为 0.53270682kg/s , 通过 BBU 出口的流量为 0.53262835kg/s , 而初始边界条件冷却液质量为 1.55kg/s , 大约有 $1/3$ 的冷却液进入了 BBU 内部, 只有这一部分为实际有效利用的液体。而原始设计温差 5°C , 流量变小之后为了排出相同负荷的热量, 这部分有效利用的流体温差可能达到 15°C , 虽然还在液体有效适用范围, 但是在极端高温情况下, 冷却水温度升高就有可能使得冷却液最高温度升高, 出现局部热点, 从而使 BBU 内部板卡传感器高温停机报警。而由于出液管道的温度传感器设置在 CDU 进口位置, 此处测得的温度是冷却液在

表 1 不同温度下冷却液的物理性质

物理性质	25℃	35℃	40℃	45℃
密度/(g/cm ³)	0.962	0.952	0.947	0.943
比热容/[J/(g·K)]	1.88	1.75	1.65	1.55
导热系数/[W/(m·K)]	0.161	0.158	0.155	0.15
粘度/(mm ² /s)	46.7	44	42	38

出口混合之后的温度,就有可能在 CDU 未能识别高温点的情况下,未能进行有效的冷源端调节情况下出现液体局部高温,从而导致 BBU 停机。截取 BBU 上部截面($Z=0.4\text{m}$),可以得到采用长条出口的改进模型有更好的速度分布,整体更加均匀,如图 2 所示。根据仿真结果的速度矢量分析,位于 BBU 挂耳和 Tank 支撑架之间,由于采用孔板结构,此处发生了液体的短路,液体没有进入 BBU 内部就从旁路离开进入液柜出口。同风冷数据中心类似,如果发生了冷热掺混,必然会导致冷量浪费,泵功增多,相当一部分的流量被浪费了。

图 2 $Z=0.4\text{m}$ 速度分布云图

2.2 机柜级优化模型

为了将挂耳和 Tank 支撑架之间的空位封堵,减少冷量掺混,设计 2 种盲板结构,对比方案对比分析,如图 3 所示。改进模型 1 在 BBU 挂耳与 Tank 之间设置盲板 A,在液柜进液口和 BBU 底部之间设置了盲板 B,这样强迫所有冷却液体进入 BBU 内部,这样就解决了冷却液短路的问题。改进模型 2 同改进模型 1 的盲板位置相同,但是改变了进液口位置,将原有位于液柜底部的进液口改到液柜侧壁,期望用垂直于 BBU 进口截面的液流充分扫掠 BBU 内部各发热元件,获得比改进模型 1 更加充分发展的入口流。但是观察速度分布并没有较大改善,主要是由于改进模型 2 仅改变了入口的位置,出口位置仍然是液柜上表面一侧,基于浮力效应,部分液体加热之后进入高温区域,上浮到 BBU 顶部。从图 3 温度分布来看,改进模型 2 的高温部分更少,符合预期。此外,BBU 实际上表面也有部分空隙可以供部分高温液体进入液柜上部液面,这样 2 种方案的区别可能被均衡了。2 种方案都能改善整体液流流动和热环境,但是基于成本来讲,改进模型 1 更加简单。再者,改进模型 1 的盲板 B 可以采用长方体填充块的形式填满 BBU 底板与液柜底板之间的空间,这样能够满足盲板功能的同时,减少液柜的液体使用量。此国产冷却液体价格远低于进口氟化液,还能节约 2000 元左右的用量。

2.3 设备级模型

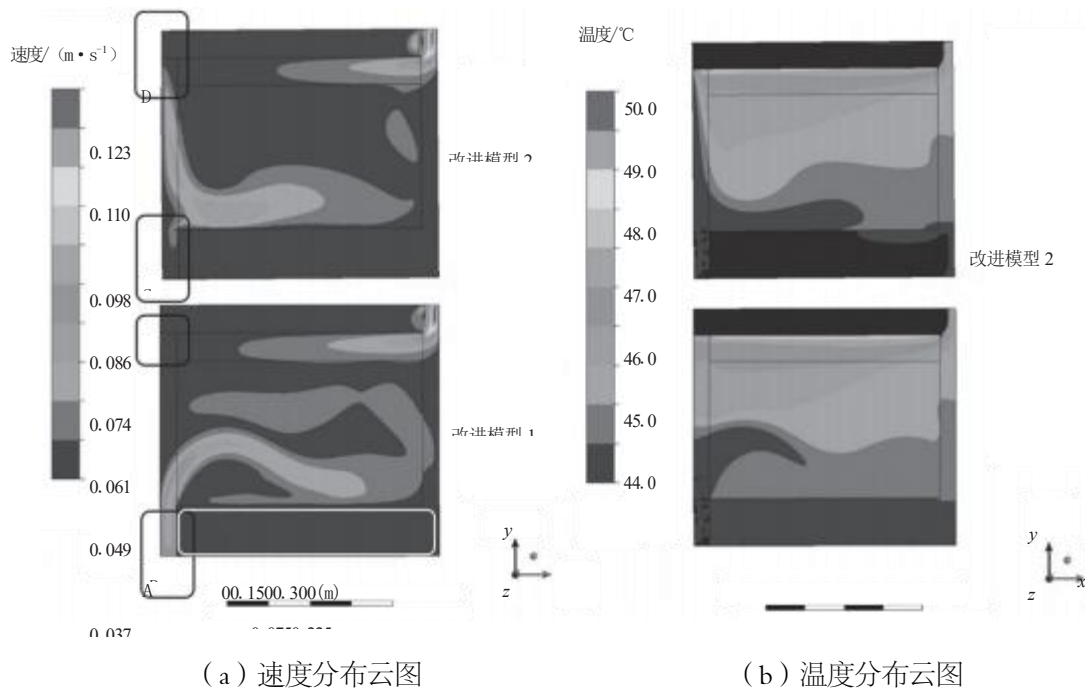


图 3 $X=-0.42m$ 速度和温度分布云图

基于计算成本的考虑，对服务器内部的器件进行多孔介质区域的简化，这种方法在机柜级模型是可以接受的。但是如果考虑空气介质和液体介质的巨大性质差异，基于空气进行热设计的 BBU 在冷却液里的散热性能需要考虑内部各元件和散热翅片的精细模型。模型采用空气和冷却液 2 种介

质，以相同的冷负

荷对 BBU 进行热交换，可以发现冷却液中散热器部位的流线稀疏，更多的冷却液从旁路流走，而空气模型并没出现类似现象，如图 4 所示。这说明，在浸没式液冷应用中，基于空气进行热设计的设备并不能很好地适合液冷环境。

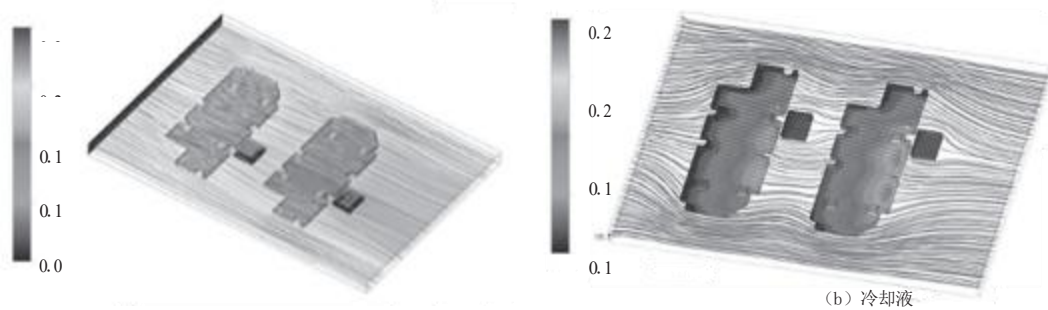


图 4 不同冷却介质流线图

基于此，设计了一种对比的方案，运行极限思维，将散热器的翅片去除，只留下一个 1mm 厚铝合金平板作为均温板。去除了散热器翅片的均温板模型，由于减少了翅片的阻力，更多的冷却液扫掠发热体，这样能够使得翅片的最高温由 94.08°C 下降到 58.98°C，整个板卡的温度场更加均匀。当

然，这只是一种理想的情况，实际板卡上各个元件发热体并不均匀，模型只是考虑均匀发热条件。但是此研究的结论可以作为一个有益的参考。

3 结论

以一个运行中的单相浸没式技术的液

冷 BBU 机柜为研究对象,建立了机柜级和设备级 2 种模型。对于机柜级模型,将 BBU 内部板卡和散热器组件简化为多孔介质模型,着重分析了原机柜 Tank 内部的流场,并针对其存在的局部热点问题,提出 2 种优化方案进行对比分析,得出改进技术方案。对于设备级模型,使用空气和冷却液进行仿真分析,并设置了去除散热翅片的均温板模型,显示翅片的最高温度下降了 35℃左右,这说明在进行液冷应用时对于原有采用空气热设计的设备需要进行一定的优化改造才能有更好的适用性。

参考文献:

- [1]付继焱,张泉,孟凡希,等.5G 机柜式热管空调一体机过渡季动态性能实测[J].科学技术与工程,2022(22):9616-9622.
- [2]张磊,苏俊,李佳.5G BBU 竖装对降低能耗的影响分析[J].电信工程技术与标准化,2021,34(1):27-32.
- [3]王飞,李毅潇,袁朋.基于 5G BBU 的风冷系统与液冷系统对比分析[J].长江信息通信,2021(3):221-224.
- [4]温亮.液冷技术在通信机房的标准化规模化运用研究[J].广东通信技术,2022,42(10):73-75.
- [5]侯福平.CRAN 模式下 5G BBU 供电及散热浅析(1):问题及思路[J].通信电源技术,2020,37(5):1-4.
- [6]刘帆,陶成,李振杰,等.5G 路由器液冷系统的设计和探究[J].电子机械工程,2021,37(4):30-34.
- [7]董雅洁,叶锐.某机载机箱的热仿真分析研究[J].电子机械工程,2022,38(4):44-47.

基于 MEC 打造 5G 智慧工厂

陈锋 李岩 潘晓宇

(中国联合网络通信有限公司福州市分公司)

摘要：本文主要研究如何通过部署 5G 网络基站进行厂区的全方位信号覆盖，使得厂区底层工控设备数据通过 5G 基站的 5G 网络传输至 UPF，UPF 再将数据传输流转至用户机房服务器，实现数据不出厂，数据用户自己管控的目的。并通过 5G+MEC 替代厂区内有线或无线 WiFi 网络，来达到高带宽，低时延的客户需求，为提升用户 5G 体验感知与品牌树立夯实的基础。

关键词：5G 网络基站；MEC；高带宽；低时延

1 概述

本项目拟在冠捷福建公司 F3 厂区的 A2 一楼 FA 车间和待验区、零配件加工车间部署 5G 网络基站进行全方位信号覆盖，底层工控设备数据通过 5G 基站的 5G 网络传输至 UPF，UPF 再将数据传输流转至用户机房服务器，实现数据不出厂，数据用户自己管控。A2 一楼 FA 车间和待验区、零配件加工车间的底层工控设备数据就近接入客户机房服务器部署，本地决策、实时响应、安全可靠，通过智能引擎分布到网络边缘，实现业务本地智能处理，以 5G+MEC 来替代厂区内有线或无线 WiFi 网络，达到高带宽，低时延要求。

2 MEC 建设方案

2.1 5G MEC 架构及部署方案

➢ 流量疏导方案

5G 核心网可通过 SMF 灵活的会话管理机制，实现本地流量疏导。5G 网络可采用 DNN 分流、“上行分类”功能和 IPv6 Multi-Homing 等实现本地流

量卸载。

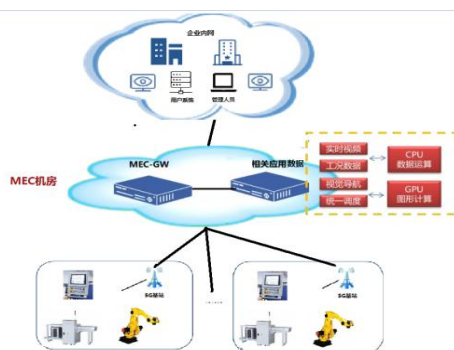


图 1 DNN 分流方案

1) DNN 分流方案：如图 2 所示，DNN 分流方案与传统 4G 时代的 APN 方案差异不大。网络侧 5GC SMF 基于用户所签约的 DNN 为用户选择特定 UPF，基站基于不同 DNN 面向不同的 UPF 建立 N3 接口，从而实现特定 DNN 流量由特定的 UPF 疏通。^[1]

作者简介：

陈锋：高级工程师，现任职于中国联合网络通信有限公司福州市分公司云网中心，云网专家，从事无线网络规划建设优化工作

李岩：现任职于中国联合网络通信有限公司福州市分公司云网中心，从事算力网建设工作

潘晓宇：现任职于中国联合网络通信有限公司福州市分公司云网中心，网络优化工程师，从事无线网络优化工作

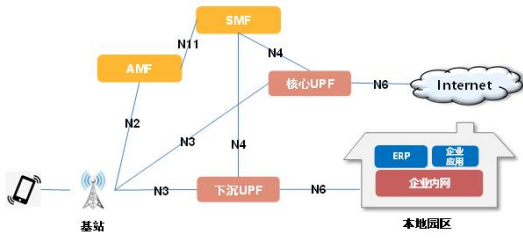


图 2 DNN 分流方案

2) 上行分类 (UL CL) 方案: 如图 3 所示, UL CL 的增加、删除由 SMF 依据切换过程中的终端位置决定, 当终端移入 MEC 覆盖区域时, SMF 通过 N4 接口对 UPF 增加 UL CL 功能和 PDU Session Anchor 完成本地流量通路的创建。SMF 可以在一个 PDU 会话的数据路径上引入多个支持 UL CL 功能的 UPF。PDU Session 可以是 IPv4 或 IPv6, UL CL 通过识别业务流的传输特征信息实现分流。[2]

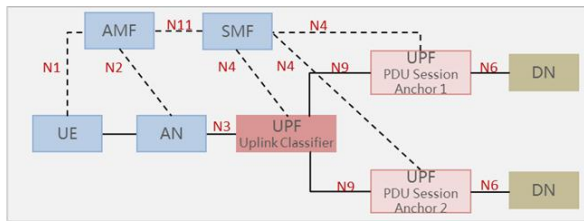


图 3 “上行分类”方案

3) IPv6 Multi-Homing 方案: 如图 4 所示, Multi-Homing 场景下通过对 Branching Point 的增加、删除完成对本地业务 Anchor 的创建, 并完成分流功能。SMF 通过 N4 接口对 UPF 功能进行控制。当会话为 IPv6 类型时通过 Branching Point 将需要分流的本地流量疏导到本地 Anchor 上。PDU session 可以与多个 IPv6 前缀关联, 提供多个 IPv6 PDU 锚点接入数据网络 DN。

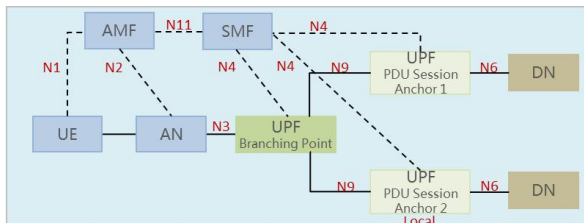


图 4 “IPv6 Multi-Homing”方案

► 业务连续性方案

为了支持移动性下会话与业务的连续性, 5G 网络提供三种不同的 SSC 模式。

SSC Mode1: UE 移动过程中, 无论 UE 所采用何种接入技术, PDU 会话建立时的 Anchor UPF 保持不变。这种模式类似于 LTE 网络中 PDN 锚点不变更的方式。此时 UE IP 不会发生变化。

SSC Mode2: 当终端离开当前 UPF 的服务区域, 网络会触发释放掉原有的 PDU Session, 指示 UE 立即建立与同一数据网络的新的 PDU 会话。建立新会话时, 可以选择一个新的 UPF 作为 PDU 会话 Anchor UPF, 此时需要保证新建立的 Session 信息和原 Session 信息的 UE IP 相同。

SSC Mode3: 当终端离开 Anchor UPF 的服务区域, 保持原有的 PDU Session 及 Anchor UPF, 同时通过选择新的 Anchor UPF, 并在该 Anchor UPF 上建立新的 PDU Session, 此时 UE 同时拥有到 2 个 Anchor UPF 的 PDU Session, 最后释放掉原有的 PDU Session, 在这个过程中 UE IP 保持不变。

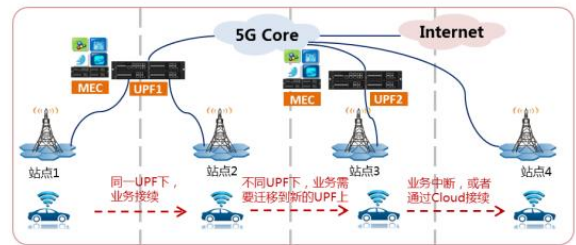


图 5 5G 网络会话与业务连续性示意图

根据运营商网络配置 SSC 模式选择策略, UE 可以为一个应用或者一组应用选择合适的 SSC 模式。在该策略中, 可以为所有应用配置一个默认 SSC 模式。如果 UE 没有为应用选择 SSC 模式, 网络可以根据签约信息、本地配置和应用请求等, 为该应用选择一个合适的 SSC 模式, 以支撑边缘计算业务连续性。例如图 2-4 所示, UE 移动到 UPF1 覆盖的区域内, 5G 核心网采用业务连续性 SSC Mode1, 并通过上行分类或 IPV6 Multi-Homing 的方式, 保持本地分流业务的连续性。当 UE 移动到 UPF2 覆盖的区域内, 5G 核心网采用业务连续性 SSC Mode3, 将业务迁移到新的 UPF2, 业务不中断。当 UE 移动到 MEC 覆盖的区域之外, 5G 核心网采用业务连

续性 SSC Mode2，业务中断或者通过 Cloud 接续。

➤ 智能感知与优化方案

在 5G 网络中，MEC 将与 5G 网络架构深度融合，其业务分流、策略控制、Qos 保证等功能，都将通过标准的 5G 网络功能实现。如图 6 所示，边缘计算应用(ME APP)通过 NEF(Network Exposure Function, 网络能力开放)与 5G 网络进行实时交互。一方面,NEF 将感知的 UE 和业务流相关测量信息,比如 UE 实时位置、无线链路质量、漫游状态等传递给 MEC 服务器,MEC 服务器基于上述测量信息通过智能分析和抽象,对应用的业务性能进行优化(例如调整视频播放码率),进而提升服务质量;另一方面,NEF 将感知的应用服务相关信息,比如业务时长、业务周期、移动模式等传递给网络,网络感知分析应用提供的这些信息,进一步优化其 UE 资源配置(例如为 VIP 用户分配合适的带宽资源)与会话管理。

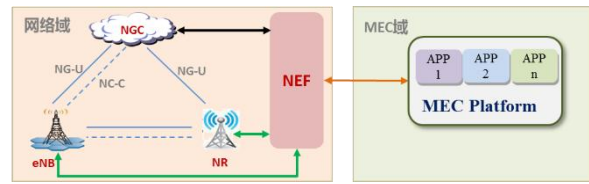


图 6 5G 智能感知与优化方案

针对 MEC 的各种应用，用户可经由 NEF 查询网络的各种状态，如计费策略、网络能力等，也可提交各种任务请求，如定位、带宽等。

2.2 MEC 平台部署

结合 ETSI 和 3GPP 对 MEC 的总体架构定义和要求，边缘平台主要的功能模块包括：MEAO、MEPM、MEP、ME-IaaS、ME-CaaS、以及以 Service 形式存在的 Paas 能力及产品等。[3]

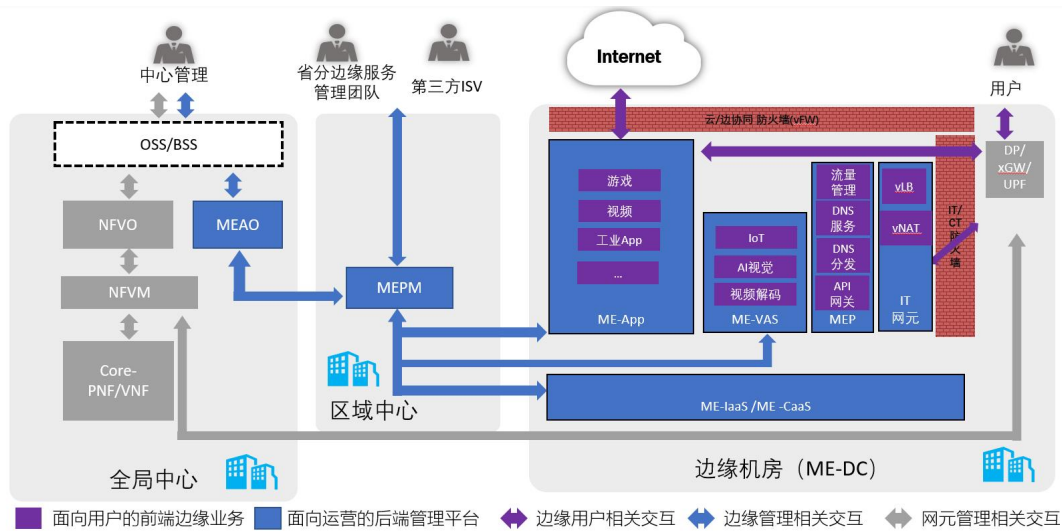


图 7 MEC 平台总体架构

2.2.1 平台运营管理功能

边缘平台的运营管理主要实现了以下功能：

实现边缘应用和服务的统一编排，包括应用引入、上线、部署、状态查询、升级、下线等生命周期管理；

实现边缘资源的统一管理，包括边缘应用包、边缘基础架构包等资源的管理，支持边缘机房的注册、上线、启用、下线等生命周期管理，支持边缘

基础架构的发现、部署、状态查询、资源监控、扩容、升级、下线等等生命周期管理。根据边缘资源的资源局限，支持不同形态的部署模式，实现边缘资源远程推送、无人值守、故障恢复、远程巡检等边缘运维能力；

提供边缘基础架构服务、平台服务和应用服务，和边缘网元协同，向外暴露平台服务和应用服务；

提供标准接口，实现边缘云和通信云、统一运

维平台等系统实现集成和协同,实现边缘服务的自动化,管理边缘服务状态,确保服务质量。

提供边缘应用描述符 MEAD (ME-App Descriptor) 标准规范,支持基于 IaaS (虚拟机方式) 和 CaaS (容器方式) 两种部署模式。

2.2.2 平台 ME-IaaS 功能

ME-IaaS 主要实现了以下功能:

- 接受 MEPM 的应用资源配置指令,识别 MEAD 描述符,实现配置并管理虚拟网络资源,配置并管理 ME-IaaS 资源

- 支持边缘-IaaS 的轻量化和多样化设计,根据边缘侧机房能力,提供单机、双机和三机及以上集群的四种边缘形态,实现拉远和下沉两种的不同部署方式。边缘侧分布式存储实现在资源紧凑模式下与控制、计算共享宿主机的集成模式。

- 基于 OpenStack 优化边缘侧的故障管理,在节点发生故障的时候,具备主动报警功能,能够显示关键、重要、轻微、告警等事件信息,统计不同高进的数量,可以汇报 ME-IaaS 故障,以及自动恢复相关进程。具备针对虚拟资源的操作记录,能够支持操作审计。

- 支持边缘 ME-IaaS 的资产管理,实现对控制节点、计算节点的信息收集,可以收集各个计算节点的主机名、状态、运行事件信息,节点的 CPU/内存/存储/网络/端口/接口等信息,显示网络拓扑和存储信息,显示网卡类型、端口、MTU 等信息。

- 支持显示底层节点的网络拓扑结构,暴露网络的状态,并支持对网络拓扑进行编辑和配置。

- 基于 OpenStack 优化边缘 ME-IaaS 高可用管理的 VM 高可用,实现更高效的 VM 切换效率;实现在边缘侧资源紧凑模式下的双节点模式下的平台高可用。

- 虚拟机内置代理,实现对 ME-App 部署时候的参数透传和配置,并实现部署后的进程级别状态远程监控和远程管理。

2.2.3 平台 ME-CaaS 功能

ME-CaaS 主要实现了以下功能:

- ME-CaaS 提供容器的能力,实现对容器化应用的承载,要求符合 CNCF Certified Kubernetes Conformance Program 规范,根据模板快速创建 Kubernetes 容器集群服务,提供 Kubernetes 服务管理配置管理能力,提供 kubernetes 服务管理配置。

- 支持容器基本功能,包括:创建、启动、停止、删除容器,应用服务启动、停止、重启、升级、回滚、自动注册、扩缩容,多租户管理、可集成 LDAP、用户授权,镜像管理、日志管理、故障恢复等。

- 支持微服务基本功能,包括服务自动注册、发现和故障剔除等。

2.2.4 MEP 主要功能

MEP 主要实现了以下功能:

- 边缘流量控制功能 (Traffic rule control): 支持定义应用流量从 UE 到 ME-App/ME-VAS 的流量管理,支持激活、更新流量规则,将特定的业务流路由至某个 ME-App/ME-VAS,允许将流量策略关联/取消关联边缘应用。边缘流量控制功能 (Traffic rule control) 包括对 UPF/DP/xGW 等通过 3GPP 标准规范的流量下发,也包括对边缘内部相关 IT 网元提供的二次分流能力的调度。

支持定义管理流量,允许将流量策略关联/取消关联到网络端口 Interface (虚拟网卡/物理 PCI 网卡),定义与 DP/UPF/xGW 对接的接口、与 internet 公网对接的接口、与 Management 管理网络对接的接口和与本地虚拟网络接口对接的接口等可选底层管理接口。

- 边缘 API 网关 (API Gateway):

支持对 API 进行统一管控,实现服务的注册认证、鉴权、协议转换、监控等。ME-App 可以发现并定位所请求的 ME-VAS 服务的 end-point 信息,VAS 也可以将相关的接口和版本信息通知给其他的 服务或应用,支持 ME-App 到 ME-VAS 的服务访问控制,包括访问频率、最大次数、最大连接数等。实现边缘服务 API 层访问的熔断控制、访问控制和访问统计,降低服务能力对接复杂度,实现更精准的能力资源分配,提升平台的稳定性。

- 边缘 DNS 域名解析服务 (DNS Library):

支持建立边缘机房的本地 DNS 域名解析服务,实现 DNS 的注册、移除、变更、查询的功能。

- 边缘 DNS 分发 (DNS Handling):

配合 DNS 分流能力,实现相应的边缘分流处理。

2.2.5 平台网络能力开放及 Paas 能力

平台的网络能力开放及主要 Paas 能力包括:

- 平台可以实现针对 LBS、RNIS、带宽管理

能力的对外开放。

平台目前可以实现的主要 PaaS 能力包括：AR/VR 渲染、视频转码、AI 人脸识别、vCDN 等。

3 MEC 落地部署模式分析

模式一：UPF/分流网关与平台同步下沉部署
该种模式下具体组网如下图所示：



图 8 UPF/分流网关与平台同步下沉部署

该种模式类似于客户专供模式，UPF/分流网关、边缘平台均下沉部署，可以实现客户流量不出园区，客户应用本地部署。

该种模式优点是业务时延体验好，可以较好的满足客户个性化需求，但是成本相对较高，需要客

户具备较高的支付能力。同时由于部署位置比较靠下，对网络搭建、运营维护要求都会比较高。

模式二：UPF/分流网关共享，边缘平台下沉
该种模式下具体组网如下图所示：



图 9 UPF/分流网关共享，边缘平台下沉

该种模式采用共享UPF/分流网关、边缘平台下沉部署，UPF/分流网关与边缘平台之间采用承载网或者专线对接。该种模式适用于客户无流量不出园区要求，时延不太敏感，但是应用需部署在客户本地的场景。

该种模式优点是平台及应用可以部署在客户侧，同时UPF/分流网关采用共享模式，因此成本相对较低，客户支付压力相对较小。

模式三：UPF/分流网关、边缘平台均共享建设
该种模式下具体组网如下图所示：

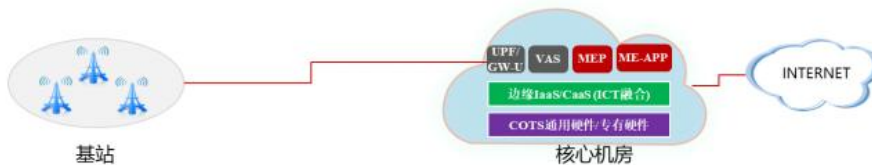


图 10 UPF/分流网关、边缘平台均共享建设

该种模式采用UPF/分流网关、边缘平台均共享模式，UPF/分流网关与边缘平台均部署在核心或者

汇聚机房。该种模式采用全共享模式，设备及平台部署层

级相对较高，业务覆盖范围广，便于维护管理，且客户付费压力小。适用于 2C、2B2C 场景或者 2B 但是客户无流量不出园区要求、时延不太敏感的场景。

景。

模式二：UPF/分流网关共享，边缘平台下沉
该种模式下具体组网如下图所示：



图 11 平台下沉部署并集成 DP 功能

该种模式边缘平台下沉部署并采用平台自带的 DP 功能进行分流，无需分流的流量向上透传至核心网。

该种模式仅适用于 4G/NSA 场景，对于客户要求流量及应用不出园区但是不采用流量计费及对安全监听无要求的场景，可以采用该种模式。

该种模式优点是不需专门部署分流网元，规避了 4G/NSA 下分流网关厂家报价虚高的情况，缺点是方案安全保障较差，也无法提供监听及流量计费，需酌情选用。

4 建设内容

本期 MEC 节点部署内容及组网如下图所示：

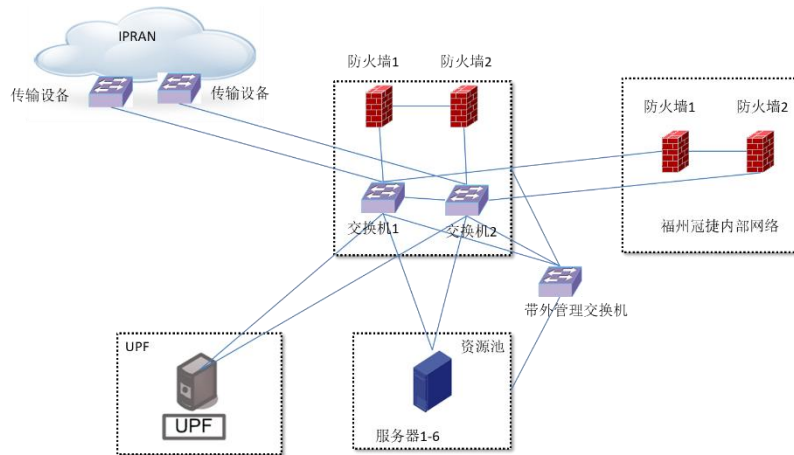


图 12 福州冠捷 MEC 节点组网图

现场级 MEC 节点主要内容如下：

本期 UPF EOR 交换机和站点 SPINE 交换机共用 1 对；

SPINE 交换机对接智能城域网 MER 设备或者 IPRAN 设备，在保证安全的前提下，将 SPINE 与客户交换机直连；

MEC 资源池 <-> 边缘 UPF，MEC 资源池至其他网络，均需要采用防火墙做安全隔离；

防火墙组网：防火墙侧挂在 SPINE 交换机上；

4.1 服务器硬件

边缘平台可基于 x86 通用服务器及 ARM 服务器部署，针对特殊场景，具备特殊尺寸服务器、GPU 服务器等的提供能力。

目前各厂家虚拟化 UPF/GW 主要采用专用硬件的方式，集团正在推动厂家虚拟化 UPF/GW 基于通用服务器部署，目标是可实现 UPF/GW 与平台共硬件资源池部署。现阶段 UPF 可暂时采用厂家专用硬件的方式。

本期工程拟利旧现有服务器硬件建设。

4.2 UPF/GW

边缘分流网关，5G SA采用标准UPF分流：

本项目按需新建边缘UPF；

本期工程新建MEC（利旧设备）和UPF主要满足福州冠捷相关的MEC行业业务需求。

4.3 交换机/防火墙

边缘节点需要通过交换机连接各上层网络（B网、IP RAN、智能城域网等），需要防火墙隔离各个系统（分流网关、平台等）。

本期工程涉及交换机/防火墙建设。

4.4 平台 MEP/MEPM

各节点均需部署MEP，部分核心节点涉及MEPM建设。MEP/MEPM集团均已完成开发，各省可直接使用，本项目不涉及平台部署。

4.5 ME-APP

各节点均需部署ME-APP，ME-APP可为自研或者客户APP（需与平台对接验证），ME-APP部署原则上不涉及费用，本项目不涉及APP部署费用。

4.6 设备选型要求

本项目硬件设备选型主要涉及UPF设备，UPF依据5GC采购结果，一般采用厂家软硬一体化设备。

5 网络安全防护

根据YD/T 1729-2008《电信网和互联网安全等级保护实施指南》对移动核心网系统进行安全等级定级，采用数法确定该系统的安全等级，分析如下：

（1）本系统主要为用户提供移动语音和数据等业务服务，本系统被损害后对国家安全不产生影响，也不会对其他重要行业、企事业单位造成影响，仅影响用户对该业务的使用，因此社会影响力的赋值为2；

（2）本系统为厦门冠捷提供移动网业务，相关核心网设备被损害后会对区域内（核心网网元覆盖的区域）的业务正常使用造成影响，因此规模和服务范围的赋值为3；

（3）移动网相关业务对于提升品牌价值形象，保持用户粘性，提高市场竞争能力、巩固市场的占有率有重要意义，因此本系统所提供服务的的重要性为3；

（4）采用对数法，取 $\alpha = \beta = \gamma = 1/3$ ，经计算得本系统的安全等级值为2.7，则本系统的安全等级是第3.1级。

6 结束语

本文提出的优化部署策略可以较好的融合移动网相关业务，对于提升品牌价值形象，保持用户粘性，提高市场竞争能力、巩固市场的占有率有重要意义。

参考文献：

- [1] 张芳,伍林伟,蒋永彬等.基于专用DNN的5G双域专网校园网部署方案[J].移动通信,2023,47(01):12-17+76.
- [2] 严会霞,曲云盛,王刚等.PCF基于TA下发动态规则实现5G ULCL分流方案[J].通信与信息技术,2023(02):94-97.
- [3] 唐海东.中国联通MEC边缘云平台建设分析[J].数字通信世界,2023(06):17-19+40.

一种基于六维业务特征分析的 PCDN 流量检测工具

陈倩雯 林伟 杨慰民

(中国移动通信集团福建有限公司)

摘要: PCDN, 即在 CDN(内容分发网络)的基础上扩展传输方式, 多数扩展方法都和 P2P 技术相关, 主流技术包含 PCDN、HCDN 等, 该技术在一定程度上降低了互联网厂商的成本, 但换来的是用户业务感知的不稳定, 运营商利益的被侵害, 信息安全的漏监管, 因此本文通过一种基于 netflow 流量加 DNS 解析日志识别的 PCDN 识别方法, 挖掘违规 PCDN 用户, 督促行为整改, 实现互联网业务品质提升。

关键词: P2P; PCDN; 流量; 分析

引言

PCDN 即 P2P CDN, 是以 P2P 及 CDN 为技术基础, 通过挖掘利用运营商低价互联网流量开展内容分发的行为, 其中, CDN 是 Content Delivery Network 的缩写, 即内容分发网络, CDN 系统通过专用网络及带宽设备, 存储互联网中的热点内容或特定内容, 网络稳定有保障。而 PCDN, 是利用 P2P 下载来提升 CDN 性能, 也就是分布式的碎片化存储, 在本地下载好资源后也可以作为一个下载节点供其他用户下载。

在普通用户的上网流程中, 用户通过常规的 http 请求、DNS 请求等完成互联网访问, 而通过引导网内用户的访问请求至 CDN 系统中, 由 CDN 系统为用户

提供所需的互联网访问数据, 避免直接从互联网网站下载数据, 从而降低客户访问互联网的时延, 有效提升用户访问速度和业务质量、解决互联网出口的拥挤问题。

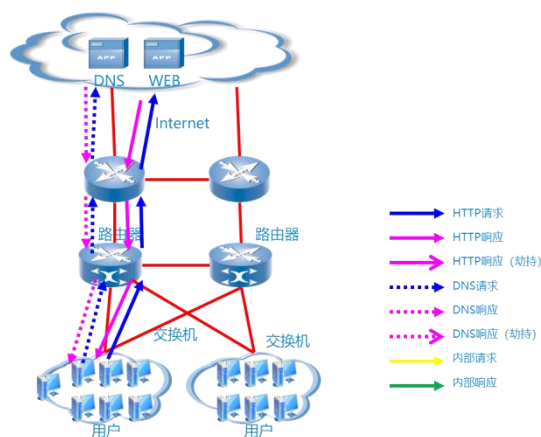


图 1 普通用户上网流程 (资源使用 CDN)

而 PCDN 是针对 CDN 的一个突破, 最初用在 BT 下载等应用, 在一定程度上减少了 BT 内容的资源存储也提升了用户下载速率, 但现阶段, 该技术已被滥用, 不少互联网公司, 通过大量采购运营商互联网接入服务产品 (如家庭宽带、企业宽带、商务快线、互联网专线等), 通过 P2P 技术将上行带宽聚合用于内容分发, 替代、冲击运营商高价值

作者简介:

陈倩雯: 中国移动通信集团福建有限公司网络部网管中心员工

林伟: 中国移动通信集团福建有限公司网络部网管中心员工

杨慰民: 中国移动通信集团公司首席专家; 中国移动通信集团福建有限公司网络部网管中心副总经理

IDC 带宽（即前文提到 CDN），造成运营商整体利益损失。同时，PCDN 目前大量应用于视频类业务分发，因资源点未使用专用网络，导致视频业务质量受损，严重影响用户感知，同时因为私自提供 CDN 商业服务，缺乏相关 CDN 运营资质，违反了工信部的相关要求，存在监管和信息安全问题。

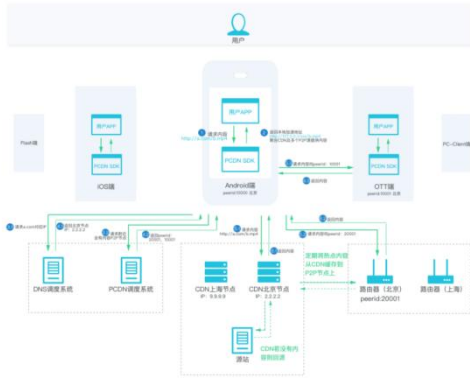


图 2 普通用户上网流程（资源使用 PCDN）

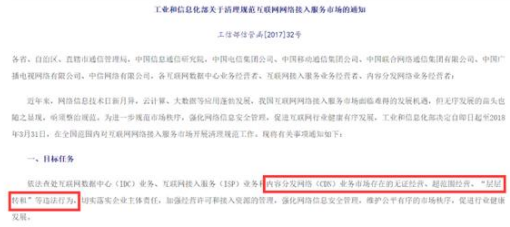


图 3 监管部门对于 PCDN 的相关要求

基于此，本团队进行分析研究，并得出一种基于六维业务特征分析的 PCDN 流量检测工具，方案对现有网络管道流量进行精细化分析，搭建 PCDN 业务特征分析模型，通过 FLOW 流量日志、DNS 解析日志对 PCDN 业务的流量进行识别与分析挖掘，6 项流量特征主要包括用户上下行流量比值分析、用户常用目的端口占比分析、用户上行 TOP 源端口流量占比分析、用户服务本省用户数量占比分析、用户上行 UDP 流量占比分析和用户访问已识别特征库域名匹配分析。

1.PCDN 识别模型

本方案构建 PCDN 特征识别模型,基于 FLOW 日志、DNS 日志进行分析，主要功能架构包括数据底座、数据处理和数据应用层。



图 4 底层数据交叉关联逻辑示意图

1.1 数据底座：基于城域网出口的 NetFlow 日志、省 DNS 日志以及 Radius 数据进行数据分析与挖掘。各数据源作用如下：通过城域网用户 NetFlow 流量能够分析用户 IP 上行流量、下行流量和端口流量等数据。通过 DNS 日志能够进一步分析 TOP 风险用户交互的风险域名。Radius 数据实现用户 IP 与用户带宽账号关联，实现用户级别的定位。

1.2 数据处理

在对 FLOW、DNS 日志进行预处理基础上，开展 PCDN 行为特征分析。

首先，对初步筛选出 PCDN 违规用户 IP 资源池。主要通过两个分析渠道筛选出 PCDN 异常用户。第一个是基于城域网用户流量和用户 ACL，筛选出 ACL 地址段中符合用户 IP 的上行流量、下行流量

及上下行流量比值大于指定阈值的。第二个是基于城域网用户流量和已识别 PCDN 特征库,筛选出特征库用户 IP 中符合用户 IP 的上行流量、下行流量及上下行流量比值大于指定阈值的。其中,通常设置过滤出上行流量大于 10M,且上行流量/下行流量 > 5 的用户 IP(这两个阈值根据实际场景需求设置)。将两个渠道的筛选结果进行汇聚获得初步筛选出 PCDN 违规用户 IP 资源池。

其次,基于构建的 PCDN 特征项进行统计分析,包括用户上下行流量比值分析、用户常用目的端口占比分析、用户上行 TOP 源端口流量占比分析、用户服务本省用户数量占比分析、用户上行 UDP 流量占比分析和用户访问已识别特征库域名匹配分析。

(1) 用户上下行流量比值分析:将用户上下行流量比值作为 PCDN 特征项之一,由于 PCDN 用户的上行带宽使用远高于正常用户,且上行带宽相对下行带宽使用越多则风险越高。

(2) 用户常用目的端口流量占比分析:PCDN 用户相当于 CP 的边缘服务节点,其作为服务端,上行流量中的大部分流量的目的端口随机性较强,而常用目的端口的流量比例较小。统计用户上行流量中常用目的端口流量的占比对于 PCDN 用户特征分析具有重要意义,常用目的端口流量占比越小则用户风险越高。

(3) 用户上行 TOP5 源端口流量占比分析:正常用户的源端口通常是随机分配的,各源端口流量呈现均匀分布状态。而 PCDN 用户的上行流量通

常聚集在一些特定的源端口上,呈现正态分布状态。将用户上行流量按源端口从大到小排序,统计 TOP5 源端口流量及占比,占比越大则用户风险越高。

(4) 用户服务本省用户数量占比分析:该项的统计方式是该用户服务本省用户的数量/所有待分析异常用户服务本省用户的数量*100%。根据上行流量中目的 IP 的归属判断是否是服务本省用户。在所有待分析异常用户中,服务本省用户的数量越多则风险越高。

(5) 用户上行 UDP 流量占比分析:基于历史抓包经验,PCDN 用户上行流量中的 UDP 协议流量占比通常高达 60%,且主要传输的是 data 数据。通过 FLOW 日志中上行流量中 UDP 流量占比越多则风险越高。

(6) 用户访问已识别特征库域名分析:若待分析异常用户的 IP 在已识别 PCDN 特征库中,则该项的风险系数为 100。反之,则该项的风险系数为 0。

对上述特征数据进行模型参数调优,加权计算,为疑似 PCDN 用户计算综合得分。

1.3 解决方案

面对传统 DNS 方案抓包成本高不具有可持续性、DPI 识别协议需要持续升级等问题,本成果构建 PCDN 基于流量的特征识别机制,规避了 PCDN 协议经常变更调整导致模型失效的问题。该模型基于 PCDN 流量模型特征,正常家宽用户与 PCDN 业务用户呈现特征如下表所示:

表 1 正常家宽用户与 PCDN 业务用户呈现特征

特征类别	正常家宽用户	PCDN 业务用户
1、上下行流量特征	上行流量远低于下行流量	上行流量远高于下行流量
2、源端口特征	源端口数值分散随机	源端口有聚合
3、目的端口特征	目的端口聚合,常访问 8080、443、53 等	目的端口分散随机
4、服务本省家宽用户特征	数量较少	数量较多
5、上行 UDP 流量特征	上行 UDP 流量较少	上行 UDP 流量较多
6、已识别 PCDN 盒子交互域名特征	与 PCDN 盒子无交互,没有相关域名	与 PCDN 盒子交互,有域名交互数据

6 项 PCDN 风险特征项,包括用户上下行流量比值项、用户上行 TOP5 源端口流量占比项、用户

常用目的端口占比项、用户服务本省用户数量占比项、用户上行 UDP 流量占比项和用户访问已识别

特征库域名项。并根据 PCDN 用户表现所在区间进行判断打分，实现多维度、综合、全面发现 PCDN 服务并评估 PCDN 服务风险等级。

计算综合风险系数=各项风险系数*分项权重之和

表 2 风险系数判定表

序号	类别	参数设置
1	设定流量维度异常模型，上下行流量占比区间算法（%）	0-50 风险系数 0 50-100 风险系数 20 100-200 风险系数 40 200-300 风险系数 60 300-400 风险系数 80 400 以上 风险系数 100
2	设定目标端口流量维度算法模型（常用目标端口与非常用目标端口的流量占比（%））	0-10 风险系数 100 10-20 风险系数 80 20-30 风险系数 60 30-40 风险系数 40 40-80 风险系数 20 80-100 风险系数 0
3	设定源端口流量维度算法模型（TOP5 源端口流量与其他源端口流量占比区间（%））	0-15 风险系数 20 15-30 风险系数 40 30-50 风险系数 60 50-75 风险系数 80 75-100 风险系数 100
4	设定服务本省用户占比算法模型（服务本省家宽用户数占比（%））	0-10 风险系数 0 10-50 风险系数 20 50-100 风险系数 40 100-150 风险系数 60 150-200 风险系数 80 200 以上 风险系数 100
5	设定上行 UDP 流量占比算法模型（%）	0-10 风险系数 0 10-20 风险系数 20 20-40 风险系数 40 40-60 风险系数 60 60-80 风险系数 80 80-100 风险系数 100

2.系统应用与实践

通过系统部署，进行部分流量接入试点，将 A 地 DNS 和 netflow 数据接入系统。其中，DNS 日志在已采集的省 DNS 进行过滤筛选，通过 A 地市地

址段过滤出 A 地市 DNS 日志；netflow 日志通过接收 A 地市城域网 netflow 流量进行采集形成。

设备资源需求需 3 台服务器，其作用及配置如下：

表 3 硬件需求情况

服务器数量	用途	配置
1	负责接收和预处理 A 地市 DNS 日志	24 核 CPU、256G 内存、20TB 存储硬盘、10GE 网卡
1	负责接收 A 地市城域网 netflow 流量，形成 netflow 日志	24 核 CPU、256G 内存、20TB 存储硬盘、10GE 网卡
1	负责综合 DNS 和 netflow 进行综合分析 PCDN 异常用户	24 核 CPU、256G 内存、20TB 存储硬盘、10GE 网卡

设备资源到位后，部署相关 DNS 预处理、netflow 采集、以及 PCDN 分析软件，并进行 PCDN 异常特征分析参数配置。

通过将 A 地市 DNS 日志和 netflow 日志融合分析，综合模型算法进行计算，将风险系数高于 50% 的用户提取出来，并结合 PCDN 异常域名库进行核对，验证其 PCDN 的检测准确性。

2.1 支撑 PCDN 用户识别及关停

通过该系统，支撑业务部门关停违规开通 PCDN 的专线 323 条，涉及带宽 180Gbps，下线企宽 6562 条，涉及带宽 446Gbps，有效打击违规 PCDN 业务的发展。

对于关停的 PCDN 流量，评估约 30% 回流省内正规 IDC 及 CDN 分发流量，助力无序流量归位，实现提质增收。

用户IP	用户类型	用户名称	归属地市	用户地址	总流量	上行平均流量	上行峰值流量	下行平均流量	下行峰值流量	上下行流量占比	上行UoP流量	上行UoP流量占比
总计	-	-	-	-	66.73Gbps	58.43Gbps	157.89Gbps	7.30Gbps	30.77Gbps	814.54	7.33Gbps	12.33
112.5	专线	-	市	-	1.20Gbps	1.15Gbps	2.36Gbps	47.33Mbps	107.39Mbps	2431.08	2.89Mbps	0
183.252	专线	-	市	-	803.50Mbps	561.18Mbps	788.92Mbps	242.33Mbps	438.09Mbps	231.58	2.03Mbps	0
112.51	专线	-	市	-	590.96Mbps	518.56Mbps	847.13Mbps	72.40Mbps	384.02Mbps	716.23	497.21Mbps	0
112.5.1	专线	-	市	-	498.89Mbps	403.09Mbps	710.27Mbps	95.80Mbps	157.38Mbps	420.76	4.35Mbps	0

图 5 PCDN 用户列表

2.2 支撑评估 PCDN 处置成效

近期通过开展 PCDN 网络流控，结合业务部门专项打击关停工作，有效压降 PCDN 相关流量，从 A 地市的试点系统来看，可明显看出 A 地市 PCDN 流量及用户的下降趋势。

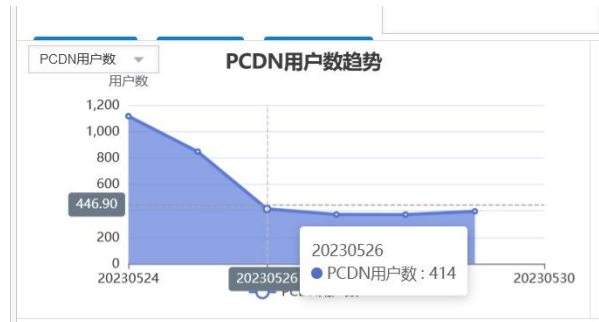


图 7 PCDN 用户变化示意图

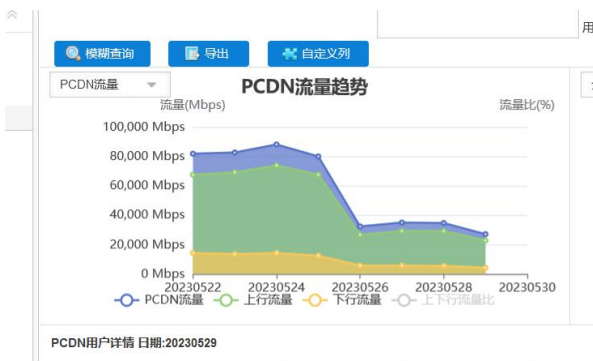


图 6 PCDN 流量变化示意图

2.3 支撑疑似 PCDN 用户证据链

本系统不仅实现疑似 PCDN 用户清单的列表输出，同时针对每个疑似用户，均有明确的证据链数据进行支撑，含上文提到的 6 维业务特征，提供更加详细用户清单，支撑各部门核查。



图 8 疑似 PCDN 证据链数据

3.应用价值

本系统的落地应用，切实解决了前端业务部门对于 PCDN 用户精准识别的需求，以及后端业务维护部门对于 PCDN 业务流量整体情况、变化情况的把我，同时，良好的交互体验，层层下钻分析的系统逻辑，开放解耦灵活配置条件，给后续的特异性流量识别奠定了良好的基础。

3.1 商业价值

PCDN 的发展，对于运营商的高价值 IDC 及 CDN 资源是极大的冲击，如何精准识别 PCDN，并进行有效控制，是各个运营商亟需解决的难题。

本系统通过深入研究 PCDN 的业务流程及行为特征，建立一套成熟的 PCDN 异常流量检测模型。基于全省 DNS 日志、城域网出口的 NETFLOW 流量以及已识别的 PCDN 特征库，过滤筛查 PCDN IP 及 PCDN tracker IP，多维度呈现 PCDN IP 的异常特征，指导运营商对此类违规行为进行排查治理。

通过模型评估，在试点地市完成流控后，能够实现 30%流量转换为省内高价值 IDC 流量，按照 1.5T 的 PCDN 流量进行换算，预估可提升运营商收入约 200 万/月。

3.2 网络价值

在网络运维方面，PCDN 对于承载网上行压力较大，可能造成承载网利用率较高，形成丢包等相关网络问题，通过 PCDN 流量压降，上行流量有效降低 51%，确保承载网络有序运行。

4.3 用户价值

1、本方案的 6 维业务特征可为解耦的分布式评估方案，对于用户的运维有着极大的便利，对于任何一项的业务特征评估标准，均可以根据 PCDN 业务的发展、运维的需要、打击力度的需要进行修订，更好的服务于当下的定位精准度的需求。

2、使用便捷，用户界面简洁明了，图形化界面，简单的查看使用说明，可轻松上手使用该系统，可根据需要查询的维度和报表类型，输出结果，系统提供了 API 的接口，外部系统可通过 API 接口实时调用定制的查询结果，大大降低了使用难度。

4.结束语

内容流量的有序调度，是确保视频、网站、下载业务体验感知良好的前提，现今各互联网产商通过 PCDN 的技术手段压降流量成本，一是导致用户的业务体验下滑，二是内容分发的安全性脱离了监管，三是违背了与运营商签约的合同条款，严重冲击了运营商的高价值带宽收入。

因此，本系统通过结合 DNS 域名解析日志数据、netflow 流量数据，并建立六维业务风险指数评估精准定位 PCDN 用户，通过网络限制、业务部门督促整改，进行 PCDN 流量压降，引导内容资源往有序的 IDC 及 CDN 流量迁移，从用户感知、信息安全、带宽收入三大维度协同提升。同时，本系统在 PCDN 用户的定位外，还提供整体 PCDN 流量及用户数量的趋势变化分析，有助于专项运维人员的高效运维，并进行 PCDN 限制操作的后评估，从宏观把握整体 PCDN 业务发展形势。

智能算力与医疗产业的融合发展

李 晨

(中国电信股份有限公司福建分公司)

摘 要： 本论文深入研究了智能算力在医疗产业中的应用和创新模式，并探讨了其对医疗资源整合和诊断辅助的影响。在当今数字化时代，智能算力作为一种新型生产力正逐渐渗透到各个行业，医疗产业也不例外。智能算力的快速发展为医疗产业带来了许多机遇和挑战，对推动医疗产业转型升级具有重要意义。

关键词： 智能算力、医疗产业、医疗资源整合、诊断辅助、创新模式

本文首先分析了智能算力在医疗资源整合方面的应用和效果，探讨了如何通过智能算力实现医疗资源的合理配置和优化利用。其次，研究了智能算力在诊断辅助方面的应用，探索了智能算力在医学影像分析和辅助诊断中的效果和优势。进一步探讨了智能算力在远程医疗方面的应用，分析了智能算力对远程医疗服务的改进和提升作用。最后，探索了智能算力在医疗产业中的创新模式，为医疗产业的融合发展提供了有益借鉴和启示。

本研究采用了多种研究方法，包括文献综述、案例分析和数据收集与分析。通过对相关文献的综述和代表性案例的深入分析，全面展示了智能算力在医疗产业中的应用和创新模式。同时，收集来自医疗机构、相关统计数据 and 已有研究报告等多个来源的数据，通过合适的统计方法进行分析，评估智能算力在医疗产业中的应用效果和潜力。

研究表明，智能算力在医疗产业中的应用和创新模式具有巨大的潜力。从实时数据分析与决策支持到个性化医疗与精准诊疗，再到远程医疗与智慧健康，智能算力为医疗产业的发展带来了新的机遇和挑战。智能算力在医疗资源整合方面的应用可以实现医疗资源的合理配置和优化利用，提高医疗服务的效率和质量。智能算力在诊断辅助方面的应用可以帮助医生进行准确的诊断和治疗决策，提高医疗的精准性和个性化水平。智能算力在远程医疗方面的应用可以改进远程医疗服务，提高患者的便利性和医疗体验。智能算力在医疗产业中的创新模式为医疗产业的融合发展提供了有益借鉴和启

示。

引言

在当今数字化时代，智能算力作为一种新型生产力正逐渐渗透到各个行业，其中医疗产业也不例外。智能算力的快速发展为医疗产业带来了许多机遇和挑战。智能算力的应用在医疗领域展现出巨大的潜力，可以加速医疗资源整合、改善诊断准确性、提升远程医疗服务质量，推动医疗产业的融合发展。因此，深入研究智能算力在医疗产业中的应用和创新模式对于推动医疗产业转型升级具有重要意义。

(1) 研究背景和意义

医疗产业是人们生活中不可或缺的一部分，随着人口老龄化、疾病谱变化和医疗需求的增加，医疗资源的整合和优化利用成为一个迫切的需求。智能算力作为一种新兴技术，具有强大的计算和分析能力，可以帮助医疗机构实现医疗资源的合理配置和高效利用。智能算力与医疗产业的融合发展不仅可以提升医疗服务的质量和效率，还可以推动医疗产业的创新和可持续发展。

近年来，随着人工智能、大数据和云计算等技术的快速发展，智能算力在医疗产业中的应用呈现出爆发式增长的态势。通过智能算力的支持，医疗机构能够更好地利用和分析海量的医疗数据，提高诊断准确性和效率，改善患者的医疗体验，推动医疗资源的合理分配和优化利用。智能算力还为远程医疗提供了技术支持，使患者能够更方便地获得专业的医疗服务。因此，研究智能算力在医疗产业中

的应用和创新模式对于提升医疗服务质量、改善患者体验和推动医疗产业的可持续发展具有重要意义。

(2) 研究目的

本研究旨在深入探讨智能算力在医疗产业中的应用和创新模式，具体目标如下：

分析智能算力在医疗资源整合方面的应用和效果，探讨如何通过智能算力实现医疗资源的合理配置和优化利用。研究智能算力在诊断辅助方面的应用，探索智能算力在医学影像分析和辅助诊断中的效果和优势。探讨智能算力在远程医疗方面的应用，分析智能算力对远程医疗服务的改进和提升作用。探索智能算力在医疗产业中的创新模式，为医疗产业的融合发展提供有益借鉴和启示。

(3) 研究方法和数据来源

本研究将采用多种研究方法，包括文献综述、案例分析和数据收集与分析。首先，通过对相关文献的综述，了解智能算力在医疗产业中的应用现状和发展趋势。在研究背景中已经提到的智能算力在医疗产业中的应用案例和趋势将作为研究的基础。其次，选取具有代表性的案例进行深入分析，以揭示智能算力在医疗资源整合、诊断辅助和远程医疗方面的应用实践和效果。

在数据收集方面，本研究将收集来自医疗机构、相关统计数据 and 已有的研究报告等多个来源的数据。这些数据将通过合适的统计方法进行分析，包括描述性统计、回归分析和相关性分析等，以评估智能算力在医疗产业中的应用效果和潜力。

在确保数据的可靠性和有效性方面，本研究将采取以下措施：首先，确保所选取的案例具有代表性和典型性，能够反映智能算力在医疗产业中的应用现状和发展趋势。其次，对数据进行严格的筛选和验证，排除不准确或不完整的数据。此外，合理选择适当的统计方法，以确保数据的分析结果具有统计学上的可靠性。

通过以上研究方法和数据来源的综合应用，我们将深入探讨智能算力与医疗产业的融合发展，为医疗产业转型升级提供实证研究和政策建议。

1. 智能算力在医疗产业中的应用概述

1.1 智能算力的定义和特点

智能算力技术正在快速发展，人工智能、大数据、云计算等先进技术的融合使计算机系统具备自主学习、推理、决策和解决问题的能力。智能算力具有高速计算、大规模数据处理和智能决策等特点，为医疗产业带来了革命性的变革和创新。

智能算力在医疗产业中的应用领域包括实时数据分析与决策支持、个性化医疗与精准诊疗以及远程医疗与智慧健康等。在这些领域，智能算力的技术和方法发挥着关键作用，为医疗产业的发展提供了新的机遇和挑战。

1.2 医疗产业中的应用领域

1.2.1 实时数据分析与决策支持

实时数据分析与决策支持是智能算力在医疗产业中的重要应用领域。通过智能算力技术，医疗机构可以快速处理大规模的医疗数据，并提取其中有价值的信息，为医疗决策提供科学依据。具体应用包括病人监测、病情预测、疾病风险评估和治疗方案优化等。

1.2.1.1 案例 1：智能算力在医疗数据分析和决策支持中的应用

一家医疗机构利用智能算力技术对大量的病历数据进行分析，通过关联规则挖掘等数据挖掘算法，发现了一些潜在的病情关联和治疗模式。这些发现为医生提供了更准确的诊断和治疗建议，提高了医疗决策的科学性和精准性。

1.2.2 个性化医疗与精准诊疗

个性化医疗与精准诊疗是智能算力在医疗产业中的另一个重要应用领域。通过智能算力技术，医疗机构可以根据患者的个体特征和基因信息，制定个性化的治疗方案和预防策略，实现精准诊疗。

1.2.2.1 案例 1：智能算力在个性化医疗和精准诊疗中的应用和效果

一项研究利用智能算力技术对患者的基因测序数据进行分析，根据患者的遗传变异和基因表达数据，预测患者对特定药物的反应和治疗效果，制定个性化的用药方案。这种个性化医疗的应用大大提高了治疗的效果和可行性，为患者提供了更精准的诊疗服务。

通过以上案例，我们可以看到智能算力在医疗产业中的应用和创新模式具有巨大的潜力。从实时

数据分析与决策支持到个性化医疗与精准诊疗,智能算力为医疗产业的发展带来了新的机遇和挑战。同时,随着智能算力技术的不断进步和创新,我们可以期待更多智能算力在医疗产业中的广泛应用和突破。

1.2.3 远程医疗与智慧健康

智能算力在远程医疗和智慧健康领域的应用也具有重要意义。通过智能算力的支持,患者可以在家中进行远程医疗服务,包括在线问诊、远程监测和远程手术等。同时,智能算力还能够为患者提供智慧健康管理和个人健康监测服务,例如通过智能设备实时监测患者的健康状况,并提供个性化的健康建议。

1.2.3.1 案例 1: 智能算力在远程医疗和智慧健康领域的应用

一项远程医疗项目利用智能算力技术实现在线问诊和远程监测。患者可以通过手机或电脑与医生进行实时的视频咨询和诊断,医生可以对患者的病情进行远程监测,并提供相应的治疗建议。这种智能算力的应用可以提高医疗资源的利用效率,减轻患者就医的负担。

通过以上案例,我们可以看到智能算力在医疗产业中的应用和创新模式具有巨大的潜力。从实时数据分析与决策支持到个性化医疗与精准诊疗,再到远程医疗与智慧健康,智能算力为医疗产业的发展带来了新的机遇和挑战。

2. 智能算力对医疗资源整合的影响

医疗资源整合是指通过有效的协调和管理,将医疗机构、医疗设备、医疗人员以及医疗信息等各种医疗资源进行有机整合和优化配置的过程。医疗资源整合的目的是提高医疗服务的效率、质量和可及性,满足人们多样化的医疗需求,降低医疗成本,提升整体医疗体系的综合竞争力。

2.1 医疗资源整合的概念和意义

医疗资源整合是指通过有效的协调和管理,将医疗机构、医疗设备、医疗人员以及医疗信息等各种医疗资源进行有机整合和优化配置的过程。医疗资源整合的目的是提高医疗服务的效率、质量和可及性,满足人们多样化的医疗需求,降低医疗成本,提升整体医疗体系的综合竞争力。

2.2 智能算力在医疗资源整合中的应用

2.2.1 算法和模型在资源整合中的应用

智能算力通过算法和模型的应用,为医疗资源整合提供了新的方法和手段。通过智能算力的技术支持,医疗机构可以进行更准确的资源评估和规划,优化医疗资源的配置和调度,提高医疗服务的效率和质量。

2.2.1.1 案例 1: 智能算力在医疗资源整合中的算法和模型应用

一项研究利用智能算力技术开发了一套医疗资源整合的算法和模型。该算法和模型综合考虑了医疗机构的地理位置、医疗设备的分布和医疗人员的专业能力等因素,并通过智能算力的支持进行快速优化和决策。通过智能算力的分析和计算,该算法和模型能够评估不同医疗资源的利用率和供需情况,进而提供科学依据和建议,使医疗资源的配置更加合理和高效。

例如,在研究中,利用智能算力的支持,研究团队收集了大量的医疗数据,包括医院的就诊记录、医生的诊断结果和患者的医疗需求等。通过对这些数据进行分析和建模,研究团队能够发现医疗资源的利用瓶颈和不平衡问题,并通过智能算力的支持提出相应的改进措施和优化方案。这些算法和模型的应用使得医疗资源的利用更加合理和高效,提高了患者就医的便利性和满意度。

2.2.2.1 案例 1: 智能算力在医疗数据管理和共享平台构建中的应用

一项研究利用智能算力技术构建了一个医疗数据管理和共享平台。该平台通过智能算力的支持,集中存储、管理和共享医疗数据,包括患者的电子病历、医学影像和实验室检验结果等。通过智能算力的分析和处理,该平台能够实现医疗数据的安全存储和隐私保护,同时提供了便捷的数据查询和共享交流渠道。

在该平台中,智能算力的应用不仅可以实现医疗数据的集中管理,还能够进行数据的智能化分析和挖掘。通过智能算力的技术,医疗机构可以更好地利用这些数据资源,进行临床研究、医学决策和疾病预测等方面的工作。此外,该平台还可以促进医疗机构之间的合作和信息交流,提高医疗服务的质量和效率。

这些案例展示了智能算力在医疗资源整合和

数据管理方面的应用和效果。智能算力的应用使得医疗资源的配置更加科学和高效,提高了患者的就医体验和医疗服务的质量。同时,智能算力的技术也为医疗数据的管理和共享提供了新的解决方案,促进了医疗机构之间的合作和信息共享,推动了医疗产业的发展和创新发展。

2.2.2.2 案例 2: 智能算力在医疗资源共享和协同工作中的推动作用

一家医疗机构利用智能算力技术建立了一个医疗资源共享和协同工作平台。通过这个平台,医疗机构可以实时共享医疗资源的信息和状态,包括医疗设备的使用情况、医疗人员的工作安排和患者的就诊需求等。这种智能算力在医疗资源整合中的应用促进了医疗机构之间的合作和协同工作,提高了医疗服务的综合效能。

3. 智能算力在诊断辅助中的应用

3.1 诊断辅助系统的发展和挑战

随着智能算力技术的快速发展,诊断辅助系统在医疗领域得到了广泛应用。诊断辅助系统利用智能算力的高速计算和数据处理能力,能够辅助医生进行疾病诊断和治疗决策。然而,诊断辅助系统的发展还面临着一些挑战,包括数据隐私和安全性、算法的准确性和临床可应用性等方面的问题。因此,如何克服这些挑战并推动诊断辅助系统的应用成为当前的研究重点。

3.2 智能算力在医学影像分析中的应用

医学影像是诊断辅助的重要工具,而智能算力在医学影像分析方面的应用具有广阔的前景。智能算力可以利用深度学习和图像处理等技术,对医学影像进行自动分析和识别,从而帮助医生进行准确的诊断和治疗决策。

3.2.1 案例 1: 智能算力在医学影像分析和辅助诊断中的应用案例

一项研究使用智能算力技术开发了一种基于深度学习的医学影像分析系统。该系统能够自动分析患者的影像数据,识别异常区域并提供定量的测量结果。通过智能算力的支持,该系统能够高效地处理大规模的影像数据,并为医生提供准确的诊断辅助。在临床应用中,这种系统已经被用于肿瘤检测、心脏病诊断和脑部损伤评估等领域,取得了显

著的成效。

3.2.2 案例 2: 智能算力对医学影像分析准确性和效率的提升作用

多项研究表明,智能算力在医学影像分析方面具有超越人类医生的潜力,能够提高诊断的准确性和效率。以下是一些相关研究数据的例子:

一项研究中,研究人员使用智能算力技术对肺部 CT 影像进行肺癌检测。结果显示,智能算力的准确性达到了 90% 以上,相比之下,人类医生的准确性为 80% 左右。这表明智能算力在肺癌检测中具有较高的准确性,并且能够超过人类医生的表现。

另一项研究探索了智能算力在乳腺癌检测中的应用。研究人员使用深度学习算法训练智能算力系统,用于分析乳腺 X 射线和乳腺超声影像。结果显示,智能算力系统的准确性达到了 90% 以上,而人类医生的准确性为 70-80% 左右。这说明智能算力在乳腺癌检测中能够提高诊断的准确性,并且有望成为辅助医生进行乳腺癌筛查和诊断的有力工具。

这些研究结果表明,智能算力在医学影像分析中的应用能够显著提升诊断的准确性和效率。智能算力系统通过深度学习和图像处理等技术,能够自动分析影像数据、识别异常区域,并为医生提供可靠的诊断辅助。然而,尽管智能算力在医学影像分析中取得了令人鼓舞的成果,其在临床实践中的应用仍需进一步研究和验证,以确保其准确性和可靠性。

3.3 智能算力在辅助诊疗中的应用

除了医学影像分析,智能算力还在辅助诊疗方面发挥着重要作用。通过分析患者的临床数据和医疗记录,智能算力可以辅助医生进行疾病诊断、制定治疗方案和预测疗效。

3.3.1 案例 1: 智能算力在辅助诊疗中的应用案例和技术支持

一项研究利用智能算力技术开发了一种辅助诊疗系统。该系统通过分析患者的临床数据和病历记录,结合大数据和机器学习算法,可以为医生提供个性化的诊断和治疗建议。通过智能算力的支持,该系统能够实现对患者的疾病风险评估、治疗方案优化和疗效预测等方面的辅助。

3.3.2 案例 2: 智能算力对辅助诊疗结果的影响

和推动作用

另一项研究调查了智能算力在辅助诊疗中的应用对诊疗结果的影响。研究结果显示,智能算力的应用能够提高辅助诊疗的准确性和效率,减少医疗错误和误诊率,并提升患者的治疗效果和满意度。智能算力的技术支持为辅助诊疗带来了新的可能性,有助于改善医疗服务质量和提升患者的健康水平。

这些案例展示了智能算力在诊断辅助中的应用和效果。智能算力的技术在医学影像分析和辅助诊疗方面具有巨大潜力,能够提高诊断的准确性和效率,改善医疗资源的利用和医疗服务的质量。然而,智能算力的应用还需要进一步研究和验证,以确保其在临床实践中的可靠性和有效性。

4. 智能算力在远程医疗中的推动作用

4.1 远程医疗的定义和优势

远程医疗是通过信息和通信技术将医疗服务延伸到距离患者较远的地点进行诊疗和监护的医疗模式。远程医疗的优势在于克服了时间和空间的限制,为患者提供了便捷的医疗服务。特别是对于居住在偏远地区、无法前往医院就诊或需要长期监护的患者而言,远程医疗具有重要意义。根据统计数据,远程医疗可以为全球超过 1.5 亿偏远地区的患者提供医疗服务,大大缩小了医疗资源的空间分布差异。

4.2 智能算力在远程医疗中的应用场景和技术支持

4.2.1 案例 1: 智能算力在远程医疗中的应用案例和技术支持

智能算力在远程医疗中发挥着关键作用,为多个应用场景提供技术支持。在远程影像诊断方面,智能算力系统能够自动分析和解读医学影像,为医生提供诊断意见和建议。例如,研究表明,智能算力系统在乳腺癌影像的诊断准确性方面超过了人类医生,能够提供更准确的癌症检测和早期诊断。

另一个应用场景是远程健康监护。智能算力技术结合传感器和物联网设备,可以实时监测患者的生理参数,如血压、心率、血氧等。通过智能算力系统的支持,医生可以远程访问这些数据,并进行远程监护和指导。智能算力系统能够根据患者的生

理参数数据进行分析和预测,及时发现异常情况并提供相应的处理建议。

4.3 智能算力对远程医疗数据安全和隐私保护的支持

在远程医疗中,数据安全和隐私保护至关重要。智能算力技术能够提供有效的数据加密和安全传输机制,保护远程医疗数据的隐私性和完整性。智能算力系统利用先进的加密算法和访问控制机制,确保医疗数据在传输和存储过程中的安全性。此外,智能算力系统还可以通过身份验证、访问审计和隐私保护技术等手段,确保医疗数据的合规性和安全性。

综上所述,智能算力在远程医疗中发挥着重要作用,通过提供诊断辅助、远程监护和远程手术等应用,为患者提供了更加便捷和高效的医疗服务。同时,智能算力也为远程医疗数据的安全和隐私保护提供了技术支持,推动了远程医疗的进一步发展和普及。

5. 智能算力与医疗产业融合发展的挑战与机遇

5.1 技术挑战和隐私保护

据的存储、处理和分析变得更加困难。智能算力需要应对大规模数据的挑战,提供高效的数据管理和处理能力。例如,采用分布式计算和存储技术,可以有效处理海量的医疗数据,并保证数据的可靠性和安全性。同时,医疗数据的安全性和隐私保护也是一个重要的问题。智能算力系统需要确保医疗数据的安全传输、存储和访问,保护患者的隐私权益。为了解决这一问题,可以采用加密算法、权限控制机制和隐私保护技术,确保医疗数据的安全性和隐私性。

5.2 人才培养和业务模式创新

智能算力与医疗产业融合发展需要具备相关技术和专业知识的人才支持。医疗领域需要拥有医学知识和智能算力技术的交叉人才,能够理解医疗需求和应用场景,将智能算力技术应用于医疗实践中。为了培养这样的人才,需要加强医学和计算机科学等领域的交叉培训和教育。此外,新的业务模式和商业模式创新也是一个挑战。智能算力为医疗产业带来了新的机会和方式,需要创新的思维和商

业模式来推动产业的发展。为了实现这一目标，可以通过产学研合作、创业孵化等方式，促进医疗产业与智能算力技术的深度融合，探索新的商业模式和创新路径。

5.3 智能算力与医疗产业融合发展的机遇和前景

智能算力与医疗产业的融合发展也带来了巨大的机遇和前景。首先，智能算力可以提高医疗资源的利用效率和质量，优化医疗服务的提供和流程管理。通过智能算力系统的支持，医疗机构可以更好地利用医疗数据，提高医疗决策的科学性和准确性，改善医疗服务的质量和效果。例如，根据统计数据，远程医疗可以为偏远地区的患者提供便捷的医疗服务，大大缩短了患者就医的时间和成本。

其次，智能算力为医疗产业的创新带来了新的机会。通过结合人工智能、大数据和云计算等技术，医疗产业可以开展个性化医疗、精准诊疗、智慧健康等新业务和新模式，满足人们对健康的需求。例如，智能算力在远程医疗中的应用案例展示了智能诊断和治疗方案的优势，提供了更加精准和个性化的医疗服务。

综上所述，智能算力与医疗产业的融合发展面临着技术挑战和隐私保护问题，同时也提供了人才培养和业务模式创新的机会。克服这些挑战，合理应用智能算力技术，将为医疗产业带来更多的机遇和前景，推动医疗服务的创新和升级，提升人们的健康水平和医疗体验。

6. 结论

6.1 主要研究发现总结

通过本研究，我们得出了以下主要研究发现：

首先，智能算力在医疗产业中的应用对医疗资源整合和提升医疗服务质量起到了积极的作用。通过智能算力技术的应用，医疗资源的配置和调度变得更加高效和合理，提高了医疗资源的利用效率和质量。例如，数据挖掘和机器学习算法的应用可以帮助医疗机构更好地预测病情、优化治疗方案，并

为患者提供个性化的医疗服务。

其次，智能算力在医学影像分析和诊断辅助方面展现出巨大的潜力。智能算力系统能够自动分析和解读医学影像数据，提高医学影像分析的准确性和效率。它可以帮助医生快速发现病变、提供辅助诊断建议，有效改善了医学诊断的精确性和及时性。

另外，智能算力在远程医疗方面的应用也带来了诸多优势。远程医疗通过智能算力技术的支持，使得患者可以享受到便捷的医疗服务，尤其是对于偏远地区的患者来说。智能算力系统能够远程监测患者的生理参数、提供远程诊断和治疗指导，实现了医疗资源的跨地域共享和远程协同工作。

6.2 对智能算力在医疗产业中应用的建议和展望

基于以上研究发现，我们对智能算力在医疗产业中的应用提出以下建议和展望：

首先，需要加强技术研发和创新，推动智能算力技术的进一步发展。这包括不断提升算法和模型的精度和效率，加强数据管理和隐私保护的技术支持，开展前沿技术研究，以应对医疗产业的需求。

其次，需要加强人才培养和跨界合作。医疗产业与智能算力的融合需要具备医学和算法技术的交叉人才。因此，我们应加强相关领域的人才培养，培养具备医疗知识和智能算力技术的专业人才。同时，促进医学和计算机科学等学科的跨界交流与合作，形成良好的人才生态系统。

最后，需要加强政策支持和合作共建。政府部门应出台相关政策和规范，引导和支持智能算力技术在医疗产业中的应用和发展。同时，促进医疗机构、高校和企业之间的合作，共同推动智能算力与医疗产业的融合发展。

展望未来，随着智能算力技术的不断进步和应用场景的拓展，我们有理由相信智能算力将在医疗产业中发挥更加重要的作用。通过持续的研究和实践，我们可以进一步优化智能算力的应用模式，提升医疗服务的质量和效率，为人们的健康福祉做出更大的贡献。

基于 AI 算法的无人机巡河目标检测应用

黄炳塔 董帝焱 杜丕加 吴晶晶

(中国联合网络通信有限公司泉州市分公司)

摘要: 随着河长制的推进落实,无人机以其高效率、高安全性的优势被广泛应用于河道巡检。传统工作流程存在人工检测目标效率低、人工标注费力费时、难以批量识别、难以实时监控等缺点,无人机巡河任务的增长与人工效率低下的矛盾日益突出。经过将近半年时间的探索和实践,基于开源框架自主开发出了无人机巡河目标智能检测平台,并在两个区县落地使用,实现对特定目标的智能化识别,显著提高效率。

关键词: 无人机;巡河;智能;目标;检测;YOLO

1 背景

“绿水青山就是金山银山”是时任浙江省委书记习近平于 2005 年 8 月在浙江湖州安吉考察时提出的科学论断,河长制工作责任制就是牢固这一发展理念的重要抓手。随着河长制的推进落实,无人机以其高效率、高安全性的优势被广泛应用于河道巡检。中国联通泉州市分公司与当地相关部门签订的政企项目就实现了运用无人机对辖内两条流域的日常巡检。

2 问题

现有的工作流程主要分为三步:一、工作人员周期性操控无人机,开展流域巡检任务,采集视频资料存储于硬盘等介质;二、人工读取视频资料,以快进或时间段记忆的方式,找到特定目标点如污水排放、河面漂浮物、违规挖掘等,并对其截图保存,标框说明异常情形;三、将整理的截图汇总成文档报告,发送至相关部门。按照现有工作流程,耗时最大的工作是人工对图像资料的查阅、特定目标的标框、标注说明,存在人工检测目标效率低、人工

标注费力、难以批量识别、难以实时监控等缺点。无人机巡河任务的增长与人工效率低的矛盾日益突出,需要一种更加智能的解决方案来提高效率。



图 1 无人机巡河现有工作流程存在的问题

3 解决思路

根据现有的无人机巡河目标检测流程,结合目前目标检测领域的技术发展趋势,针对耗时费力的视频人工查看、目标边框标注、标注说明、目标保存等步骤,寻求通过采用 AI 算法实现目标智能检测,以期大幅提高巡河工作效率,节省时间和人力成本。

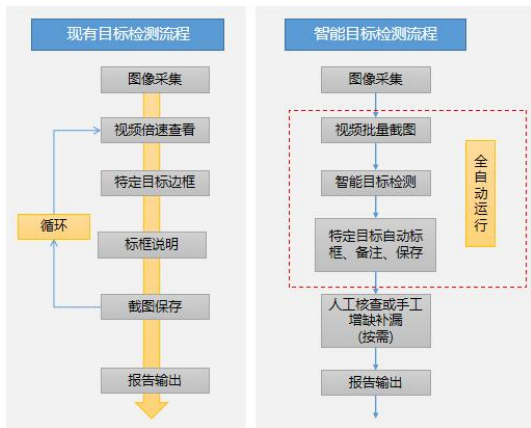


图 2 无人机巡河新旧检测流程对比

4 目标检测算法

目标检测 (Object Detection) 的任务是找出图像中特定的目标, 确定它们的类别和位置, 是很多计算机视觉任务的基础。由于各类物体有不同的外观、形状和姿态, 加上成像时光照、遮挡等因素的干扰, 目标检测一直是计算机视觉领域最具有挑战性的问题。近些年来, 目标检测的算法有了很大的发展。具体可以分为 two-stage 和 one-stage 两大类。对于 two-stage 来说, 这类算法将检测问题划分为两个阶段, 首先通过 CNN (卷积神经网络) 等方法产生候选区域 (Region Proposal), 然后再对候选区域进行分类。典型代表为: Faster-RCNN, 虽然这种算法的错误率低, 但同时漏洞识别率低, 速度慢, 不适合实时监测场景。而对于 one-stage, 这类算法可以仅用一个 CNN 网络来直接产生物体的类别概率和位置坐标值, 且经过单次检测即可以直

接得到最终的检测结果。虽然准确性较 two-stage 低, 但速度快, 代表算法有: SSD, YOLO 等。

本应用采用 YOLO 算法, 它是一种使用深度卷积神经网络学习的特征来检测物体的目标检测器。该算法的核心思想就是利用整张图作为网络的输入, 将目标检测看作一个回归问题来解决, 直接在输出层回归预选框的位置和类别。

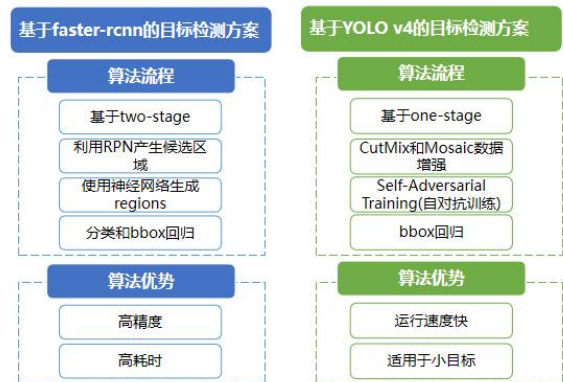


图 3 两种主流算法对比

5 YOLO 算法

YOLO V4 发布于 2020 年 6 月。是在原有 YOLO 目标检测架构的基础上, 采用了近几年 CNN 领域中最优秀的优化策略, 在数据处理、主干网络、网络训练、激活函数、损失函数等各个方面都有着不同程度的优化, 最终取得了性能和速度的提升。通过单个卷积神经网络(one-stage)遍历整个图像, 回归目标的类别和位置, 实现了直接端到端的目标检测。

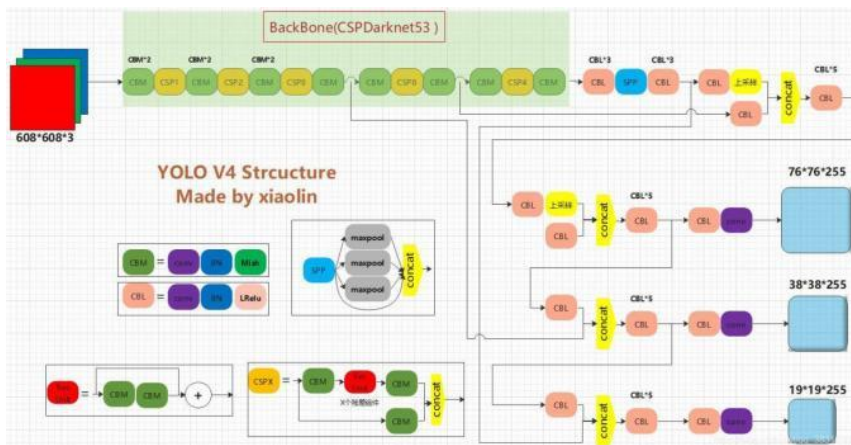


图 4 YOLO V4 网络结构图

6. 模型的训练与调整

训练模型所使用的数据集一共 268 张图片，包含有不同情况下的人、漂浮物、船、排水口等特定目标。其中 200 张用来训练模型，68 张用来验证模型，比例约为 4 : 1。

6.1. 数据集预处理

在图像中均匀选取不同情况下的人、漂浮物、船、排水口等特定目标，采用 labeling 工具用矩形框对图像中的特定目标进行人工标注，用于 YOLO 的训练和测试。标注信息采用 PASCAL VOC 数据集的格式进行保存，其中包含的信息有原图路径、类别、边框位置、文件名等信息。然后进行归一化处理，将目标的实际数据除以图像的宽度和高度，这样得到的数据都在 0 ~ 1 的范围之内，使得在训练的时候能够更快地读取数据，并且能够训练不同尺寸的图像。实际上深度学习需要大量的数据，268 幅图像远远不够使用。因此在训练之前对数据集进行数据增强来增加训练数据，其中色调的变化范围从 1 到 1.5 倍，曝光的变化范围从 1 到 1.5 倍，色量的变化范围从 0.9 到 1.1 倍，以此最终生成了 26800 幅图像以供训练使用。在训练时还需要用 K-means 聚类计算出当前数据集锚点的预训练数值，其中 K-means 使用的是欧氏距离。



图 5 labeling 工具图像手工标注

6.2. 模型的测试与评估

针对模型调优迭代，分别采用全手工操作和全自动操作调优迭代两种方案，利用 VOC 数据进行训练以及模型迭代，获取到第 5 个最佳新模型后，在 VOC 验证集上进行验证，结果如下：

方案	训练时长(h)	TOP-1识别率	Top-5识别率
全手工操作	20.16	90.58%	95.33%
全自动操作	4.67	90.26%	94.56%

表 1 两种模型调优迭代方案对比

全自动操作方案的训练时长是全手工操作的 23%，识别率较全手工操作低，但是相差均在 1% 之间。因此选择全自动操作模型调优迭代方案。

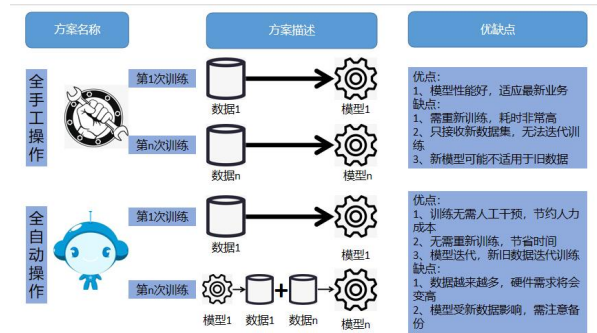


图 6 模型调优迭代方案对比

6.3. 效果验证

利用无人机现场拍摄回传至服务器的视频，采用 python 调用 opencv 库编程实现对视频的批量截图，每隔 3 秒将对应帧保存为 jpg 格式。

```
def save_image(image,addr,num):
    address = addr + str(num) + '.jpg'
    cv2.imwrite(address,image)

path = "/VIDEO_check/VIDEO5/"
video_list = os.listdir(path)
i=1
for mp4 in video_list:
    video_root = path + mp4
    print("截取视频中...")
    video = cv2.VideoCapture(video_root)

    # Find OpenCV version
    (major_ver, minor_ver, subminor_ver) = (cv2.__version__).split('.')

    if int(major_ver) < 3 :
        fps = int(video.get(cv2.cv.CV_CAP_PROP_FPS))
        print ("Frames per second using video.get(cv2.cv.CV_CAP_PROP_FPS): {0}".format(fps))
    else :
        fps = int(video.get(cv2.CAP_PROP_FPS))
        print ("Frames per second using video.get(cv2.CAP_PROP_FPS): {0}".format(fps))

    #读帧
    if video.isOpened():
        success, frame = video.read()
    else:
        success = False

    timeF = fps*3
    j=1
    while success :
        j = j+1
        success, frame = video.read()
        if (j % timeF == 0):
            i = i + 1
            save_image(frame, "/VIDEO_check/VIDEO_2_IMG/",mp4+str(j))
            print("save image: ",j)
            cv2.waitKey(1)
        video.release()
```

图 7 批量截图代码

通过定时执行自适应模型调优模块，按需更新数据集，实现模型的自动优化、迭代。具体如图 8 所示，分为模型库与参数库，确保新模型的最优化。

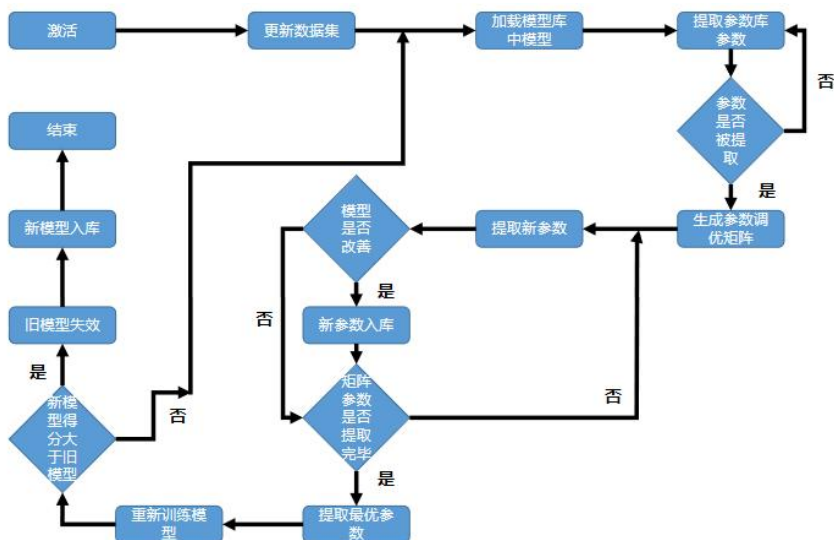


图 8 识别模型自适应优化，自迭代流程图

采用更新数据集的方式，分 10 次执行自适应模型调优模块，结果显示每次都成功选择最佳模型。

序号	新增数据集	调优前后	模型名称	新模型是否更佳	是否最佳模型
1	500	调优前	ep001-loss33.412-val_loss24.182	是	是
		调优后	ep002-loss31.592-val_loss22.767		
2	400	调优前	ep002-loss31.592-val_loss22.767	是	是
		调优后	ep003-loss28.139-val_loss18.972		
3	600	调优前	ep003-loss28.139-val_loss18.972	是	是
		调优后	ep004-loss26.629-val_loss19.959		
4	200	调优前	ep004-loss26.629-val_loss19.959	否	是
		调优后	ep004-loss26.629-val_loss19.959		
5	1000	调优前	ep004-loss26.629-val_loss19.959	是	是
		调优后	ep006-loss24.498-val_loss17.611		
6	700	调优前	ep006-loss24.498-val_loss17.611	是	是
		调优后	ep007-loss23.834-val_loss17.227		
7	600	调优前	ep007-loss23.834-val_loss17.227	是	是
		调优后	ep008-loss22.498-val_loss15.503		
8	300	调优前	ep008-loss22.498-val_loss15.503	否	是
		调优后	ep008-loss22.721-val_loss15.678		
9	1200	调优前	ep008-loss22.721-val_loss15.678	是	是
		调优后	ep010-loss18.629-val_loss12.972		
10	100	调优前	ep010-loss18.629-val_loss12.972	否	是
		调优后	ep010-loss18.629-val_loss12.972		

表 2 识别模型自适应优化，自迭代验证

运用算法模型对无人机巡航时长为 60 分钟的视频进行目标检测，自动截图 1200 张作为数据集做目标检测测试，得到结果如图 9。从结果可看出，训练好的模型能有效识别出特定目标。



图 9 目标检测结果

选取 20 个长度分别为 30 分钟左右的视频进行目标检测，共生成 12381 张截图。经目标检测后，共保存 1469 张图片，人工核查缺漏为 4 张。结果

如图 10 所示。从综合结果来看，YOLO 算法模型

能够胜任不同目标的检测。

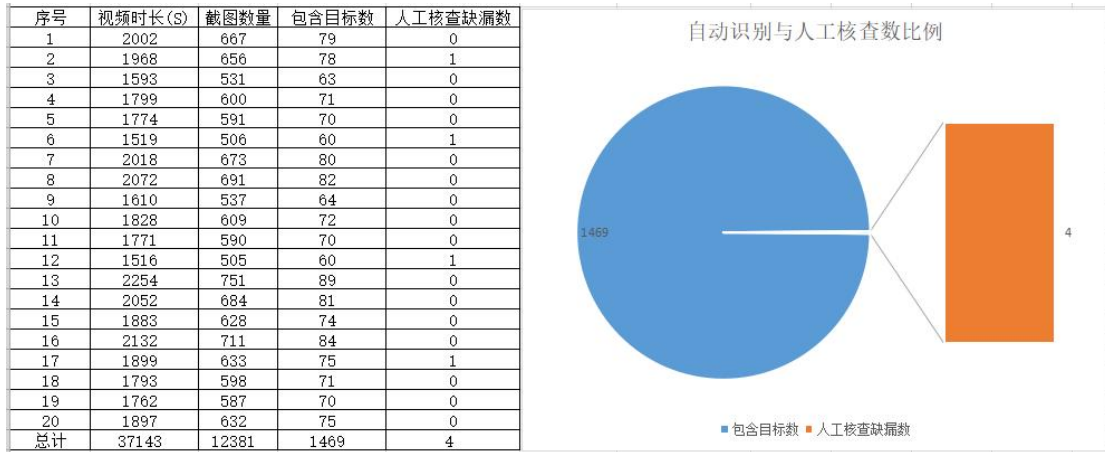


图 10 目标检测应用批量验证

7. 结束语

基于深度神经网络 YOLOv4 算法的特定目标识别方法，在保证效率与准确率兼顾的情况下实现了复杂环境下无人机巡河的特定目标智能检测。在不影响精度的情况下通过压缩网络规模来加快检测速度，在不同拍摄时间、不同河道、不同气候、不同光线等干扰场景下分别进行不同目标的识别，经过训练的模型在验证集下的 mAP(mean average precision)为 87.71%，准确率为 97.02%，召回率为 90.51%，实践表明该模型的检测精度高、速度快、在复杂环境下鲁棒性高。

经过近半年时间的探索和实践，基于开源框架自主开发出的无人机巡河智能目标检测平台，能够实现对特定目标的智能化识别，60 分钟视频目标检测工作时长由 50 分钟缩短至 5 分钟，效率提升 10 倍。该平台已应用于鲤城区和永春县的无人机河道巡检，显著降低人工劳动强度，提高工作效率。

参考文献：

[1] [美]Alexander T.Combs. Python 机器学习实

践指南 [M]. 北京：人民邮电出版社，2017.5：187-206.

[2] [德]安德里亚斯·穆勒，[美]莎拉·吉多，张亮译. Python 机器学习基础教程 [M]. 北京：人民邮电出版社，2018.1：21-22.

[3] [爱尔兰]米洛万诺维奇，颀清山译. Python 数据可视化编程实战 [M]. 北京：人民邮电出版社，2015.5：66-87.

[4] [美]埃里克·马瑟斯. Python 编程 从入门到实践 [M]. 北京：人民邮电出版社，2016.7：44-56.

[5] 董洪义. 深度学习之 PyTorch 物体检测实战 [M]. 北京：机械工业出版社，2020.1：167-183.

[6] [挪]芒努斯·利·海特兰德. Python 基础教程 [M]. 3 版. 北京：人民邮电出版社，2018.2：192-194.

[7] [美]比斯利，[美]琼斯. Python Cookbook [M]. 3 版. 北京：人民邮电出版社，2015.5：142-150.

5G 全连接工厂关键技术的研究与实践

陈锋 吴帆 洪林梦涵

(中国联合网络通信有限公司福州市分公司)

摘要: 本文主要介绍围绕三大 5G 全连接工厂关键技术“Massive MIMO、5G 网络切片、5G 专线”的研究与实践,并应用于长乐凯邦锦纶科技“5G+工业互联网”智慧工厂实例中。凯邦锦纶科技智慧工厂项目依托上述 5G 技术,打造 5 大业务系统,构建 3 大基础平台,创建 2 大智能应用,建立 7 大应用场景,打造全连接工厂。

关键词: Massive MIMO; 5G 切片; 5G 专线

1. 概述

近年来随着我国工业智能制造发展,越来越多的车间设备,如各类传感器、控制器、监控摄像头、机床控制臂、AGV 小车、各类机器人等开始接入工厂内网,尤其是 AGV 等移动设备的通信,有线网络难以满足,对工厂内网的灵活性和带宽要求越来越高。传统工厂有线网络可靠性带宽高,但是灵活性差、管理困难,无线网络灵活性较高,但是可靠性,覆盖范围,接入数量等都存在不足,如 WIFI 网络易受干扰、Lora 带宽不足。兼具灵活性、高带宽和多终端接入特点的 5G,成为承载工厂内设备接入和通信的新选择。同时 5G 新技术赋能垂直行业各项应用,给 5G 带来无限商机,也推动国家新基建布局中的 5G 核心技术良性发展。本文主要介绍如何通过 5G 技术打造凯长乐邦锦纶科技“5G+工业互联网”智慧工厂的实例。

2. 凯邦锦纶科技“5G+工业互联网”智慧工厂简述

长乐凯邦锦纶科技有限公司是福建省纺织业集群地的龙头企业,是福州联通重要战略合作伙伴。面对政府规划和企业需求,福州联通联合长乐凯邦锦纶科技有限公司共同打造工业互联网项目标杆。

凯邦锦纶科技“5G+工业互联网”智慧工厂以 5G 技术为核心,建设 5G 虚拟专网,实现 5G 全工厂连接,提供“专网+平台+终端+应用”的 5G 一体化解决方案。建成了“1+4”的新型 5G+工业互联网虚拟专网,其中,“1”为 5G 虚拟专网,“4”为 5G 虚拟专网中运用的 5G 分布式 Massive MIMO、5G 网络切片、5G 边缘计算、5G 专线等核心技术。5G 虚拟专网适应众多工厂智能设备不同级别的高带宽、大容量及低时延需求,并能确保 5G 专网专用。

作者简介:

陈锋: 高级工程师,现任职于中国联合网络通信有限公司福州市分公司云网中心,云网专家,从事无线网络规划建设优化工作

吴帆: 现任职于中国联合网络通信有限公司福州市分公司云网中心,网络优化工程师,从事无线网络优化工作

洪林梦涵: 现任职于中国联合网络通信有限公司福州市分公司云网中心,网络优化工程师,从事无线网络优化工作

3.采用 5G 分布式 Massive MIMO 技术，实现“5G+工业互联网”高带宽、大容量接入。

3.1 基于 5G Beam Sweeping 的广播信道波束赋型，实现场景化的精准覆盖

长乐凯邦锦纶科技公司占地面积大，楼宇类型多样、场景丰富，其工业互联网设备在众多场景下均有部署，对5G网络覆盖要求较高。5G在设计阶段就引入了天线新特性，其中基于 Beam Sweeping 的广播信道波束赋型技术，由多个窄波瓣波束轮发，形成宽波束覆盖效果，并支持多种广播波束权值配置，生成不同组合的赋型波束，满足不同场景的覆盖要求。[1]



图 1. 凯邦厂房建筑环境实景

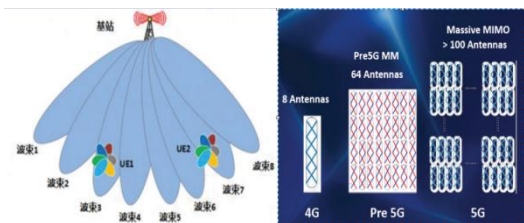


图 2. 5G 波束扫描和波束权值示意

SSB 波束可以根据不同的场景配置不同的波束覆盖方案，目前主流的波束覆盖方案如下：

序号	水平波宽	垂直波宽	Azimuth	Digital Beam tilt	应用场景
1	65°	6°	[-20 +20]	[-3.15]	标准宏覆盖，低层覆盖需求
2	65°	12°	[-20 +20]	[0.12]	宏覆盖，中层楼宇
3	65°	25°	[-20 +20]	6	宏覆盖，高层楼宇
4	45°	6°	[-30 +30]	[-3.15]	热点场景，低层楼宇，干扰场景
5	45°	12°	[-30 +30]	[0.12]	热点场景，中层楼宇，干扰场景
6	45°	25°	[-30 +30]	6	热点场景，高层楼宇，干扰场景
7	30°	6°	[-40 +40]	[-3.15]	热点小范围场景，低层楼宇
8	30°	12°	[-40 +40]	[0.12]	热点小范围场景，中层楼宇
9	30°	25°	[-40 +40]	6	热点小范围场景，高层楼宇
10	15°	12°	[-45 +45]	[0.12]	独特热点场景，中层楼宇
11	15°	25°	[-45 +45]	6	独特热点场景，高层楼宇
12	90°	6°	/	[-3.15]	水平覆盖要求较高场景，广场及宽大的低层建筑
13	90°	12°	/	[0.12]	广覆盖立体覆盖，广场及宽大的中层建筑
14	90°	25°	/	6	水平和垂直覆盖要求较高场景，广场及宽大的高层建筑

图 3. SSB 主流波束覆盖方案

经过现场实地查勘，将工业园场景划分为以下三类主要场景：

(1) 低层建筑覆盖场景 (以生产车间为代表)，楼层高度为5-10米之间，楼间距约为20米。该种场景5G基站覆盖范围主要保证水平宏覆盖。

(2) 中低层混合场景 (以研发楼为代表)，主要以中低层 (小于30米) 建筑为主，需要兼顾高低建筑物的水平和垂直覆盖。

(3) 高层建筑场景 (以办公自动化大楼为代表)，需要保证高层 (大于30米) 建筑室内垂直覆盖需求。

对于不同场景，采用网规仿真规划工具 (Smart Hippo&ACP) 分别输出天线权值规划建议，并通过现场测试确定不同场景权值设置建议。各场景的覆盖对比结果如下：

表 1. 不同覆盖场景的波束规划方案效果比较

场景	不同方案对比	覆盖提升的测试点位比例	质量提升的测试点位比例	下行速率提升的测试点位比例
低层覆盖	7波束(smart hippo规划) VS 单波束的H65V6方案	100.00%	90.91%	81.82%
	7波束(ACP规划) VS 单波束的H65V6方案	90.91%	90.91%	90.91%
	7波束(ACP规划) VS 7波束(smart hippo规划)	36.36%	72.73%	54.55%
中低层混合覆盖	7波束(smart hippo规划) VS 单波束的H65V6方案	100.00%	84.62%	84.62%
	7波束(ACP规划) VS 单波束的H65V6方案	100.00%	61.54%	84.62%
	7波束(ACP规划) VS 7波束(smart hippo规划)	46.15%	53.85%	61.54%
高层覆盖	7波束(H65V25) VS 单波束的H65V6方案	100.00%	100.00%	92.86%
	7波束(H45V25) VS 单波束的H65V6方案	100.00%	100.00%	78.57%
	7波束(H45V25) VS 7波束(H65V25)	78.57%	64.29%	57.14%

结合本次分场景的方案测试结果，针对长乐凯邦锦纶科技产业园 5G 无线覆盖采用以下的覆盖方案：

- (1) 水平宏覆盖场景使用 7 波束的 H65V6 方案进行初始配置。
- (2) 中低层混合场景建议使用 7 波束的 H45V12 方案进行初始配置。
- (3) 高层建筑场景建议使用 7 波束的 H45V25 方案进行初始配置。

通过本次基于 Beam Sweeping 的广播信道波束赋型的优化研究，长乐凯邦锦纶科技产业园 5G 覆盖率达到 99.23%，平均信噪比达到 15db，其中生产区域 5G 覆盖率达到 100%，平均信噪比 20db 以上，可以满足日常工业化生产需求。

3.2 采用分布式 Massive MIMO 技术，提升 5G 专网的上下行容量

工厂车间中的工业设备数以千计，全连接工厂需要解决的一个关键问题就是需要提供高带宽、大容量接入的网络能力。传统组网中，解决大容量场景一般通过小区分裂的方式，但由于无线信号的传播特性，信号传输会经历各种复杂的传播路径，有用信号会受到各种噪声的污染，因此 NR 小区分裂后，可能出现区域内容量没有增加甚至降低的情况。

针对该问题，5G 室内组网部署中通过 5G 分布式 Massive MIMO 小区功能，将工作在相同频段上的射频模块所覆盖的 n 个连续覆盖的 2T2R 小区合并为一个 2nT2nR 小区来消除小区边界，降低小区间下行干扰，同时合并后形成的 Massive MIMO 小区可以通过 MU-MIMO 功能来提升系统的上行容量，最终实现大容量接入。

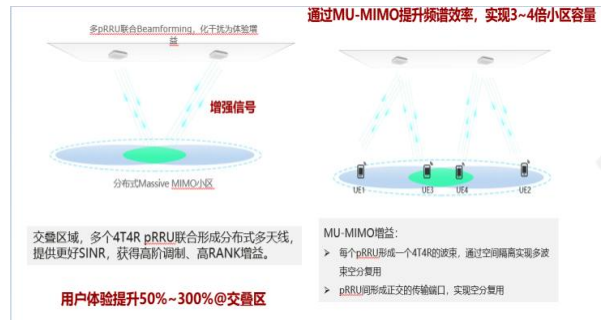
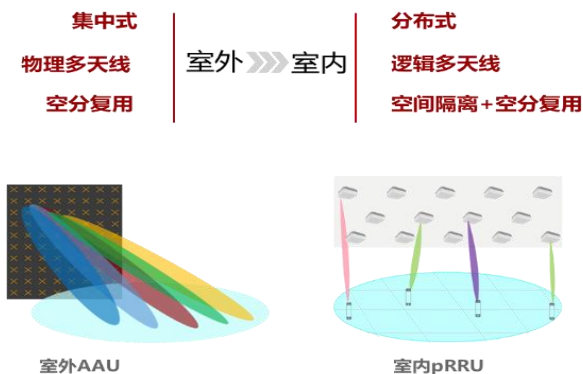


图 4. 5G 分布式 Massive MIMO 小区功能

启用分布式 Massive MIMO 功能后，基站通过测量 UE 在多个天线通道下的信号差值来识别交叠区用户。当 UE 处于交叠区，多个天线通道对该交叠区用户的业务信道数据进行联合传输，提升 UE 吞吐率。对于非交叠区用户，多个天线通道的波束可以空分复用同一个 PUSCH 或 PDSCH 时频资源，提升上行容量。

为验证该功能有效性，现场安排 4 组人员，分布在车间不同点位上，进行性能对比测试，如下图：



图 5. 车间不同测试点位

下载	整改后		整改前		增长倍数
	均值 (Mbps)	峰值 (Gbps)	均值 (Mbps)	峰值 (Mbps)	
点位1	833.14	1.01	183.58	513.27	4.82
点位2	1107.65	1.31	208.23	553.87	5.32
点位3	925.2	1.18	190.16	273.85	4.86
点位4	868.04	1.13	192.12	406.72	4.52
均值	933.51	-	193.52	-	4.82

上传	整改后		整改前		增长倍数
	均值 (Mbps)	峰值 (Mbps)	均值 (Mbps)	峰值 (Mbps)	
点位1	182.87	199.91	63.88	105.04	2.88
点位2	230.04	288.42	67.76	116.95	3.43
点位3	237.98	349.36	77.04	170.23	3.07
点位4	186.81	261.67	48.51	93.91	3.87
均值	209.43	-	64.29	-	3.26

图 6. 车间不同测试点位实测性能对比

采用 5G 分布式 Massive MIMO 技术后，从现场实测对比来看，设备下行吞吐率均值为 933Mbps，峰

值达到 1.3Gbps，较整改前提升 4.82 倍；上行吞吐率均值为 209Mbps，峰值达到 349Mbps，较整改前的提升 3.26 倍，有效提升了厂区的网络容量。

4. 以 5G 端到端切片技术为核心，建设凯邦锦纶科技智慧工厂“5G+工业互联虚拟专网”

长乐凯邦锦纶科技项目采用基于无线网、核心网、承载网的端到端网络切片实现 5G+工业互联专网。无线网运用 GBR、PMBR、调度优先级三种无线切片方案，核心网采用 UPF 下沉实现基于业务需求的切片全生命周期管理，承载网通过 DSCP 实现传输差异化调度，满足 5G 专网设备差异化场景的业务需求。[2]

4.1 无线网通过基于 5QI/QCI 切片，满足智慧工厂多样化的业务场景

无线侧通过不同的 QCI/5QI 映射实现无线资源切片，通过三类解决方案：PMBR、GBR 和调度优先级，实现差异化的速率保障。保障方案具体如下：

表 2. PMBR、GBR 和调度优先级差异化的保障方案

切片类型	保障方案	优点	缺点	推荐方案优先级
无线切片	PMBR 保障方案	基站可灵活设置保证速率，使设备速率稳定在设置值附近	无法确保最小保障速率，且优先级低于 GBR	高
	GBR 保障方案	可设置最小保障速率，且优先级最高	由核心网分配专用 QCI，基站侧无法灵活设置，易造成资源浪费	中
	调度优先级差异化保障	可更改加权因子优先调度	基站仅能根据加权因子优先调度，无保证速率	低

PMBR、GBR、调度优先级切片方案优先级验证：

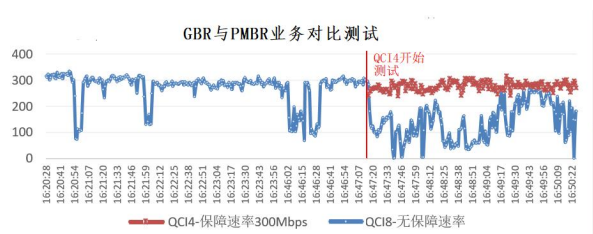


图 7. PMBR、GBR、调度优先级切片方案优先级验证

当只有 PMBR（QC18）业务时，设备速率基本维持在 300Mbps，当 GBR 业务（QC14）接入时，开始抢占 PMBR 资源，PMBR 设备下

载速率远低于 300Mbps。

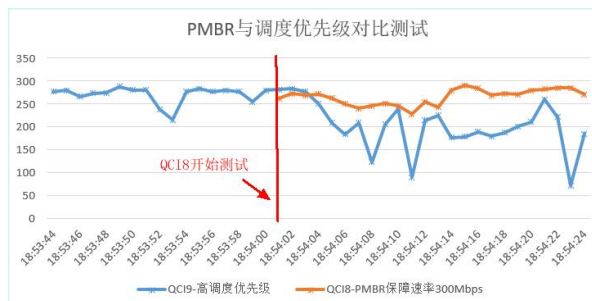


图 8. PMBR 与调度优先级对比测试

设置加权因子的 QC19 进行业务，下载速率在 300Mbps 左右，当设置 PMBR（QC18）设备进行业务时，QC19 业务速率下降较明显。即 PMBR 保障方案的优先级高于调度优先级保障方案。

表 3. 调度加权因子对用户速率的影响

下行调度加权因子		下载速率(Mbps)		专线/普通用户速率比
专网设备(QCI=8)	普通用户(QCI=9)	专网设备(QCI=8)	普通用户(QCI=9)	
1000	800	349	301	86.25%
1000	600	371	228	61.46%
上行调度加权因子		上传速率(Mbps)		专线/普通用户速率比
专网设备(QCI=8)	普通用户(QCI=9)	专网设备(QCI=8)	普通用户(QCI=9)	
1000	800	92	76	82.61%
1000	600	95	64	67.37%

专网设备按加权因子提升调度优先级，专网设备与普通用户的速率比例基本与设置的加权因子相同。该方案仅提升调度优先级，无最低保证速率。

无线切片实验证明：GBR 优先级最高，当带宽不足时 GBR 承载设备则会抢占 PMBR 和调度优先保障的设备带宽，PMBR 设备资源不足时可以抢占非调度优先级保障的资源，调度优先级保障设备仅可以抢占普通用户的资源。

根据上述切片优先级验证，结合厂区内设备带宽需求，厂区控制中心采用 GBR 方案确保和厂区内千台设备的数据互联，厂区监控采用 PMBR 方案确保高清视频及时回传至指挥部备份，厂区内办公及其他速率业务则可采用调度优先级进行保障。

4.2 核心网基于 UPF 下沉组网提供智慧工厂低时延高可靠解决方案

传统互联网安全更多关注网络设施、信息系统软硬件及应用数据安全，工业互联网安全

扩展延伸至企业内部，包含设备安全、控制安全、网络安全、应用安全及数据安全。为了实现上述需求，5G 核心网可通过 SMF 灵活的会话管理机制，将 UPF 下沉，采用 DNN 分流技术，实现本地流量疏导，并提供可靠的移动性解决方案。[3]

DNN 分流方案介绍：网络侧 5GC SMF 基于用户所签约的 DNN 为用户选择特定 UPF，基站基于不同 DNN 面向不同的 UPF 建立 N3 接口，从而实现特定 DNN 流量由特定的 UPF 疏通。

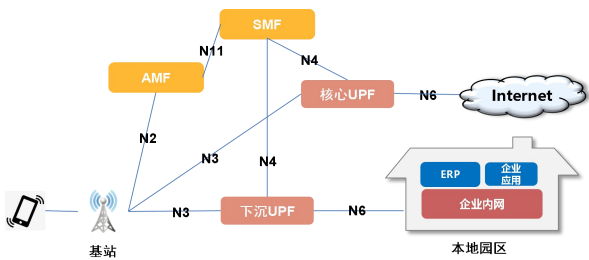


图 9. DNN 分流方案

同时为了保证园区内移动设备的业务连续性，5G 网络提供三种不同的 SSC 模式（SSC Mode1、SSC Mode2、SSC Mode3），如下图所示：

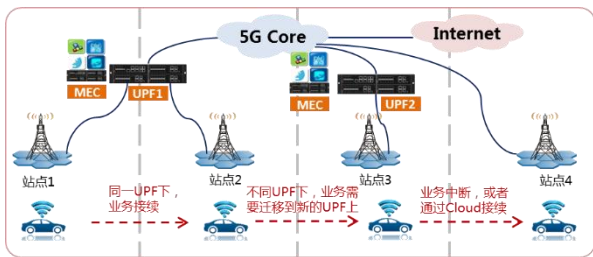


图 10. SSC 的三种模式

根据网络配置 SSC 模式选择策略，终端可以作为一个应用或者一组应用选择合适的 SSC 模式，以支撑边缘计算业务连续性。结合长乐凯邦产业园区的业务需求最终配置 SSC Mode1 与 SSC Mode2 结合的模式：当设备在园区内（即 UPF 覆盖区域内）移动，5G 核心网采用业务连续性 SSC Mode1，保持本地分流业务的连续性；当设备移动到园区外（MEC 覆盖区域外），5G 核心网采用业务连续性 SSC Mode2，业务中断或者通过 Cloud 接续。

本项目在长乐凯邦部署 1 套 MEC UPF，通过智能引擎分布到网络边缘，实现业务本地智能处理，以 5G+MEC 来替代厂区内有线或无线 WiFi 网络，达到高带宽、低时延要求。

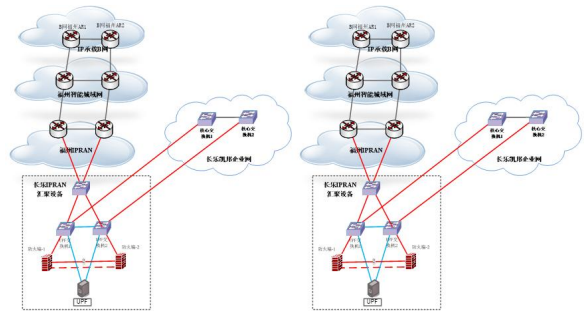


图 11. 厂区 5GMEC 组网方案

通过 UPF 下沉，将本地业务的数据直接分流到凯邦科技内部网络，避免了流量在核心网的迂回，减少了业务传输时延，并将机器视觉、低速移动驾驶等功能下沉到 MEC 边缘计算平台，提升计算效率，业务数据不出园区，安全可控。同时采用 UPF 下沉组网方案后，产区内设备时延降低至 10ms 以内，满足工厂日常低时延控制生产要求。

4.3 基于 DSCP 的承载网切片保障智慧厂区业务差异化调度

传输下沉光端机（利旧设备）至无线基站上，并双路由上连至长乐下沙机房和长乐东湖 VR 机房。厂区内传输路由按无线基站及 UPF 部署需求进行光缆布放建设，同时采用 5QI 来区分切片，承载网通过 5QI 映射到 DSCP 实现不同优先级差异化调度。组网拓扑图如下：

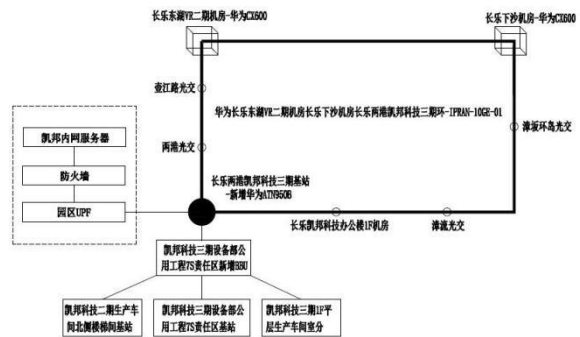


图 12. 厂区内传输路由的组网拓扑

基于 DSCP 实现差异化调度：当设备传输数据超出自己的最大负荷时，部分服务将被丢弃，但在某些时候必须让网络通过放弃传输相对不重要的数据来保证重要用户，因此，就需在网络中实施 Quality of Service (即 QoS)，通过 QoS 可以为特定业务提供保证带宽，DSCP 就是给网络中的流设置优先级的方式，DSCP 设置取值越大（优先级越高）。[4]

DSCP 切片验证：

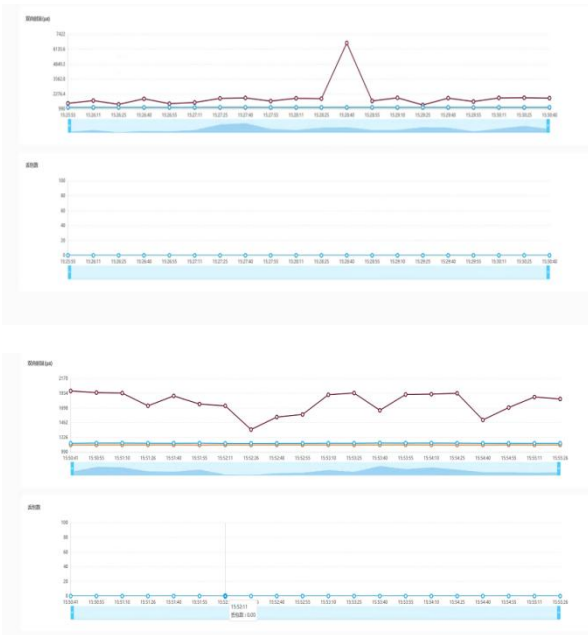


图 13. 基于 DSCP 实现差异化调度验证

相同优先级下：不切片场景下，专网设备业务时延高，有抖动，无丢包；切片场景下，专网设备业务时延低，无抖动，无丢包。

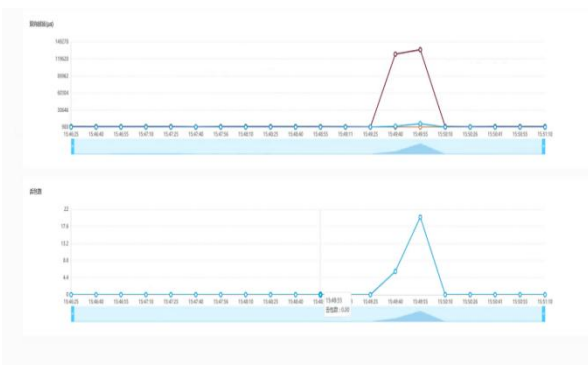


图 14. 相同优先级调度验证

不同优先级下：不切片场景下，专网设备业务时延高，抖动大，有丢包；切片场景下，专网设备业务时延低，无抖动，无丢包。

验证结果：相同优先级情况，在流量突发场景下，通过 DSCP 切片能很好的隔离突发，一个切片内的突发对其他切片无影响；不同优先级情况，当高优先级重载突发时候，对其他切片的低优先级业务无影响。

厂区采用 DSCP 切片组网实现了上千设备同时工作的重载场景下，厂区办公业务、视频回传业务等大带宽业务保持上行速率 336Mbps，下行速率 1.16 Gbps。

创新点三：构建 5G 专网设备级风险管控平台，确保 5G 专网专用。

工业互联网场景中部分设备存在移动性，为确保 5G 网络在厂区内专网专用，项目构建 5G 专网设备级风险管控平台，实现 5G 专网设备在指定区域实现专网专用，5G 专网终端使用风险管控平台相较传统的机卡绑定技术安全性提升 40%。

5.构建 5G 专网设备级风险管控平台，确保 5G 专网专用

工业互联网场景中部分设备存在移动性，为确保 5G 网络在厂区内专网专用，项目构建 5G 专网设备级风险管控平台，实现 5G 专网设备在指定区域实现专网专用，5G 专网终端使用风险管控平台相较传统的机卡绑定技术安全性提升 40%。

5.1 研发 5G 专网机卡分离管控技术

通过用户使用过程中平台上报的 IMEI 号进行监控，将用户开户登记的 IMEI 号与之匹配，判定用户是否有过机卡分离行为。针对机卡分离情节严重的，后台可限制其使用。

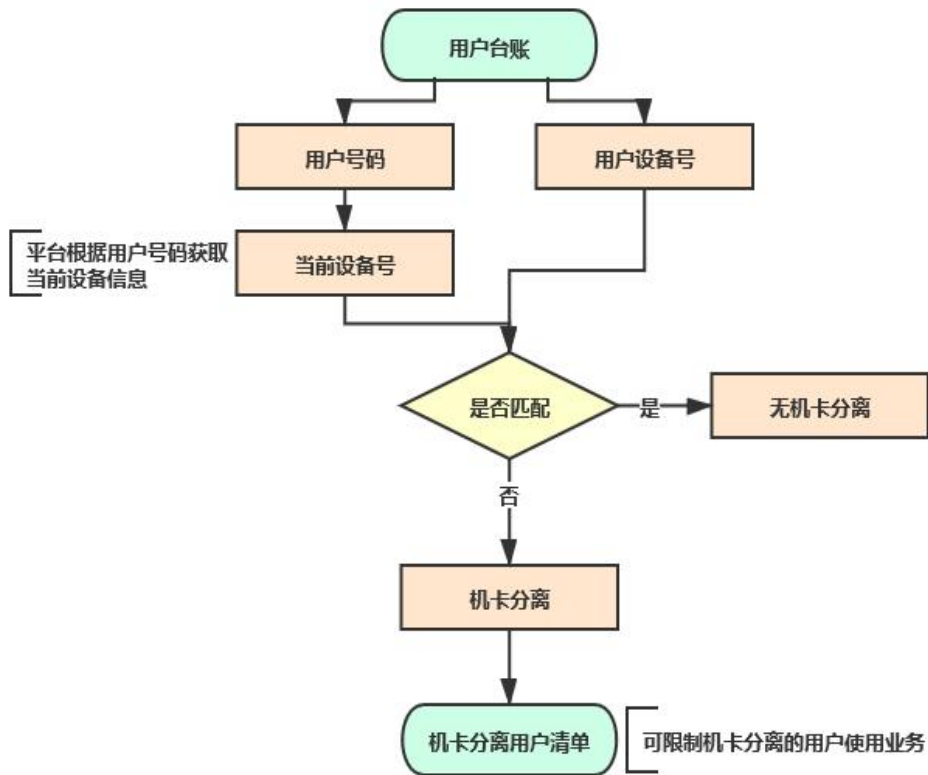


图 15. 机卡分离用户监控流程

上述流程主要通过 PCRF 网元来实现限制：设备开户时在 PCRF 静态设置设备的 IMSI 和 IMEI，并配置相应的策略。设备入网或者发生位置更新，UGW 上报终端的 IMEI，并由 PCRF 来检查 IMEI 与预设的策略是否匹配，如果不匹配则执行 block，业务无法使用。

清单；即专网设备可将产业园区内所有 5G 专网小区记录在设备的小区列表里，当专网设备后续注册的小区不在此列表内，则不允许访问网络。

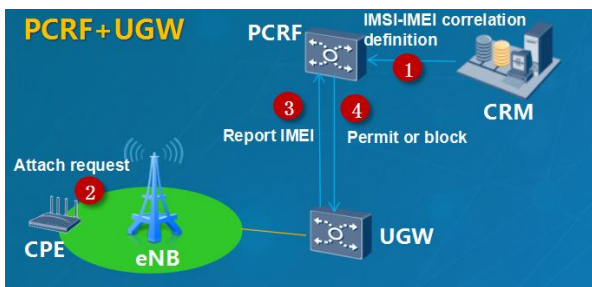


图 16. 基于 PCRF 网元的锁卡流程

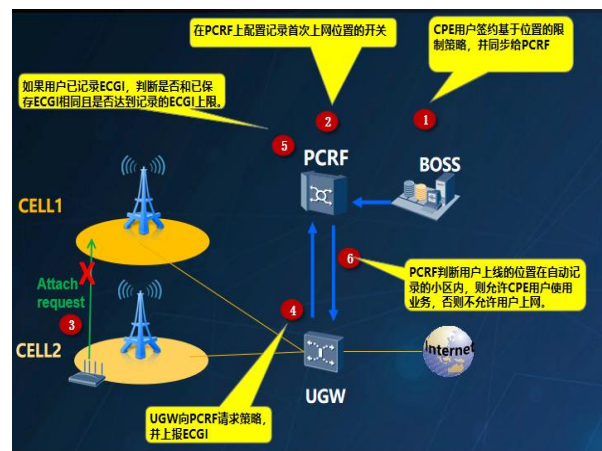


图 17. 基于累计驻留小区数量锁设备流程

5.2 研发 5G 专网基站锁定技术

专网设备在长乐凯邦产业园区开机注册后，PCRF 会开始记录设备注册过的小区，并形成小区

6. 结束语

本文主要介绍了 5G 三大技术“Massive MIMO、

5G 网络切片、5G 专线”在凯邦锦纶科技“5G+工业互联网”智慧工厂是如何应用的,研究成果在福州联通进行落地试点,并进行了全省推广。其中 5G 核心关键技术在前期的“5G+智慧港务建设”项目、“海心一号 5G 智慧医疗”项目、“5G+远程会诊/远程医疗教学方案”都进行了大量的推广和应用。

参考文献:

[4] QI C ,DONG P ,MA W , et al.Acquisition of channel state information for mmWave massive MIMO: traditional and machine learning-based

approaches[J].Science China(Information Sciences),2021,64(08):5-20.

[5] 刘惜吾,邢剑卿,王笃炎等.基于车联网的网络切片 GBR 动态专载触发方案 [J].移动通信,2022,46(08):60-66.

[6] 文旭桦,鄢欢,肖荣奋等.基于边缘 UPF 下沉的 5G 定制网在电子制造业的应用 [J].广东通信技术,2022,42(06):2-5+15.

[7] 方宇,李若朋,孙颖等.5G 切片技术在 5G toB 业务中的应用研究 [J].数字通信世界,2021(10):55-57.

云网融合算力基础设施应用与发展

陈邦文

(中国电信股份有限公司福建分公司)

摘要: 本文首先介绍算力的发展历程及趋势, 然后结合中国电信福建公司的自身实践, 对云网融合的算力基础设施体系规划布局以及算力赋能行业应用情况进行阐述, 并对智能算力的发展及其风险进行探讨。

关键词: 云网融合; AIGC; 智能算力; 行业应用;

引言

数字经济时代, 算力已经成为推动社会进步和经济发展的重要生产要素和基础设施。以云计算为技术底座的算力基础设施在通信、金融、医疗、教育、社会治理到智能制造等各个领域都有广泛应用, 为推动产业数字化发展、社会生产力变革方面发挥重要作用。近几年来, 随着人工智能技术的发展, 算力基础设施也开始向智能化方向不断发展, AIGC 成为时下最热门的话题之一。本文首先介绍算力的发展历程及趋势, 然后结合中国电信福建公司自身实践, 对云网融合的算力基础设施体系规划布局以及赋能行业应用情况进行阐述, 并对智能算力的发展进行探讨, 最后总结并展望。

1. 算力发展历程及趋势

1.1 算力发展历程

云计算是算力发展的基础技术底座之一。云计算是一种基于互联网的计算方式, 通过共享硬件和软件资源, 为用户提供随时随地的数据存储和计算服务。云计算的特点包括弹性扩展、按需使用、随时随地连接等。云计算的发展历程可追溯到 20 世纪 90 年代, 当时互联网技术正在逐步普及, 网站和网络应用开始大量出现。为了解决这些网络服务的存储和计算问题, 人们开始探索将计算机资源集中起来, 提供按需使用的计算服务。这个阶段被称为云计算的萌芽阶段。到了 21 世纪初, 一些企业开始提供云计算服务, 如国际上, 亚马逊的 AWS

和 IBM 的蓝云等, 这些服务提供了基本的计算、存储和网络服务, 但还没有形成大规模的商业市场。随着时间的推移, 云计算技术不断发展和普及, 云计算开始形成大规模的商业市场, 越来越多的企业开始进入这个领域。如国外微软的 Azure 和谷歌的云平台, 国内的阿里云、腾讯云、华为云、电信天翼云等, 这推动了云计算进入快速发展阶段。随着云计算技术的不断成熟和发展, 越来越多的行业开始应用云计算技术, 企业的传统 IT 进行上云转型, 云计算成为了企业数字化转型的重要支撑, 这是云计算的行业应用阶段。随着人工智能技术的不断发展, 云计算技术也开始向智能化方向发展, 如智能云、智能客服等, 以 ChatGPT 为首的生成式人工智能更是点燃了智能算力发展的浪潮。云计算技术开始提供更加智能化、个性化的服务, 这是云计算的智能化发展阶段。

除了虚拟化, 算力的应用领域还包括大数据、人工智能等。其中 AI 算力是人工智能发展的重要支撑, 它是指计算机能够进行人工智能计算的能力。随着人工智能技术的不断进步, AI 算力的需求也在不断增加。AI 算力的应用领域包括图像识别(人脸识别、车牌识别、动作识别、物品检测、周界检测等)、自然语言处理(机器翻译、语音识别、文本生成、情感分析等)、智能推荐、辅助驾驶、趋势预测等等, 其中图像识别和自然语言处理是 AI 算力应用的主要领域之一。通过 AI 算力, 企业可以快速实现人工智能的计算和处理, 从而提高企业的创新能力和竞争力。

1.2 算力发展趋势

算力的战略价值在全球范围内持续提升。美国、欧洲和亚洲等主要国家均发布了国家战略或计划，旨在推动云计算在各行各业的应用布局，并挖掘算力的产业赋能价值。我国政策指引云计算应用创新，持续推动云计算与实体经济的融合。算力的技术也在不断推陈出新，助力高质量发展。算力的发展趋势主要有以下几点：

算力形式更加多样，算力资源泛在分布。算力形式正在变得更加多样化，包括人工智能、云计算、大数据等。这些不同的算力形式使得算力资源更加普及化。高速率、低时延、泛在接入的应用需求也使得分布式算力成为基本要求。

基础设施建设加速，算网融合持续深入。“东数西算”工程的推进，应用的云边互联、边边协同的需求，驱动算力和网络之间的深度融合，云网融合、算网一体的服务需求开始产生，这种服务将算力和网络服务整合在一起，为不同领域和行业提供更加便捷、高效、可靠的云网算力服务。

创新应用竞相涌现，产业生态日渐多样。基于算力的创新应用正竞相涌现，以 ChatGPT 为代表的 AIGC 大模型应用被誉为是一场工业革命，结合行业应用要素融合互促，算力将成为多技术融合、多邻域协同的重要载体。

2. 云网融合的算力基础设施体系

早在 2016 年，中国电信率先在业内提出云网融合。并按照“网是基础、云为核心、网随云动、云网一体”的思路不断探索实践，并不断迭代升级。随着 5G、物联网和工业互联网等技术的发展，海量边缘数据爆发式增长，以云计算为核心的算力时代，在网络延迟、隐私安全和能效等方面，已无法满足边缘数据处理需求，因此有必要构建云网融合的底层基础设施作为承载基础。本章将结合中国电信福建公司的具体实践，从承载算力的数据中心、多云共生的算力体系和连接算力的基础网络等三个方面阐述云网融合算力基础设施体系。

2.1 承载算力的数据中心

数据中心(IDC)是承载算力资源的基础设施载体，其布局和规划将深刻影响算力资源的布局。中国电信福建公司坚持数据驱动和效益导向，在协同集团做好枢纽节点区域布局的同时，立足省内业务，

以市场需求为导向，配合政策指引和技术要求等，统筹考虑省内数据中心的中长期科学布局，做好重点区域的资源储备，按照全省“3+9+X”架构，构建“中心集群+浅边缘+深边缘”的层次化云网融合基础设施能力。

由福州东南信息园和厦门海峡数据中心、福州仓山科技园 IDC 三大园区 IDC，作为全省性算力业务主要载体。结合省内网络重构后的通信机房建设，将 IDC 基础资源拓展到全省 9 地市。每个地市至少 1 个星级数据中心。以县区边缘为单位，建设小心 X 节点，作为 IDC 的边缘化补充，满足边缘算力的承载需求。

此外，围绕园区节点打造数据中心集群，优化双 AZ 资源布局，打造同城双活、省内两地三中心的高可用算力基础设施，为各类政企业务上云提供更安全、更优质的服务。满足多种解决方案和多种保障交付需求。



图 3.1-2 福建双 AZ 网络基础设施布局图

2.2 多云共生的算力体系

协同数据中心布局规划，加快从云计算为主向基于算力的多数字要素协同布局转变，统筹算力、云、IDC/DC、大数据、AI、安全、平台、绿色等云网要素的布局，满足云网融合 3.0 发展要求。以福州、厦门为省级双节点，完成“2+9+X”省市县三级梯次化算力体系布局。

云网统筹布局规划，采用基于中国电信自研云产品，建设统一架构、统一承载的天翼云算力体系，承载对外服务客户以及中国电信内部业务上云需求，其中智能算力占比 32.2%。可满足各类数字要素的业务承载需求，赋能经济社会和民生发展。



图 3.2 多云共生的算力体系

2.3 连接算力的基础网络

以算力为中心，统筹数据中心和客户数字化转型需求，提升网络和运营能力，构建带宽随选，路径可调，智能调度的算力互联和算力接入网络。

高速：围绕福州、厦门双省级算力中心，业务全光接入，优化基础光缆路由及承载网络。实现各地市至福厦 3ms 时延圈。

泛在：以云 POP 为锚点，连接算力和网络，构建天翼云云网。用户可通过互联网、专线、SDWAN 等多种方式获取泛在算力资源。

3. 算力赋能千行百业

算力正在赋能千行百业，在政务、工业、交通、医疗等领域得到了广泛应用，成为推动经济发展的重要力量。本章主要结合福建电信的具体落地案例，对算力赋能千行百业的情况进行阐述。

3.1 赋能社会治理

实现“综治 AI 视频一张图”，利用算力，通过人脸门禁与枪机的配合应用能够对流动人员进行分析，通过一标三实数据碰撞，区分住户、访客及漏登人员。对人、车等信息采集和分析，实现信息数据的利用及合理预警，对城中村人员信息进行采集和登记，形成全系档案，使公安详细了解城中村情况，增强公安对城中村的管理能力；同时城中村借助算力提升城中村自身防控能力，利用科技手段，弥补城中村管理的先天性不足，通过前端设备的建设及权限的管理，实现出入城中村实时监控，人员、车辆信息实时查询，让违法犯罪分子不敢出

现在城中村内，减少治安隐患，让居住环境更安全。



图 4.1 综治 AI 视频一张图

3.2 赋能餐饮行业

基于 AI 算力的明厨亮灶，可满足监管部门、餐饮管理方、消费者的多方需求，提供“Web 管理系统+PC 客户端+监管 APP+微信公众号”整体解决方案。利用明厨亮灶后台管理系统实现：企业档案管理、消费者评价记录管理、监管报告管理、食安政策管理、评分项管理；人员档案管理、食材留样记录等管理,结合客户端实现人员变动管理、食品溯源管理。采用“AI 平台+普通探头”的模式，结合“厨师帽识别”“厨师服识别”“口罩识别”及“人脸识别”等算法更加贴近用户“互联网+明厨亮灶”的理念，并级联各区各单位的视频信息，通过 AI 算法，对各单位进行智慧监管，达到加大监管力度、提升监管效率、关键环节透明化的目的。



图 4.2 明厨亮灶

3.3 赋能环保行业

污水厂主要排污设备周界的安全作为重要防护区域设置视频监控装置，结合视频智能识别技

术和设备，对污水处理的相关技术参数和周界入侵事件进行实时监控，一旦发现异常，立刻报警，并将视频信号传输到控制中心，供相关管理人员进行相应处理处置。通过视频 AI 算力，可有效提升污水处理厂对于水质、水位的跟踪效率，能够及时跟进水质的变化进行技术调整，提升了整体厂区内的工作效率和管理效率，并通过人脸考勤系统对机房等敏感区域进行了更有效的人员管理，保障污水厂区内的安全。



图 4.4 污水处理厂水质、水位跟踪

3.4 赋能工业智造

智能算力在制造业中的应用广泛，围绕提升效率、降低成本、增加产品和服务价值以及探索新业务模式等价值定位产生了不同的细分应用场景，如 AGV 避障、流量控制、货物校验、人工质检行为质量综合评价、动力设备渗透检测等制造业生产辅助 AI 应用。

通过 5G 网络，深度融合云边端+AI 智能，聚焦生产端的安全、质量、管理等方向，5G+AI 生产辅助集成解决方案，联动 5G+AI 智能应用的集成建设框架、提供 SaaS 化应用产品标准化、软硬终端一体化，以及云网多数字要素编排调度，实现“场”“人”、“货”、“设备”四维管理、生产的效率提升和智能化创新，助力产品质量提升和生产设备安全管理，真正做到垂直行业的深耕和规模复制推广。

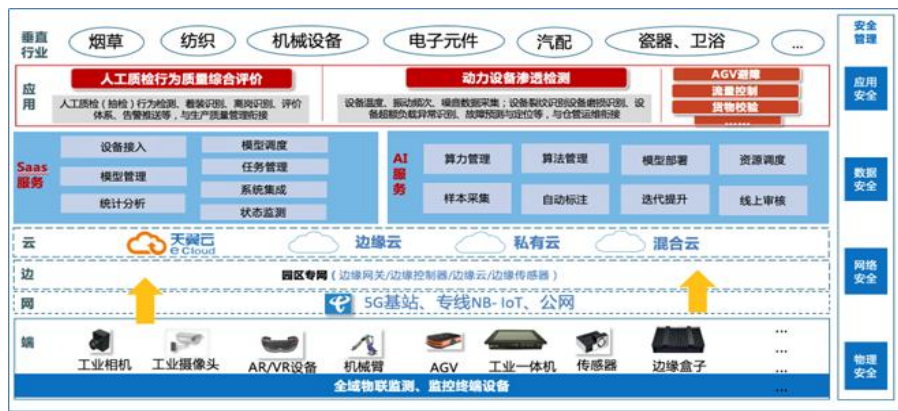


图 4.4 智能制造系统架构

3.5 赋能交通行业

船舶碰撞桥梁安全隐患预警：聚焦闽江通航后，以消除“船舶桥梁碰撞隐患”为目标，解决“航道空间感知设备不足、桥梁水域船舶航行动态监测信息化不足、信息资源整合不足”等问题，融入 5G、大数据、人工智能、北斗导航、物联网的综合应用，打造以基于人工智能技术的船舶碰撞桥梁安全隐

患预警监测为核心、通过激光扫测过桥船舶、5G 数据采集、边缘计算等，提供基于人工智能识别的精准预警分析。以闽江航运数字化安全监管平台为底座，融合预警监控系统触发、预警信息发送、信息沟通等需求，提供船舶智能识别、主动式预警、语音告警、船舶防撞、偏航识别等功能，提高船舶航行安全管理工作的效率和质量。

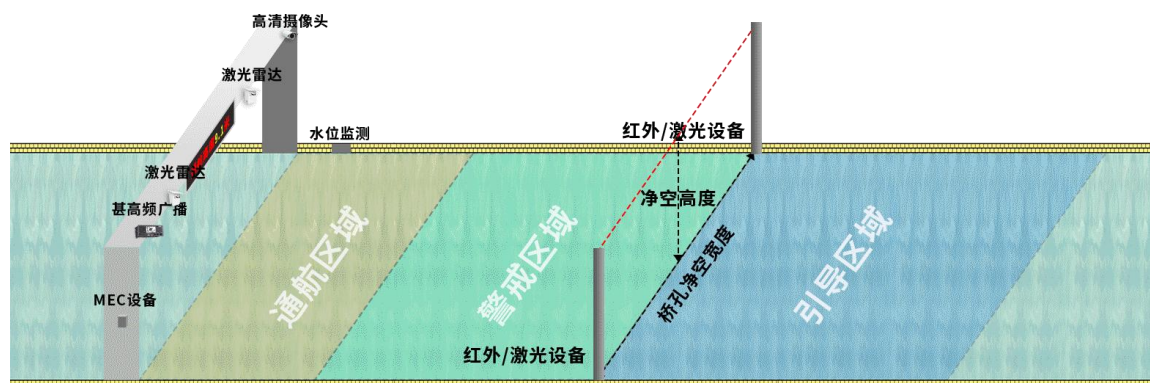


图 4.5 船舶碰撞桥梁安全隐患预警平台

4. 智能算力发展初探

4.1 智能算力飞速发展

ChatGPT（人工智能聊天机器人）一推出便火爆全网。上线仅仅 2 个月，其活跃用户就突破了一亿，对于知识性工作者来说，这就是一场工业革命，全球包括微软、谷歌、百度、阿里巴巴、京东等在内的大互联网公司也看见了机遇，纷纷加入这场 AIGC（人工智能生成内容）“革命”中。而在背后支撑 ChatGPT 的是人工智能大模型，必将进一步带动算力需求的增长。2023 年初，IDC 与浪潮信息联合发布《2022-2023 中国人工智能算力发展评估报告》，报告显示，2022 年中国智能算力的规模首次超过通用算力的规模，预计到 2026 年，中国智能算力的年复合平均增长率达到 52.3%，是三倍于通用算力规模的增长速度。

4.2 智能算力数据中心建设

AI 服务器是算力的承载体。由于 AI 服务器运算速度快，能耗需求和产生的热能比普通服务器更大，CPU 和 GPU 的平均功耗从 300W 跃升至 1000W 以上。用于 AI 训练的机器单机功率或将高达 2.6kW，未来 AI 训练 GPU 单机预计最高可达 10kW，单机架功率近百 kW。同时，工业和信息化部下发《新型数据中心发展三年行动计划（2021-2023 年）》要求，到 2023 年底，新建大型及以上数据中心 PUE 降低到 1.3 以下。智能算力对数据中心基础设施建设提出了新的挑战，为满足智能算力基础设施绿色低碳运营能力建设，智能算力中心建设方案主要分为全风冷架构智算中心建设方案和全液冷架构智算中心建

设方案。

（1）全风冷架构智算中心建设方案

废除列头柜等建设方式，全面采用小母线、弹性方舱、列间空调、2N UPS 等，封闭冷/热通道为智算机柜提供高质量环境。

（2）全液冷架构智算中心建设方案

制冷模式采用风冷+液冷模式，其他建设要求同全风冷架构。

4.3 监管要求

智能算力 AIGC 在赋能人类生产、生活同时，也带来了一系列的安全风险，这些安全风险包括数据风险、算法风险、网络风险和其他风险。特斯拉 CEO 埃隆·马斯克曾表示，大数据的安全隐患比核弹更危险，监管者应当留意，不能让犯罪分子有可乘之机，每个人都应保护个人数据。针对 AIGC 的快速发展，国家网信办联合国家发展改革委、教育部、科技部、工业和信息化部、公安部、广电总局七部门于 2023 年 7 月 13 日联合公布，自 2023 年 8 月 15 日起施行《生成式人工智能服务管理暂行办法》，该办法旨在促进生成式人工智能健康发展和规范应用，维护国家和社会公共利益，保护公民、法人和其他组织的合法权益。国家坚持发展和安全并重、促进创新和依法治理相结合的原则，采取有效措施鼓励生成式人工智能创新发展，对生成式人工智能服务实行包容审慎和分类分级监管。

5. 结束语

算力是数字经济快速发展的重要推力，它通过多种技术的集成创新、协同创新和开放创新，构建

了一个以数据为中心、安全可靠、自主可控的数字化信息基础设施，为数字经济的发展提供了强大的支撑。

算力建设正在向着智能化、绿色化、高密度等方向发展，对相关配套基础设施建设提出了更高的要求。另外，全面推进以算力为核心的云、网、数、智、安等多要素升级，满足算力资源高效连接、按需分配、灵活调度，也是未来算力应用、发展的重要方向。需要建立完善的云网融合的产业生态，以便更好地推动数字经济的发展。

参考文献：

- [1] 邢文娟,雷波,赵倩颖.算力基础设施发展现状与趋势展望[J].电信科学,2022(6):51-61
- [2] 李洁,王月.算力基础设施的现状、趋势和对策建议[J].信息通信技术与政策,2022(3):2-6
- [3] 韩淑君,穆域博,柴瑶琳,宋平,毕立波.算网基础设施发展现状及建议[J].信息通信技术与政策.2022(1):24-29
- [4] 许幸荣,刘琪,夏俊杰,梁鹏,王铁敬.大型数据中心低碳解决方案的现状、趋势和对策建议[J].信息通信技术,2022(4)16-22
- [5] 孙钰坤,张兴,雷波.边缘算力网络中智能算力感知路由分配策略研究[J].无线电通信技术.2022(1):60-67
- [6] 周旭,李琢,范鹏飞,李菁菁.面向科学计算的算力网络技术发展展望[J].信息通信技术.2023年(1):37-43
- [7] 王月,柯芊.智能计算中心:人工智能时代的算力基石[J].中国电业[J].2021(1):11-15

厦门联通算力赋能国贸 AI 视频云

林凯奇

(中国联合网络通信有限公司厦门市分公司)

摘要:随着视频监控技术的发展,逐渐高清化、智能化,传统的视频监控平台的各种弊端逐渐显露出来。并且在大数据时代背景下,在国家战略发展方向的推动中,数字化建设和发展是必然的。利用云计算的算力和存储能力,可以规避传统视频平台的不足,让视频平台对视频的处理能力有了重大的提升,实现统一的安全监管和视频集中存储。联通视频云基于运营商强大的网络和算力基础,提供海量的存储、强大的计算能力、丰富的设备兼容能力和视频的安全可靠。而国贸 AI 视频云正是利用联通视频云的优势,实现全国各个地区仓库的不同视频厂家的视频监控设备进行统一管理,并且强大的算力支撑中,实现 AI 智能分析和多种监控场景的智能应用,达到生产效率、便捷管理和安全性等多个方面的提升。

关键词: 算力 联通视频云 国贸 AI 视频云

1.什么是算力

算力,顾名思义就是计算能力。小至手机、个人计算机,大到超级计算机,算力存在于各种硬件设备中,没有算力就没有各种软硬件的正常应用。以个人计算机为例,不同的配置,用户的体验效果大不相同,这主要取决于不同配置产品的中央处理器(Central Processing Unit, CPU)、内存等的差异性。高配置个人计算机的算力更强,能玩配置需求更高的游戏,运行更消耗内存的 3D 类、影音类软件。低配置个人计算机算力不够,只能体验普通游戏,运行一般的办公软件,在运行对配置要求高的大型游戏时,效果差、卡顿明显,甚至不能运行。同样玩网络游戏,算力更强的手机使用更流畅,算力弱的手机就会有卡顿现象。^[1]算力也不止存在于各种硬件设备,每个人的对事物的思考,也是一种算力的体现,在生活的方方面面中,都有算力的痕迹,每个人都离不开算力。

2.基于厦门联通算力的视频云平台

2.1 视频云平台的优势

视频监控的发展是越来越高清化、智能化,随之而来的是需要传送的数据越来越多、处理能力越来越强。这对数据传输、视频存储、视频平台的处

理性能也有了很大的挑战。传统视频平台的兼容能力低,系统开发维护成本较高,技术处理难度大,特别是企业自己搭建的平台,对于前端设备的兼容性是个很大的问题,各个地区的监控设备往往是不同厂家,在平台的开发和维护上需要耗费大量的人力物力,以往企业都是在各个地区建立机房,打通各个机房的服务器等设备的网络,以此来进行集中管理,这样极大的增加企业的建设和维护成本。而基于云计算的视频云通过云端服务器,使用配套的管理软件集中管理不同类型的存储设备,发挥动态扩展功能,实现海量数据的协同存储。另外,即便系统出现故障,也可以短期内恢复,并实现数据快速转移拷贝,提供新服务。基于视频云平台的监控系统,可以降低技术维护难度和存储成本,将数据存储于数据库服务器集合,以便于提升数据利用效率。相较于传统的多级视频监控平台而言,视频云优势突出,可以实现数据的随取随用,提升数据利用效率。^[2]并且视频云的部分物理设备故障,对于视频监控平台的影响也不大,无疑大大提高了平台的可靠性。

2.2 厦门联通的算力基础

厦门联通在厦门软件园二期联通机房 5 楼 DC 机房建设本地云算力资源机房,机房总面积约 600

平方米,配置 7 个冷通道,可安装总机柜数 166 个。目前骨干云池含 45 台计算节点+34 台块存储+20 台对象存储+10 台裸金属服务器(带 GPU 功能)。并且建设基于国产 ARM 架构的 5 台计算节点服务器和 6 台块存储服务器组成的国产信创专区。可兼容全国产硬件,100%自主可控,是在运营商中唯一通过网信办安全评估的信创云池。目前云池共具备:VCPU7140 核,内存 22320G,存储 2094T, GPU 算力 152TFlops,弹性公网 IP 1265 个。并且在 6 个行政区分别部署了 6 套 MEC 边缘云节点,具备 3607VCPU、9600G 内存、203T 存储能力,可实现本地边缘就近转发能力,打造 10ms 时延圈。厦门联通云采用分层解耦的架构设计,IaaS、PaaS、云管平台全面提供标准化接口,支持各层能力以被集成的方式独立输出,利于构建开放生态。能够多场景部署,实现 IT 云、CT 云和业务云使用统一技术底座。安全可控,拥有三级等保标准和 24 项可信云证书。

2.3 联通视频云的能力

联通视频云聚焦于视频监控上云场景,为客户提供视频监控接入、视频存储、视频处理一体化云平台,实现海量监控视频集中上云管理,主要面向商务楼宇、医院酒店、居民社区、工业园区等场景,提供监控设备统一接入管理、视频就近接入、视频云端存储等一体化视频监控服务。视频云的远程集中监控管理平台是以 PC 电脑为基础,为远程网络用户提供一系列完整的远程监控功能的系统。方便快捷,有电脑和网络就能随地实时查看监控视频,对监控视频进行相关操作,并且可以对用户账号进行权限管理。告别传统视频监控平台依赖特定的服务器或者软件系统,多部门使用不便捷等缺点。系统主要采用 H.264/H.265 视频压缩技术;高度灵活、人性化、易于操作的可定制用户界面;强大的无极限电子地图功能;多种云台快球控制方法,可在图像上直接控制,也可通过云台控制面板控制;灵活、适用性强的分布式组件化结构;完善的用户管理及权限设置功能;系统具有很强的兼容性扩展功能。

3. 厦门联通算力赋能国贸 AI 视频云

厦门国贸集团股份有限公司作为世界 500 强企业,经营的大宗货物分散在全国不同省市区域的

100 个以上的仓库,一直未实现统一的安全监管和视频集中存储,并且目前使用的视频云平台满足不了物联网卡白名单和兼容不同硬件厂家摄像头接入等需求。基于厦门国贸集团股份有限公司需求,厦门联通视频云部署 VMS 视频平台,分别为:管理服务、流媒体服务、存储服务、LIVE 服务、转码服务、国标服务和国标流服务、服务器集群设置,建设 AI 智能分析 GPU 服务器,包括:Ubuntu 系统环境、GPU 设备英伟达 CUDA 资源、规则布署、软件试运行等功能来实现。现在通过联通视频云搭建的国贸 AI 视频云实现了对全国所有库点视频监控摄像头进行统一、标准化管理,并通过 AI 安防视频监控平台的适用性、兼容性、开放性、扩展性实现厦门国贸集团股份有限公司各系统信息互通,并通过 AI 功能提升仓储安全智能化。

厦门联通提供的 AI 视频云平台服务基于联通云提供的 IaaS 能力,厦门国贸集团股份有限公司分布在不同位置的摄像头视频接入到 AI 视频云平台上,平台将视频存储到联通云的对象存储上,AI 功能使用联通云的 GPU 服务器进行 AI 运算,并提供 AI 视频云平台协助对接二次开发工作。为提高系统安全性,云主机均开通云主机安全 vHSM 服务,为视频管理服务器开通云 WEB 应用防火墙 vWAF 服务;为提高系统可靠性,对云主机上的重要数据做定时增量备份。实现厦门国贸集团股份有限公司对视频平台的要求 1、视频平台服务具备丰富的前端设备兼容能力,满足国内主流厂商的视频监控摄像头,如海康威视、大华等主流摄像头;2、可通过流媒体转发管理服务、视频 H5 转码服务、解码上墙服务、视频权限管理服务;3、智能侦测报警与 AI 智能分析服务,需要提供人员入侵、车辆闯入、人员身份识别、车辆类型识别、车牌识别等 15 类 AI 自动识别算法,识别率高。4、实现内部管理驾驶舱的数据展示。厦门联通视频云让国贸 AI 视频云实现了功能的多样化、智能化,管理的简便化。

基于联通视频云的国贸 AI 视频云还能实现电子地图,对地图进行多种操作;对实时视频监控直接拖放、循环播放、实时视频图片抓拍、实时视频录像剪辑、图像效果设置、实时语音监听、与监控中心对讲;能够水平和垂直旋转、调焦、变倍、光圈、辅助开关控制、预置位、巡航、轨迹、自动扫

描；并且拥有各种录像回放功能；可以切换控制大屏墙、电视墙上的监控画面，实现数字矩阵功能；对视频异常报警、设备异常报警、多种方式的报警输入联动、报警输出手动或自动控制；用户帐号分组管理、权限分级管理、功能权限、地图权限、设备权限；备份管理：配置自动备份、还原备份、删除备份等各种强大的功能。

厦门联通视频云旨在充分运用数字化工具，为客户打造安全可靠的现代化、数字化、智能化的仓储运营体系，将运营体系提高效率、提升效能、提质增效落到实处。

4. 结论

随着人们对安全问题的重视，可以看到视频监控的应用场景和数量也多随之增多，但是传统视频平台的搭建难、维护难、成本高等问题也愈发阻碍视频平台的发展。而云计算的发展让这些问题迎来了转机，云计算的种种优势，仿佛都是为视频云平台的难题而生。厦门联通依托自己运营商强大的网络基础，搭建联通云池，提供大量存储和算力资源，

并基于视频平台的使用场景，搭建了智能视频云，提供视频云平台的应用场景企业可以通过联通视频云搭建 AI 视频云平台，管理全国各地的视频监控设备，实时动态的查看视频监控情况，并实现智能化的场景识别。通过国贸 AI 视频云的成功案例，可以进一步了解联通视频云的算力和应用场景。基于联通视频云的强大算力，国贸 AI 视频云可以实现佩戴安全帽、吸烟、未穿防护服、车辆倾翻、打电话和举手求救等 42 种行为和事件的监测和记录，提高生产效率和生产安全。算力是未来的发展趋势，赋能千行百业，通过对算力的应用，视频平台应用已经迎来新的活力。

参考文献：

- [1]雷波，陈运清 边缘计算与算力网络——5G+AI 时代的新型算力平台与网络连接
- [2]程挺，潘鹏志 云计算技术在公安视频监控系统中的应用研究

边缘计算在电化学储能场景中的应用探索

蔡 敏

(厦门二次方网络科技有限公司)

摘要: 随着全球能源需求的增长和环境问题的日益严重, 电化学储能技术作为一种清洁、高效的能源存储方式, 受到了广泛关注。本论文旨在探讨基于边缘计算的电化学储能场景应用研究, 以为电化学储能技术的进一步发展提供理论支持和实践指导。本文首先介绍了边缘计算的基本概念和特点, 然后分析了电化学储能技术的现状和挑战, 接着探讨了边缘计算在电化学储能场景中的应用潜力, 最后提出了一种基于边缘计算的电化学储能系统优化设计方法。

关键词: 边缘计算、边计算策略、电化学储能、EMS 能量管理、动力环境监控。

引言

背景与意义: “边缘计算”的出现背景是为了解决传统云计算模式下的数据传输、处理和存储等问题, 提高数据处理效率和响应速度, 同时保护数据隐私。

在“边云协同”计算场景中, 边缘计算发挥着重要的作用。它与云计算相结合, 形成一种分布式的计算模式。边缘计算可以在边缘设备、边缘服务器和边缘节点上执行部分计算任务和数据处理, 而将更复杂的计算任务和数据存储交给云端进行处理。

这种协同模式可以减轻云端(或数据中心)的压力, 并提高整体的计算效率。

本论文内容的结构安排: 本文在开篇, 阐述“边缘计算”的基础信息、能源领域中的“电化学储能”电站的基础信息、以及两者相结合的相关性与必要性。

阐述过程中, 明确指出“电化学储能”领域中存在的问题, 并结合自身专业、市场观察和团队同事一起实践的成果, 给出一些有助于行业发展的方法。

最后, 在“电化学储能”置于新型能源建设(虚拟电厂)的大环境下, 抒发自己的展望与对关联业态协同发展的期待。

1. 边缘计算的基本概念与特点

1.1 边缘计算的定义

在数字化的社会中, 边缘计算是一种分布式计算模型, 它将数据处理任务从数据中心转移到网络的边缘, 以满足实时、低延迟和高带宽的需求。

边缘计算(Edge Computing), 其在靠近数据源头的一侧, 集计算、存储、应用、网络等能力为一体的计算平台, 在工程实践的现场就近提供分布式计算服务, 将这些提供计算服务端设备纳入上一层网络, 实现更快速、更智能的数据处理和决策。

1.2 边缘计算的特点

1.2.1 低功耗计算技术

计算技术是边缘计算的核心。由于边缘设备的计算 CPU 普遍采用低功耗、多线程方案, 如 ARM 构架系列、X86 构架中的凌动、赛扬 J 系列等 CPU。

1.2.2 存储技术

存储技术也是边缘计算的重要组成部分。由于边缘设备的计算能力和存储能力有限, 非易失性内存(NVM)是一种理想的存储技术, 因为它具有高速读写、低功耗和长寿命的特点。

1.2.3 通讯技术

边缘计算具有低延迟、高带宽的特点。其中, 在无线传输场景, 5G 技术因其高速率、低时延和大连接数的特点, 被广泛认为是实现数据远程传输的理想选择。此外, 为了提高网络的覆盖范围使用如 Wi-Fi 6、LoRa 等近场无线通讯技术。另外, 增加共享有线网络资源以提高整体的数据传输可靠性。

1.2.4 安全技术

安全技术是边缘计算的关键保障。由于边缘设备通常位于网络的边缘，身份验证、访问控制和数据加密等技术是重要的安全技术。

1.2.5 计算逻辑扩展技术

其次，边缘计算具有高带宽的特点。由于边缘设备的处理能力有限，因此它们这使得边缘计算可以处理大量的数据流，满足大规模应用的需求。边缘计算具有弹性和可扩展性的特点。通过使用虚拟化和自动化的技术，边缘设备可以根据需求自动调整其资源配置和工作负载。这使得边缘计算可以根据业务需求快速扩展或收缩。

2. 电化学储能的集成技术应用现状

2.1 电化学储能概述

储能 (stored energy) 是指通过介质或设备把能量存储起来，在需要时再释放的过程。在电力系统中，人们常把大规模的电化学储能电站电力系统“削峰填谷”，保证供电的稳定性。

目前主流锂离子电池可分为磷酸铁锂电池、三元锂电池两大类型。不同于新能源车对能量密度（即续航）的追求，储能更关注循环寿命（充放电次数）、安全性、成本性这些特性，磷酸铁锂电池是目前电化学储能的发展趋势。

2.2 电化学储能技术分类

随着科技的发展，电化学储能技术也在不断进步，其分类方式也日益多样化。

首先，电化学储能可以根据储能原理进行分类。目前主要的电化学储能技术包括：1) 电化学电池，如锂离子电池、镍氢电池、钠硫电池等；2) 电解槽，如超级电容器、金属空气电池等；3) 电化学储能系统，如飞轮储能、压缩空气储能等。

其次，电化学储能可以根据储能容量进行分类。一般来说，储能容量越大，系统的可靠性越高，但同时也需要更大的设备空间和更复杂的控制策略。

再次，电化学储能可以根据储能效率进行分类。储能效率是指储能系统实际储存的能量与输入能量的比值。储能效率越高，说明系统的性能越好。



图释 01 – 储能电站

2.3 电化学储能的主要技术构成

锂离子电池的“超级大电池”一般采用预制仓方式，一个电站由若干电池预制仓组成。标准集装箱中包含若干电池簇，用串并联的方式组织一个电池模组，多个电池模组组成一个电池簇，若干个单体“集装箱”组成的储能电站系统。营运技术应用的基本单位：

- MW是电站功率，MWh是电站容量，功率乘以放电的时间等于电站的容量。
- 电池常用单位是KWh（俗称1度电），较大规模的储能电站通常以MWh为单位，1MWh相当于1000度电。
- 储能电站的充放电倍率（C），是指充放电的速度（时长）与电站容量的比率（互为倒数的关系），如果2小时能充满或放空，充、放电的倍率是0.5C。



图释 02 – 储能电站之储能柜

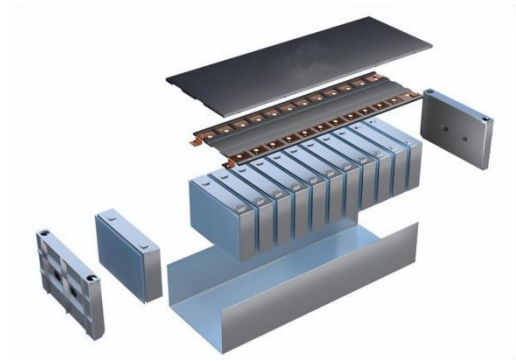
2.3.1 电池管理系统(BMS)

电池管理系统，英文名：BMS（Battery

Management System)，是电化学储能系统中的一个重要组成部分，它可以实时监测和控制电池的状态，确保电池的安全和可靠性。

电池是一个统称，而电芯、电池模组、电池包则是电池应用中的不同阶段，最小的单元就是电芯，一组电芯可以组成一个模组，而几个模组则可以组成一个包。其各阶段的具体表现：

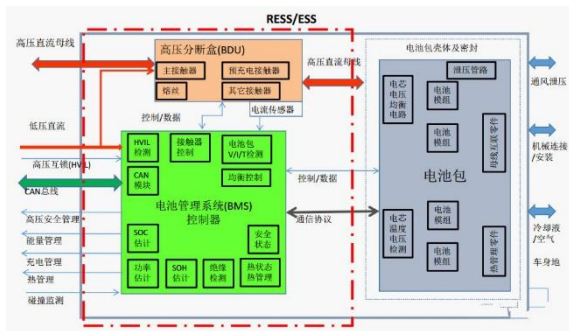
- 电芯：电芯是动力电池的最小存储单元，电芯的寿命也是最为关键的因素，任何一颗电芯的损坏，都会导致整个电池包的损坏。
- 模组：当多个电芯被同一个外壳框架封装在一起，通过统一的边界与外部进行联系时，这就组成了一个模组。比如市场上的 350 模组、390 模组、590 模组等。
- 电池包：而当数个模组被 BMS 和热管理系统共同控制或管理起来后，这个统一的整体就叫做电池包。



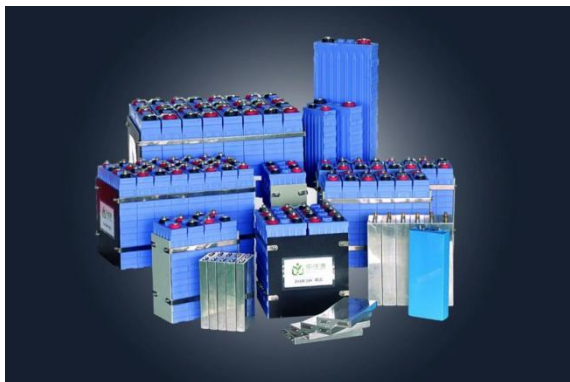
图释 05 - 电池模组



图释 06 - 电池包



图释 03 - BMS 系统结构



图释 04 - 电芯

2.3.2 储能变流器 (PCS)

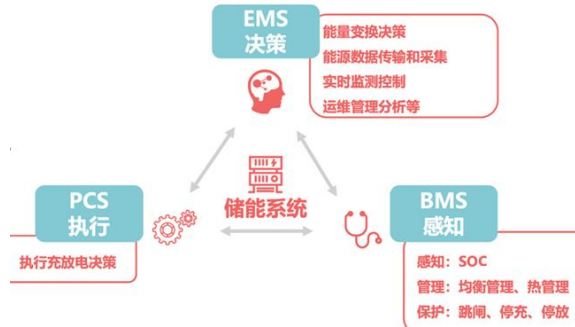
储能变流器，又称双向储能逆变器，英文名：PCS (Power Conversion System)，应用于并网储能和微网储能等交流耦合储能系统中，连接蓄电池组合电网或负荷之间，是实现电能双向转化的装置。即可把蓄电池的直流电，逆变成交流电，输送给电网或给交流负荷（用电器）使用；也可把电网等交流电，整流为直流电，给蓄电池充电。



图释 07 - 固德威 MT G2 系列 (50-80kW | 三相 | 4 MPPT) 外观图

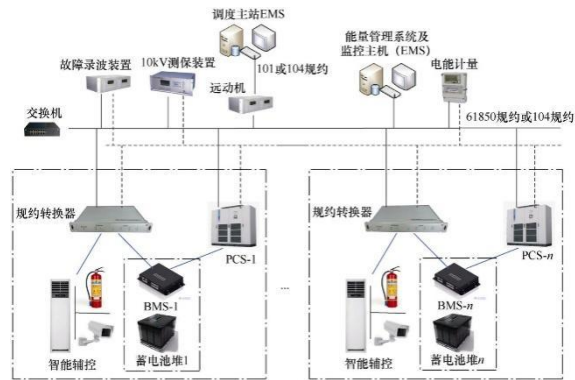
2.3.3 能量管理系统 (EMS)

能量管理系统，英文名：EMS (Energy Management System)，是储能系统的决策中枢，充当“大脑”角色。



图释 08 – 储能核心模块关系

能量管理系统包括电网级能量管理系统和微电网级(可升级为虚拟电厂)能量管理系统，储能系统中提到的 EMS 一般指微电网级(或虚拟电厂)。EMS 的核心基本功能是安全优化调度策略和可视化，储能 EMS 需要负责优化调度，给出多尺度协调控制的调度策略，并自动维持微电网的压频稳定。



图释 09 – 电化学储能 EMS 运行结构图

2.4 技术应用过程中存在的问题

电化学储能技术应用过程中存在的问题包括：

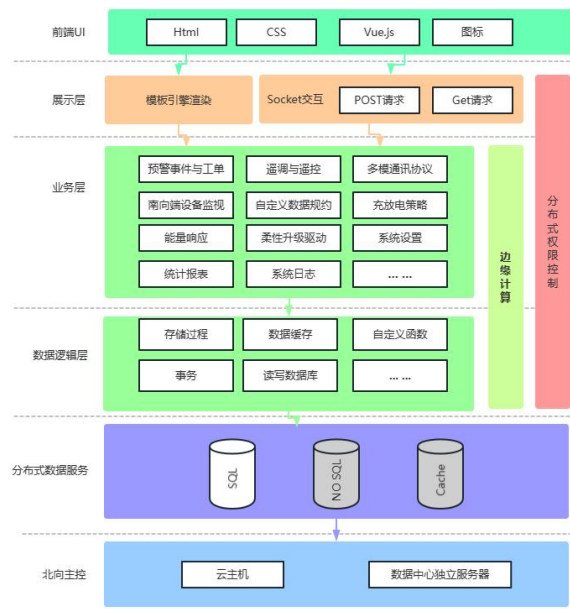
- 在当前商业模式有待健全、成本相对较高的状态下，面对“虚拟电厂”场景时，却因面向“配电侧”的响应技术升级迟缓，表现为更高盈利的水平受限等；
- 在可持续的“峰谷套利”商业模式下，储能电站的现地监控与联控中心之间，暴露出EMS系统集成软件参与盈利的程度不够高；

- 现有储能集成技术在面对大容量、长时间、跨季节等能量调节需求时，存在储能系统集成缺乏统筹规划；
- 计算资源的基础设施建设投入不平衡，注重“云计算”与“本地中心”的建设，但需要大量投资的边缘计算投资不足，导致面对“柔性”电量需求时，数据本地化逻辑处理能力不足，无法满足大规模分布式的能量“精准调控”需求；
- 现有储能系统在面对突发设备状态异常而实施“消警”、“消缺”工单时，分布式设备组织结构不明晰、端设备台账混乱、端设备历史健康状况追溯难等运维难题。直接影响了分布式储能电站的整体盈利水平、安全性和可靠性。

3.边缘计算在电化学储能场景中的优化应用

在分布式电化学储能场景中的边缘计算硬件，置于“储能柜体”中，由柜体电源供电，运行EMS系统集成软件。

3.1 分布式储能集成系统的架构设计



图释 10 – 系统集成架构

3.2 系统界面设计

EMS 软件的整体色调与布局,实施彻底的前后端代码分离,结合 Vue 3.0 方案进一步提高 UI 界面渲染速度。部分扩展绑定数据的 API,同时应用于

数据中心与云端后台与微信小程序端，提高代码的可复用性、可维护性。

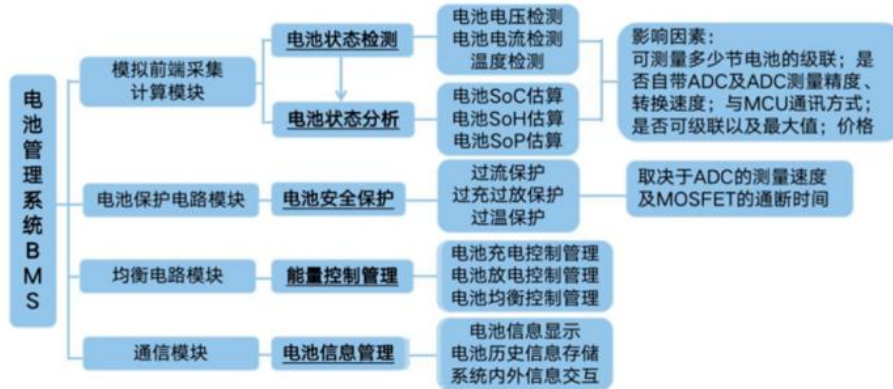
3.3 数据收集与管理

能源管理系统一般分为设备层、通讯层和应用层。设备层需要能量采集变换（BMS、PCS）做支撑。通讯层主要包括链路、协议、传输等。应用层表现形式包括APP、Web等，为管理人员提供可视化的监控与操作界面，具体功能涵盖能量变换决策、

能源数据传输和采集、实时监测控制、运维管理分析、电能/电量可视分析、远程实时控制等。

3.3.1 BMS 相关数采信息

实时监测电池的电压、电流、温度等参数；监视充放电过程，获取状态，避免过度充放电；检测和报警故障，如过压、过流、过温等；提供数据接口，方便上位机进行数据分析和管理的。



图释 11 – BMS 功能示意

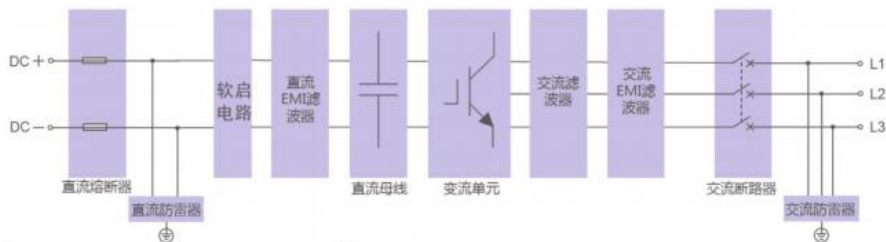
3.3.2 PCS 相关监控与数采信息

在电化学储能系统中，PCS 是负责调节能量的转换和传递的关键组件。须监测的数据类型如下：

- 电流（Current）：监测电流可用于评估系统的能量输入和输出情况，以及电池充放电状态的变化。
- 电压（Voltage）：监测电压可用于评估系统的电能状态和功率传递效率。
- 功率（Power）：监测功率可以用来评估系统的功率输入和输出情况，包括充电功率和放电功率。
- 温度（Temperature）：监测温度可以用于

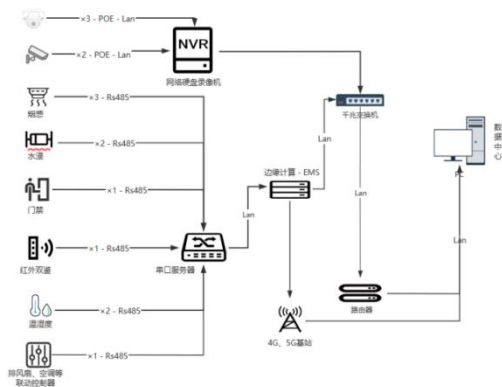
评估系统的热管理和运行状态，以确保系统在安全的温度范围内工作。

- SOC（State of Charge，电池荷电状态）：监测 SOC 可以帮助评估电池的充电状态和可用能量容量，以及预测电池的寿命和性能衰减情况。
- SOH（State of Health，电池健康状态）：监测 SOH 可以评估电池的健康状况和性能衰减情况，以确定电池的寿命和需不需要维护或更换。
- 故障诊断数据：例如电池单体电压、电流不平衡、电池组电压偏差等，用于识别和排查系统故障。



图释 12 – PCS 典型电路框图

3.3.3 柜内动力环境监控信息



图释 13 – 动环监控组网拓扑

3.4 用户管理

角色管理模块，增加组织结构属性（上下3级）。按需新增角色，并分包设置独立的操作权限组合。

- 管理员基本信息，含：操作权限继承其所属角色；因人员可能发生流动，执行“锁定”的管理员账号，将无法再次登录系统，且收回所有权限，但保留其操作数据。
- 最高权限归属超级管理，权限的管理其全权控制。角色的权限，划分为四类：登录与页面进入权限；设备查阅、数据查阅权限、接收通知；控制权限等；责任区块权限，指定负责具体区块的关联设备与事务。
- 设定与变更权限的操作，依赖“角色”的定义，并赋权给具体管理员，同一角色的管理员权限保持一致。

3.5 设备管理

为践行设备的全生命周期管理，提高储能电站的整体运维水平，完善的端、边设备基础信息是必由之路。

- 终端设备管理：分组管理，添加分组：选择设备、名称、通讯类型；子组织、采集点、设备等树形机构。
- 设备数据结构模板定义：模板名称、参量名、浮点型、报警值（可选）。
- 添加设备，含设备基本信息：名称、分类、品牌、型号、备注。
- 设备列表，搜索条件：添加时间区间、站点名、设备种类、设备型号、IMEI号；列表字段（20条每页）：ID、站点名、设备

编号、通讯类型、设备型号、IMEI号、添加时间、添加人；

- 操作：设备编辑、废除（即在系统内，保留，但其不删除）。

3.6 上位机模块

解编码，基于Rs485、CAN、Rj45等总线连接驱动，主要功能包括：实时监控储能电池组的状态、电流、电压等参数；对储能电池组进行充放电控制；对储能电池组进行故障诊断和保护；对储能电池组进行容量估算和健康状态评估；感知与控制动力环境端设备监控实现与下位机之间的通信和数据传输：

- 根据已提供的BMS的数据规约文档，执行 Modbus RTU数据服务；
- 根据已提供的PCS的数据规约文档，并说明数据包的解析约定；
- 基于Rj45接口，应用TCP协议，连接IPC；
- 其他动环监控等南向设备，依据供应商提供的数据规约文档（如：Modbus、DL/T 698-2017、IEC系列），数据包应包含：设备编号、变量点参数、采集频率；
- 北向，应用TCP协议或MQTT协议，自定义加密数据传输频率，边缘侧与云端之间的通讯与管理。

3.7 通讯管理

- 有线网络，客户现场具备有线外网联通条件，南向有Rs485、Rs232、CAN、Rj45等有线通讯方式，边缘计算设备（可选性）北向接入交换机或路由器。
- 无线网络，类型：无线4G、5G全网通模组；
- 通讯连接管理，涉及：设备IMEI号与鉴权、数据包加解密、通讯频谱制式，物联网卡的通讯状态、周期性续费、流量消耗显示。



图释 14 – IOT 通讯之南向通讯配置示意

3.7 智能控制与决策

3.7.1 告警类别

通讯故障、采集故障、其他传感设备的不良状态；预警列表：发生时间、类型、设备或设备组、预警发生值、信息接收人、是否已处理。

3.7.2 消息逻辑与工单

依据网关的通讯心跳特点、所采集的数据，筛选出设备异常、数据异常，及时发出警示消息。频繁出现的同一消息，自动转为“事务预警”。

按三个等级划分，实现降低告警次数的同时不影响“告警本身的价值”。报警升级规则等级逻辑说明：

- 普通（3 条报警及下），较急（10 条报警及下），紧急（30 条报警及下），特急（超过 30 条）；
- 随报警记录数增加，等级就会上升，等级一上升，后台与短信（4 个等级 4 条短信）就会通知，若等级不再上升，隔 5 分钟后台再报一次；
- 初始默认等级为“报警普通”，在没有去派发工单去处理（工单中有一个完成），同一设备变量点的报警都归同一事务中，工单完成后，此事务就会“已处理状态”，若后面有新报警，重新开启一个事务。

备注：为避免重复预警，以上“报警具体条数”可自定义；预警信息接收人与执行人，参照“权限管理中的通知权限”，具体参照以上“4.3 用户管理”。

3.7.3 联控策略

通过对电池的充放电进行控制，实现对用户侧电网的功率调节和稳定。具体的联控策略包括：峰谷平衡、频率调节、电压调节等。从储能电站的投资方需求出发，具体表现为：

- “安全生产”第一原则性策略；
- 峰谷平衡套利策略；
- “虚拟电厂响应”功率调节策略；
- 保障电芯寿命策略与获利策略之间的优先权定义，较为依赖“边云协同”的“遥调”指令集。

随着储能电站运营时间的累积、执行五遥（遥信、遥测、遥调、遥控、遥视）结果的持续反馈，将获得大量数据。历史大数据在“云服务端”完成

数据收集汇总、清洗、标注、训练而导出压缩版AI模型。

在“边缘侧”的EMS策略执行层，响应北向“云服务端”或“数据中心侧”的“遥控”指令，约时执行“遥调”指令，具体执行充放电的功率约束与执行时间周期约束，并将结果反馈至北向。数据流如此往返，实践“机器学习”，实现持续优化AI模型的目的。

3.8 边缘侧安全与隐私保护

- 设备选择标准：在选择边缘设备时，应考虑其安全性。例如，应选择具有安全模块的设备，以便在运行时进行安全检查和更新。此外，还应选择具有防火墙、入侵检测系统和其他安全功能的设备。
- 网络安全设计：边缘设备的网络安全设计应该包括以下方面：身份验证、授权、访问控制、数据与传输加密（SSL证书、X.509证书）、网络隔离、定期漏洞扫描等。此外，在边云协同与客户端运维场景中，还应考虑使用虚拟专用网络（VPN）等技术来保护数据传输的安全。
- 设备配置：边缘设备的配置应该包括以下方面：软件更新、硬件升级、日志记录和监控等。此外，还应考虑使用自动化工具来简化配置过程。
- 运行和维护：边缘设备的运行和维护应该包括以下方面：定期备份数据、更新软件和固件、检查日志文件以及及时响应安全事件等。

3.9 边缘侧应用柔性升级

边缘计算设备上部署本地化应用基础框架，边缘侧节点将自动从云端下载边缘镜像，并部署到边缘计算节点设备中。

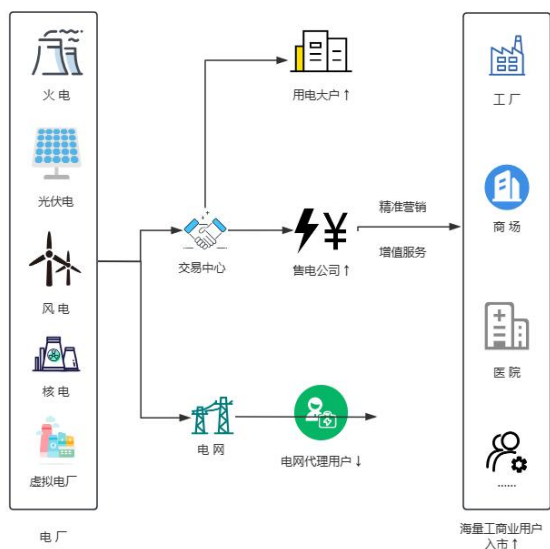
升级镜像的颗粒度细化，如：优化后通讯协议、具体南向设备的特定私有数据规约、响应具体省市“虚拟电厂”的功率遥调指令集、具体南向设备数采频率周期的修改。

镜像包文件采用热升级法与增量更新算法，在节约流量的同时，提高资源包更新成功率。

为保障所采集数据的完整性，在边设备应用升级的过程中，因计算逻辑程序升级而导致服务重启，不影响其他类型串口的数据交互服务。

4. 结论与展望

随着能源转型、结构升级和“双碳”目标的推进，由国家能源局组织11家研究机构编制而成的《新型电力系统发展蓝皮书》指导中描述：以“源、网、荷、储”多向协同、灵活互动为有力支撑，以坚强、智能、柔性电网为枢纽平台，以技术创新和体制机制创新为基础保障的新时代电力系统。



图释 15 – 电力交易市场化后的售电市场运作方式

中金公司指出，“虚拟电厂”能有效聚合分布式资源，为电网提供额外的调节能力，发展潜力大。预计到2025年，我国虚拟电厂可聚合资源空间或达到500GW以上。在发展初期，看好虚拟电厂建设中的“计算类软硬件”需求放量，带动相关厂商业绩增长，长期来看关注盈利模式改善驱动虚拟电厂运营商受益。

“电化学储能”在传统电厂、新能源电厂、输电配网、用电用户等场景的应用，从国家与地方政策性支持、市场可持续获利空间等角度分析，基于边缘计算构建的“电化学储能”系统集成产品的市场规模有望持续增长，已经成为新型电力系统中越来越不可或缺的调节资源，应用前景广阔。

5. 参考文献与 GB 标准

[1]《基于边缘计算的船舶轴系智能监测技术》，作者：王延之，吴军，陈作懿，李晓婷，邓超。发表于：兵器装备工程学报；

[2]《一种基于强化学习的车联网边缘计算卸载策略》，网络首发，作者：张家波，吕洁娜，甘臣权，张祖凡。发表于：重庆邮电大学学报(自然科学版)；

[3]《新型电力系统智慧物联感知技术标准体系研究及思考》，作者：吕顺利，丁杰，张海滨，梅德冬，侯宇。发表于：电力信息与通信技术；

[4]《面向工业物联网的移动边缘计算任务卸载与资源分配》，作者：蔡振启，李志军。发表于：工业控制计算机；

[5]《电化学储能系统方舱设计规范》，T/CEC 175-2018 ；

[6]《储能系统接入配电网技术规定》，Q/GDW 1564-2014 ；

[7]《国家电气设备安全技术规范》，GB 19517 ；

[8]《低压开关设备和控制设备成套设备》，IEC 60439 。

智能算力赋能千行百业

张欢

(福建人工智能计算中心 COO)

摘要: 从 2022 年底开始,以 CHATGPT 为代表的生成式人工智能 (GENERATIVE AI) 成为席卷全球的现象级应用,“人工智能”再度成为热词,随着 CHATGPT、Midjourney 等生成式人工智能技术的发展,人工智能已经从“执行者”提升为潜在的“协作者”,在许多行业中发挥着重要的作用,包括医疗、金融、教育、交通、低碳等。能够熟练掌握人工智能技术、充分发挥技术优势的人才受到各行各业的青睐。未来 10 年,智能算力作为智能世界的基石,将持续推动社会经济发展和科学进步。人工智能发展是一个持续演进的过程,未来福建还可以进一步加强人才培养、加大投入、推动产学研结合等方面的努力,以进一步提升在人工智能领域的竞争力和影响力。

关键词: 人工智能 产业结构 技术与产业相结合 未来发展

1. 人工智能的发展浪潮

人工智能的发展历程可以追溯到上世纪 50 年代。历经诞生与推断期、知识表示与推理期、连接主义与学习期、知识工程与应用期、大数据与深度学习期。

人工智能发展的历程中涌现出许多重要的里程碑,如 AlphaGo 击败围棋世界冠军、无人驾驶汽车的实现、语音助手的普及等。这些里程碑标志着人工智能在理论研究、技术创新和应用落地方面的重要进展。随着互联网的快速发展,大量的数据可供分析和应用。深度学习作为一种基于神经网络的机器学习方法,在图像识别、语音识别和自然语言处理等领域取得了重大突破。同时,强化学习等其他技术也得到了广泛关注。

未来,人工智能将继续推动科技、经济和社会的发展,并带来更多创新和变革。

2. 人工智能产业发展中面临的挑战

2.1 人才缺口

人工智能领域需要高度专业化的人才,如数据科学家、机器学习工程师和算法专家等。然而,供应与需求之间存在差距,使得企业难以招聘到所需的人才。

2.2 技术壁垒

人工智能领域的技术进展迅速,对于一些中小型企业而言,投入大量资源进行研发和创新可能具有挑战性。因此,它们可能无法跟上大型企业在人工智能领域的发展步伐。

2.3 数据访问和质量

人工智能系统的性能很大程度上依赖于训练数据的质量和多样性。一些企业可能无法获得足够的高质量数据来进行有效的训练。

2.4 隐私和道德

人工智能技术需要大量的个人数据来进行训练和改进。隐私保护和数据使用的道德问题在企业发展人工智能时变得更加重要。

2.5 算力资源短缺

人工智能技术的发展需要大量算力资源支持,尤其是大模型的快速发展。动辄百亿、千亿参数级的大模型,以及百 T 的训练数据,需要大量的算力资源投入。

为了克服这些不足,政府需要鼓励企业积极投入人才培养和发展,与高校、研究机构建立合作关系,共同推动人工智能领域的创新;鼓励建设基础算力资源支撑科研创新。同时,政府需要落实数据合规和隐私保护相关政策与措施,与利益相关者开

展调研,共同制定行业准则和标准,确保人工智能发展可持续的重要举措。

3.福建人工智能计算中心实践与创新

2023 第六届数字中国峰会期间,基于全国产化技术路线的建设的“福建人工智能计算中心”正式揭牌成立,首期 100P 算力对外提供服务。这预示着福建省拥有了自己的人工智能算力底座和创新平台,将成为福建人工智能产业发展原动力。

人工智能仍然是一个新兴产业,我们需要大力投入基础算力建设,同时在教学梯队、人才梯队、科研梯队、应用梯队共同发力,做大做强人工智能产业体系。

3.1 福建人工智能计算中心

福建人工智能计算新中心旨在以“普惠算力”为基础,结合“人才培养”“创新孵化”“产业聚合”构建人工智能产业生态。打造“产-教-研-学-用”一体化人工智能发展体系。

3.1.1 普惠算力

计算中心致力降低企业算力使用成本,提供普惠算力,让企业用得起,用得好。在人工智能的起步阶段快速形成生产能力。

3.1.2 人才培养

计算中心重视人才培养,与高校、科研机构和企业合作,组织开展人工智能相关的培训沙龙活动,提升相关人才的专业技能和创新能力。

3.1.3 创新孵化

计算中心为人工智能创新企业,提供技术支持和宣传推广等服务。通过创新孵化平台的建设,促进人工智能创新项目的落地和推广。

3.1.4 产业聚合

计算中心积极与高校、科研机构和企业开展产学研合作,共同研发人工智能技术和解决方案。通过搭建交流平台,促进科研成果的转化和应用。

福建人工智能计算中心的设立旨在加强福建在人工智能领域的创新实力和竞争力,助推福建经济社会的智能化发展。该中心将持续推动人工智能技术与产业的深度融合,为福建打造人工智能创新高地贡献力量。

3.2 福建人工智能计算中心创新合作实践

福建人工智能计算中心已与 170 余家企业展开合作,孵化 53 个解决方案。人工智能应用于各行各业。

3.2.1 长威信息科技有限公司--智能视觉平台方案

本方案以华为 Atlas 800-3000 作为计算核心,智能视觉平台为基础,将视频图像预览、视频图像解析、视频事件检测等视频价值挖掘的技术手段应用于智慧城市管理各行业领域并与行业职能相融合,从而充分发挥城市视频的作用和价值,致力于为智慧城市管理提供一体化的管理支撑。

3.2.2 福建新大陆--钢制品喷绘文字识别联合解决方案

本方案依托于华为 Atlas 900 AI 的算力支撑,通过前期预采集的图片训练了适用于钢铁厂的钢制品喷绘文字识别模型,将模型部署至手持式设备上,实现了复杂场景下的文字识别,在提高效率的同时可确保工人在安全位置操作。

3.2.3 福建数字孪生--气象数字孪生应用系统方案

数字乡村气象应用可视化大屏依托于华为 Atlas 900 AI 的算力支撑,借助云计算、物联网、互联网、大数据、空间地理信息和数字孪生等新型技术,构建包含综合态势、乡村治理、乡村振兴三大模块的数字气象乡村服务新模式。让气象为数字乡村振兴赋能,科技助力乡村振兴。

3.2.4 福建乐想天成--可穿戴智能执法系统方案

本方案依托于华为 Atlas 900 AI 的算力支撑,为建立能够对可能犯罪嫌疑人快速识别并及时预警的可穿戴智能执法系统,实现毫不放松地把常态化维护安全稳定工作抓紧抓实抓细,着眼于构建“核心芯片与算法”——“智能终端”——“分析系统”三级技术路线,实现“秒级速度、百路并发”的国内领先的解决方案。

3.2.5 福建星榕基--垃圾投放检测联合解决方案

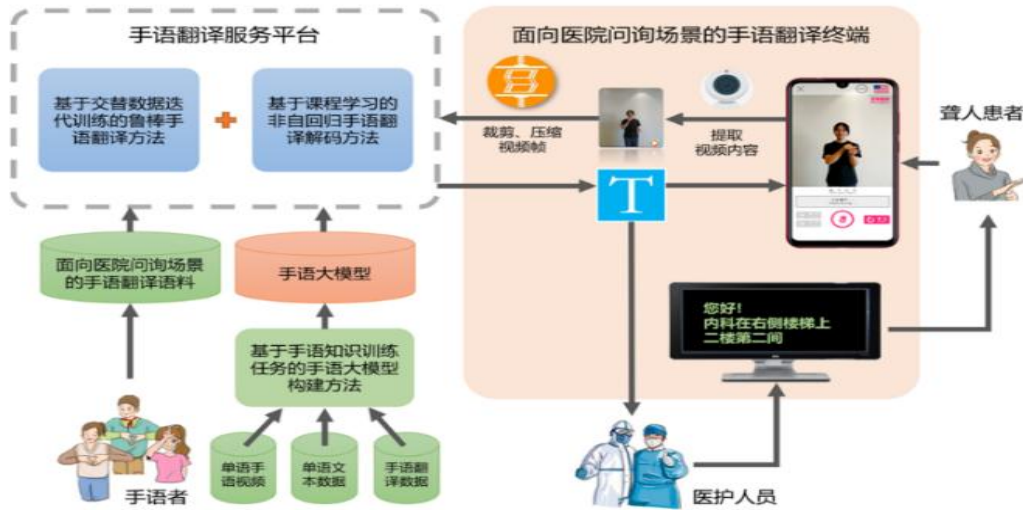
方案以华为 Atlas 200 AI 加速模块作为计算核

心，配套垃圾屋现有摄像头。通过对监控摄像头获取实时视频流，并将数据传输至 AI 加速模块，经过智能分析后，实现对垃圾投放过程中涉及到的人物、事进行全过程实时监控，优化垃圾投放管控模式，提升垃圾投放管理质量。

3.3 福建人工智能计算中心科研创新实践

3.3.1 厦门大学陈毅东团队“白鹭·迦叶”手语翻译大模型

计算中心通过算力服务孵化“白鹭·迦叶”手语大模型，对于包括手语翻译在内的手语智能处理的技术生态具有里程碑意义，将解决手语服务人员紧缺的问题，在医疗问询、银行、机场、高铁、政务办公、民生等领域拥有广泛需求。

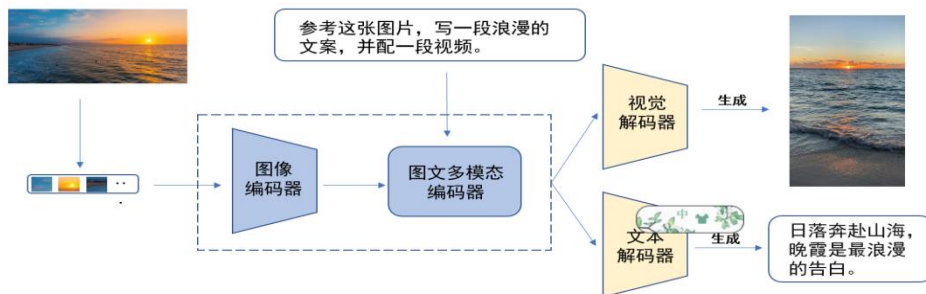


(图-01)

3.3.2 中国人民大学卢志武团队 ChatImg 大模型
计算中心通过算力服务引入 ChatImg 大模型，该大模型通过和 OpenAI 结合，使 ChatImg 不仅支

持文字聊天，还能看懂图片上的内容并根据图片内容回答问题。通过 ChatImg 多模态大模型可以为 ChatGPT 装上“眼睛”，实现了图文对话，进一步推动了数字人、机器人行业的发展。

多模态通用生成模型：基本框架



- 通用生成的定义：多模态输入，多模态输出（语言生成、图像生成、视频生成）

(图-02)

3.4 福建人工智能计算中心人才培养实践

福建人工智能计算中心着力于人工智人才梯队建设,构建“产-教-研-学-用”一体化生态体系。

3.4.1 教学师资培养

计算中心与福建省多家高校建立师资培训合作,面向福建各大高校人工智能学院的授课教师,针对昇思 MindSpore AI 框架进行全面讲解和实践培训,助力高校教师提升教学实力,丰富实践经验。目前已开展多次师资培训活动,培养 200 余名人工智能课程教师。

3.4.2 人工智能人才培养

由集美大学、华为技术有限公司、福州市电子信息集团有限公司、罗普特科技集团股份有限公司联合举办的“昇腾 AI 人才暑期班”实践取得成功。集美大学输出参培学员,华为提供昇腾 AI 技术和师资课程包,福建人工智能计算中心提供算力资源,罗普特人工智能学院、福州市电子信息集团提供学员实习就业岗位,从而实现昇腾 AI 人才培养“产-教-研-学-用”生态闭环。

4. 技术与产业相结合

人工智能技术与产业相结合已经在各个领域展现出巨大的潜力和广泛的应用。

4.1 制造业

人工智能可以应用于生产线优化、质量控制和预测性维护等方面,提升生产效率和降低成本。机器人和自动化系统也能够通过人工智能技术实现更精确和灵活的操作。

4.2 金融服务

人工智能可用于风险评估、欺诈检测、智能投资和客户服务等金融领域的任务。它可以通过数据分析和模型预测来提供更准确的决策支持和个性化的服务。

4.3 零售和电子商务

人工智能在推荐系统、个性化营销和库存管理等方面具有广泛应用。使用人工智能技术,企业可以更好地理解顾客需求,提供个性化的购物体验。

4.4 医疗保健

人工智能可以用于医学影像分析、辅助诊断和预测,帮助医生提高准确性和效率。它还可以支持药物研发、个性化治疗和健康管理。

4.5 交通和物流

人工智能可用于交通流量预测、智能交通管理和物流路线优化等领域。它可以提高交通效率、减少拥堵,并优化物流运输过程。

4.6 农业和农村发展

人工智能技术在农业生产、农作物病虫害监测和精准农业方面有着广泛应用。通过数据分析和智能决策支持,可以提高农业生产效益和资源利用效率。

除了以上领域,人工智能还在教育、能源、安全和环境等诸多领域中得到应用。随着技术的不断进步和创新,人工智能将继续与各个产业相结合,为其带来更高效、智能和可持续发展。

5. 人工智能的发展与展望

5.1 大模型将在人工智能发展中扮演重要的角色

大模型对于提高任务性能、推动技术进步和应用创新起到了关键作用。

5.1.1 提高任务性能

大模型具有更强大的学习能力和表达能力,可以处理更复杂的任务。例如,在自然语言处理领域,大型预训练语言模型如 GPT 和 BERT 已经取得了显著的成果,能够在机器翻译、文本生成和问答系统等任务上达到甚至超过人类水平的表现。

5.1.2 推动技术进步

大模型的出现激发了对模型设计、训练算法和优化方法的研究。为了满足大模型的需求,研究者不断探索如何构建更大、更高效、更可解释和更具泛化能力的模型。这些研究成果不仅推动了大模型的发展,也为其他领域的人工智能技术提供了借鉴和启示。

5.1.3 促进数据集和基准的发展

大模型需要大规模的标注数据进行训练,这推动了数据集和基准的发展和完善。为了训练大模型,

研究者和业界不断努力收集、标注和共享更大规模、更多样化的数据集，提高模型的泛化能力和适应性。

5.1.4 探索新的应用领域和商业机会

大模型的强大表现和多功能性使其能够应用于各种领域和行业，包括自动驾驶、医疗诊断、金融风控等。这为企业和研究机构带来了新的商业机会和创新方向

综上所述，大模型的出现，提高了任务性能，推动了技术进步，促进了数据集和基准的发展，推广了迁移学习和预训练技术，并探索了新的应用领域和商业机会。随着技术的不断发展，大模型将在人工智能领域中发挥至关重要的作用。

5.2 人工智能在未来将服务于更多企业。

随着人工智能技术的不断发展和成熟，越来越多的企业开始意识到利用人工智能的潜力来改进他们的业务、提高效率并探索新的商机。

5.2.1 自动化与智能化

人工智能可以帮助企业实现自动化和智能化的生产和运营流程。例如，自动化的机器人和智能化的机器视觉系统可以大大提升生产线的效率和准确性。

5.2.2 数据分析与预测

人工智能能够处理和分析海量数据，从中挖掘出有价值的信息和洞察。通过利用机器学习和数据挖掘算法，企业可以进行更精确的业务预测、市场趋势分析和客户行为模式识别等。

5.2.3 客户服务与体验

人工智能可以提供更智能、个性化的客户服务和支持。聊天机器人、虚拟助手和智能客服系统能够快速响应客户需求，解答问题，并提供定制化建议，提升客户满意度和忠诚度。

5.2.4 增强决策与管理

人工智能可以辅助企业管理层进行决策。通过数据分析、模型建立和预测算法的支持，人工智能可以帮助企业领导者做出更明智的战略和运营决策。

5.2.5 创新与研发

人工智能技术为企业创新提供了新的机会和工具。例如，机器学习和深度学习算法可以用于新产品开发、服务优化和市场推广等方面，促进企业的创新和竞争力。

随着人工智能技术的不断演进和应用范围的扩大，越来越多的企业将意识到引入人工智能的重要性，并积极探索如何将其应用于自己的业务中，以获得更大的竞争优势和商业价值。

5.3 人工智能将走入千家万户

随着人工智能技术的不断进步和普及，人工智能正逐渐走入千家万户，融入人们的日常生活中。

5.3.1 智能助理

智能助理已经成为人们生活中的常见存在，如语音助手和聊天机器人。它们可以回答问题、提供信息、控制智能家居设备等，为人们提供便利。

5.3.2 智能家居

人工智能应用于智能家居系统，使得人们可以通过语音或手机控制家庭电器、灯光、安防系统等，实现智能化的居住环境。这样的系统可以提高生活便利性、节省能源消耗，并改善居住体验。

5.3.3 智能健康监测

人工智能技术在医疗领域的应用越来越广泛，例如智能手环、智能手表等可收集个人健康数据，如心率、睡眠等，并提供健康建议和预警。

5.3.4 个性化推荐

许多平台和应用程序使用人工智能算法来分析用户的兴趣和喜好，并提供个性化的推荐内容，如电影、音乐、商品等。

5.4 人工智能算力将成为重要生产资料

5.4.1 大模型训练需求

大模型的训练离不开大量算力的支持，动辄百亿、千亿级的大模型需要百 P，千 P 的算力支持。算力是大模型快速成长的基石。

5.4.2 企业生产需求

企业的智能化升级改造，产线的智能化运行离不开大量算力的支持。特殊环境下的行业更需要定制化算力的支持。

5.4.3 智能生活需求

随着各种智能应用的普及，人们的生活将越来越智能化，对人工智能的需求将指数化增长。智能算力将会像手机一样成为人们生产生活中不可或缺的一部分。

总体而言，人工智能正逐渐渗透到人们的生活中，为我们带来更多的便利和创新。随着技术的进一步发展，未来我们还将看到更多人工智能在各个领域和方面的应用，使得人工智能走入更多家庭和生活场景。

这里做一个大胆的展望：未来 10 年，传统行业将被人工智能所取代，人们将更专注于创新工作。人工智能算力将成为重要的生产资料，中国每个家

庭都将有 1P 的算力需求并拥有自己的智能家庭管家。千行百业，千家万户，衣食住行都将产生天翻地覆的变化！

参考文献：

- [1]浪潮信息、清华全球产业院《2022-2023 全球算力指数评估报告》
- [2]中国新一代人工智能发展战略研究院 刘刚《中国新一代人工智能科技产业发展报告》
- [3]经济观察报《福建经济分析报告》
- [4]资产信息网《2023 年人工智能行业研究报告》
- [5]清华大学出版社 谷建阳《AI 人工智能》

基于互联网数据的城市 3D 建筑 射线追踪模型研究及应用

董帝焯 许绍松 黄炳塔 杜丕加 吴晶晶
(中国联合网络通信有限公司泉州市分公司)

摘要: 无线网络传播模型在网络规划和优化中非常重要,作者基于网络建筑数据的 3D 射线追踪模型,在矢量瓦片数据的基础上,针对蜂窝环境中无线接收信号特性进行了相关研究,利用镜像射线追踪算法,通过计算射线过程得到仿真结果,不但可以识别信号的来源,还可基于建筑物级别进行覆盖预测,进行模型算法的测试和验证,在 5G 网络规划和优化中具有指导意义。本方法简单,方便,相比商业仿真软件的复杂参数设置和高昂的价格,提供了另一种高性价比选择。

关键词: 5G, 覆盖仿真, 射线追踪模型、3D 建筑数据

引言

随着 5G 时代的到来,物联网、云计算、大数据等技术快速发展。信息基础设施日益完善,全球时空大数据呈爆发式增长,同时,随着多媒体业务越来越多,人们对无线通信系统传输速度的要求越来越高。为了满足高速的无线数字通信要求,无线通信系统将向高频宽带方向发展,工作频率将在几个 GHz 到几十个 GHz 之间。在移动通信无线系统设计中,一个主要的任务是在满足话务容量的要求下,使得网络能够达到满意的质量,例如覆盖率、掉话率、接通率和话音质量等,而这些系统设计指标主要由发射装置和接收装置之间传播环境决定。电磁波在环境中传播的过程中会受到环境中的物体和地形影响。因此,我们需要建立精确的传播模型,通过对无线传输信道的模拟和仿真,预测接收信号,从而为指导网络规划提供较为准确的理论依据。

传统无线网络仿真采用的模型相对单一且较为理想,实际无线信号在空间传播时一般都会受到建筑物、树木等的遮挡,信号会发生直射、反射、衍射、透射等,终端接收到的信号一般都是多径传播的结果。5G 时代,由于使用较高的频段,覆盖面积相比 4G 缩小,在网络设计、场强预测、干扰分

析时就越要更多地考虑基站周边建筑物的影响。由于城市中基站周边环境的复杂性和多变性,传统的统计模型已经基本失效,而以射线跟踪为代表的确定模型正是处理这一问题的有效方法。

1. 射线追踪模型特性

射线追踪是属于确定性电磁波传播特性预测分析方法的一种,它在无线通信系统设计具有指导作用。射线追踪技术追踪从源点出发的所有射线在环境中的传播路径,然后根据射线的影响范围确定接收点处的电波的强度、相位、电场极化方向,而且还可以结合发射天线和接收天线的方向图、天线增益和天线极化计算出接收天线接收功率。常用的射线追踪算法有:镜像法、入射反弹射线追踪技术(SBR)、射线管追踪技术。

1) 镜像法射线追踪技术: 根据源的各阶镜像确定接收点到源的射线传播路径,避免了对多余射线的追踪,所以计算速度快,成为当前国内外研究的热点。

2) SBR 追踪技术: SBR 方法适用于各种复杂环境下电波传播特性的预测,这种方法的缺点是需要进行数量巨大的相交测试,如果需要预测的接收点很多,那么接收测试的数量也会非常大,速度较慢。

3) 射线管射线追踪技术：由 SBR 方法变化而来，和入射反弹射线追踪技术相比，它的优点是不需要进行接收测试，所以，射线管的追踪技术没有重复的射线管被接收，但是在射线管的追踪过程中仍然需要射线与面的相交测试，所以这种方法的射线追踪过程也需要巨大的计算时间，硬件要求高。

综上所述，在城市环境下高楼密集，电波传播复杂环境下，为提升计算速度，提高效率，本文主要研究从互联网获取的建筑数据，利用镜像法射线追踪技术，进行建筑物级别的覆盖评估，在 5G 网络规划和优化中具有重要的指导意义，该方法不但解决 5G 覆盖仿真技术上的难题，还节省了大笔使用商业仿真软件的费用，且易于复制和推广，极具价值意义。

2. 基于建筑数据的镜像射线追踪模型的实现

2.1 互联网建筑物数据的获取

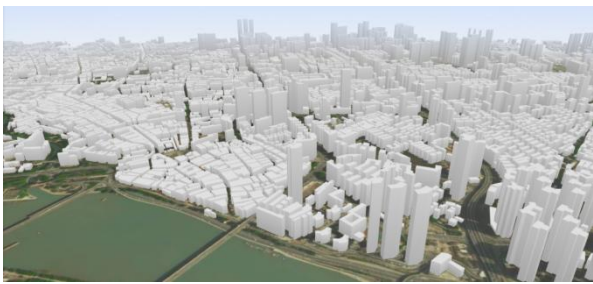


图 1:互联网建筑数据展示图(包含高度数据)

随着 webGIS 技术的发展，目前网络地图已经采用矢量瓦片的方式存储，如上图所示，矢量瓦片中使用矢量数据描述地图元素，通过渲染这些矢量数据形成地图，数据量很小，缩放的时候不会失真，可以在不同视角展示，可以展示更丰富的高度信息，例如可以拔高建筑物，其在数据库中表现为数据，而非原来的 JPG 栅格图片，矢量数据的对象类型包括：

1、利用 Python，通过网络地图的 API 开放接口和 URL 矢量瓦片地址下载城区地图的点、线、面数据，网络地图共有 18 级，只要下载精度最高的 18 级即可。

地图元素	模型信息
POI	点数据,文本, 重要度 (用于POI碰撞), 商标, 类型
公路	线数据, 名称, 道路等级, 方向性, 类型
铁路	线数据, 名称, 类型
水面	名称
绿地	名称
建筑物	名称, 高度
边界线	类型

表 1: 矢量地图数据类型

```
def __init__(self, coord, code):
    if (code == "REGION_BUILDING_ROAD"):
        self.GROUP_URL = "https://tms1.tiles.mapbox.com/v4/2s2t7y1ed5d4d1112region_built{img_road}({x},{y}).pbf"
        self.TABLE_NAME = "REGION_BUILDING_ROAD"
    elif (code == "ROADLABEL_POILABEL"):
        self.GROUP_URL = "https://tms1.tiles.mapbox.com/v4/2s2t7y1ed5d4d1112roadpoi_label({x},{y}).pbf"
        self.TABLE_NAME = "ROADLABEL_POILABEL"
    else:
        raise Exception("Error code: " + code)
    self.coord = coord
    self.rum = 0
def download(self, url, sid, total):
```

图 2: 矢量瓦片 URL 地址连接 python 模块

```
17:09:40 -> [216138, 217670, 102786, 104372] / zoom: 18 ...
17:09:40 -> total: 2432871 ...
17:09:40 -> bundles: 24800 ...

17:09:44 [i]: Downloaded GROUP: 100 / 24800
17:09:47 [i]: Downloaded GROUP: 200 / 24800
17:09:51 [i]: Downloaded GROUP: 300 / 24800
17:09:55 [i]: Downloaded GROUP: 400 / 24800
17:09:59 [i]: Downloaded GROUP: 500 / 24800
17:10:02 [i]: Downloaded GROUP: 600 / 24800
17:10:06 [i]: Downloaded GROUP: 700 / 24800
17:10:10 [i]: Downloaded GROUP: 800 / 24800
17:10:14 [i]: Downloaded GROUP: 900 / 24800
17:10:17 [i]: Downloaded GROUP: 1000 / 24800
17:10:22 [i]: Downloaded GROUP: 1100 / 24800
```

图 3: 下载 18 级矢量地图数据示意图

2、对下载的数据进行解析，还原为经纬度数据。

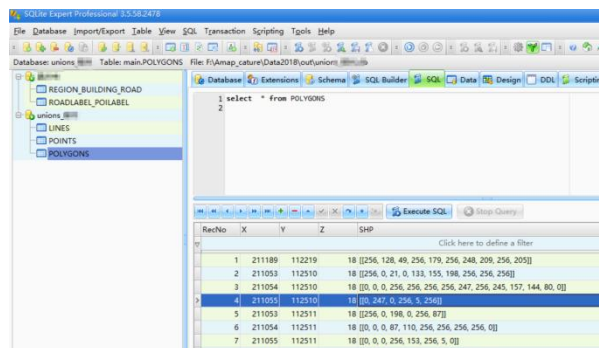


图 4: 保存到本地的矢量地图数据库

矢量瓦片 (Vector Tile) 运用 Google Protocol Buffers 作为编码方式, 通过自定义数据结构可以跨平台跨语言地实现矢量数据的存储, 而且它以二进制流形式进行存储, 大大节省了存储空间。Vector Tile 依据投影坐标及瓦片组织体系, 实现地理区域范围与瓦片编号的一一对应关系。每个 Vector Tile 至少包含一个 Layer, 每个 Layer 中又包含其图形要素和属性信息, 图形信息被分解成一系列 Line to、Move to、ClosePath 绘图命令, 一个 command id 代表一种操作, 每个 command ID 表示以下指令中的一种, 其中 1 代表 Move to, 2 代表 Line to, 7 代表 ClosePath。Line to 和 Move to 均有两个参数, 而且每个命令都有一个命令执行次数, 参数采取 zigzag 的编码方式, 小的正值和负值的编码均为小整数, 因此参数的编码值 ParameterInteger 的计算公式是 $ParameterInteger = (value \ll 1)^{(value \gg 31)}$ (value 为参数的实际值)。

指令	Id	参数	参数个数
MoveTo	1	dX, dY	2
LineTo	2	dX, dY	2
ClosePath	7	无参数	0

表 2: 矢量瓦片轮廓画图指令

假设示例点的坐标为: (25, 17), 表示它只需要一条指令: MoveTo(+25, +17):

```
编码      : [ 9 50 34 ]
| | `> 解码: ((34 >> 1) ^ -(34 & 1)) = +17
| `> 解码: ((50 >> 1) ^ -(50 & 1)) = +25
| ===== 相对地 MoveTo(+25, +17) == 创建点 (25, 17)
| `> [00001 001] = command id 1 (MoveTo), command count 1
```

假设多点的坐标为: (5, 7), (3,2)表示它只需要一条指令: MoveTo(+5,+7), MoveTo(-2,-5):

```
编码      : [ 17 10 14 3 9 ]
| | | | `> 解码: ((9 >> 1) ^ -(9 & 1)) = -5
| | | | `> 解码: ((3 >> 1) ^ -(3 & 1)) = -2
| | | | === 相对地 MoveTo(-2, -5) == 创建点 (3, 2)
| | | | `> 解码: ((34 >> 1) ^ -(34 & 1)) = +7
| | `> 解码: ((50 >> 1) ^ -(50 & 1)) = +5
| | ===== 相对地 MoveTo(+25, +17) == 创建点 (25, 17)
| `> [00010 001] = command id 1 (MoveTo), command count 2
```

假设建筑物要素的坐标为: (3,6),(8,12),(20,34),(3,6)闭合, 则指令为:

- MoveTo(3, 6)
- LineTo(5, 6)
- LineTo(12, 22)
- ClosePath

```
编码      : [ 9 6 12 18 10 12 24 44 15 ]
| | | | | | | | | | `> [00001 111] command id 7 (ClosePath), command count 1
| | | | | | | | | | ===== 相对地 LineTo(+12, +22) == 连接到点 (20, 34)
| | | | | | | | | | ===== 相对地 LineTo(+5, +6) == 连接到点 (8, 12)
| | | | | | | | | | `> [00010 010] = command id 2 (LineTo), command count 2
| | | | | | | | | | ===== 相对地 MoveTo(+3, +6)
| | | | | | | | | | `> [00001 001] = command id 1 (MoveTo), command count 1
```

3、经过上述序列算法解析后, 按网络地图开发 API 接口公式将像素坐标转化为经纬度。

```
def FromPixelToCoordinate(self, pixelX, pixelY, zoom):
    # from pixel xy in tile to gps lat lng
    tile_full_width, tile_full_height = self.GetTileMatrixMaxXY(zoom)
    mapsizex = (tile_full_width + 1) * self.TileSizeWidth
    mapsizey = (tile_full_height + 1) * self.TileSizeHeight

    xx = min(max(pixelX, 0), mapsizex - 1) * 1.0 / mapsizex - 0.5
    yy = 0.5 - (min(max(pixelY, 0), mapsizey - 1) * 1.0 / mapsizey)

    lat = 90 - 360.0 * math.atan(math.exp(-yy * 2 * math.pi)) / math.pi
    lng = 360 * xx
    return lat, lng
```

图 5: 像素坐标转经纬度坐标模块

4、通过墨卡托坐标与瓦片编号的计算解析位置信息, 多个步骤流程最终得到包含楼层高度的建筑物矢量数据:

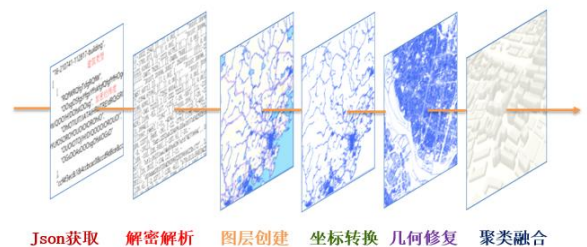


图 6: 矢量数据解析过程示意图

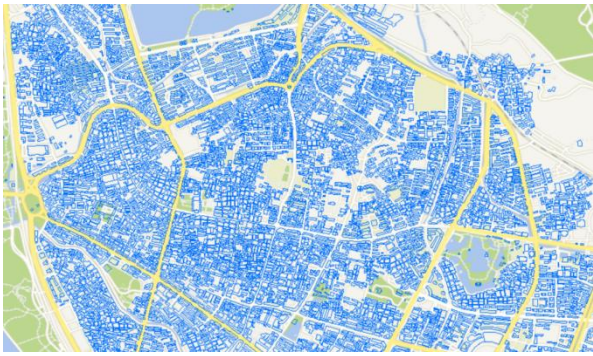


图 7：矢量地图数据生成的图层

在 GIS 中建模呈现的 3D 效果如下：

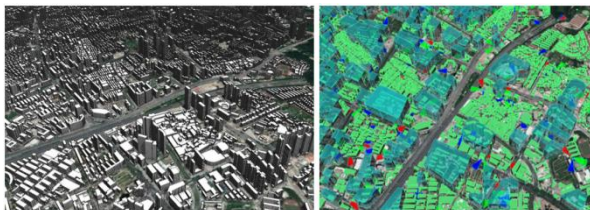


图 8：矢量建筑数据 3D 效果图

2.2 5G 链路预算和电平计算

无线网络规划是 5G 部署的重要一环，而对 5G

基站的覆盖预测又决定了 5G 网络的部署方法与规模。传播模型反映了相应频段的无线信号在不同环境传播损耗的情况，是覆盖规划中的重要输入。3GPP 38.901 文档提供了最新的适用于 0.5GHz-100GHz 频率范围内的 5G 传播模型，其包含视距&非视距概率传播模型、大尺度空间损耗模型以及穿透损耗模型。空间损耗与频段、传播路径、所处的地物、基站和终端的高度密切相关，空间损耗与距离的统计类数学模型，频率越高，损耗越大。

在 5G 覆盖规划中，主要通过链路预算测算出满足网络覆盖要求的最大路径损耗，根据最大允许路径损耗，利用传播模型测算出站点的覆盖半径。本文通过研究 5G 传播模型，并结合链路预算，讨论分析了不同频率下 5G 信号的损耗情况，给出了不同边缘速率要求下站间距参考。

链路预算是系统总增益和总损耗的计算，它是终端（UE）接收到信号电平（RxSL）；终端接收到信号电平与接收机灵敏度(RxS)进行比较，以检查信道状态是否正常，如果接收信号电平(RxSL)优于接收灵敏度(RxS)，则信道状态为“通过”(可用)，否则为“失败”(不可用)。

下图显示一些示例值，以说明计算连路预算时使用的不同输入参数：

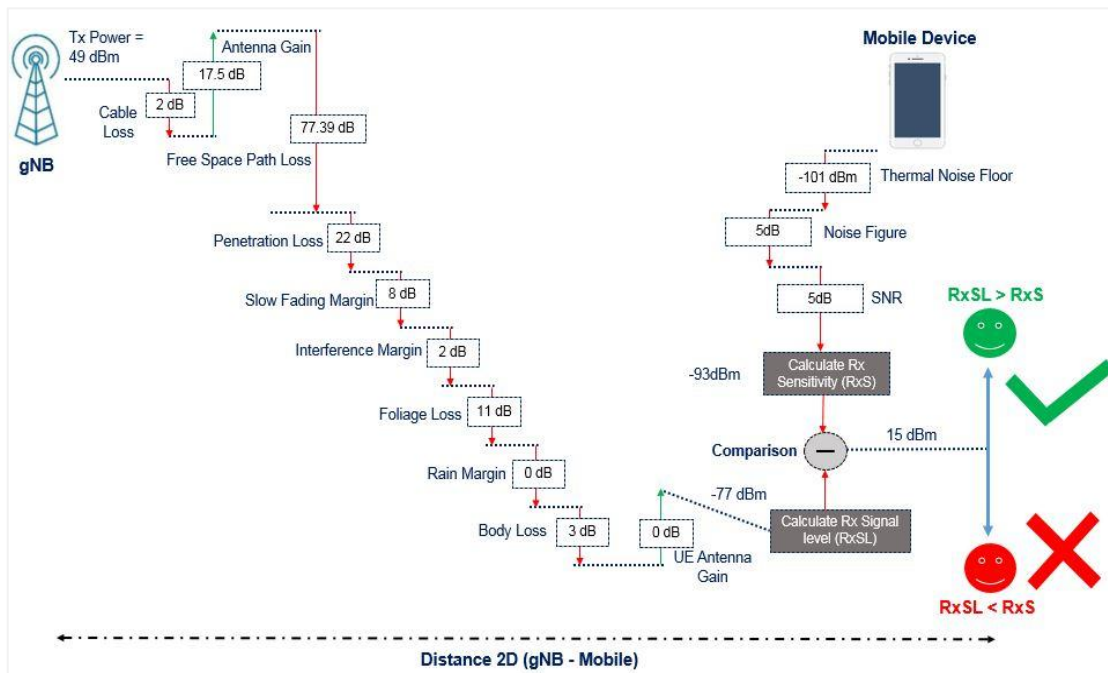


图 9：5G 链路预算示意图

5G 链路预算中计算接收信号电平(RXSL) 公式如下:

$$\begin{aligned} \text{Received Signal Level at receiver (dBm)} = & \text{gNodeB transmit power (dBm)} - 10 \cdot \log_{10}(\text{subcarrier quantity}) \\ & + \text{gNodeB antenna gain (dBi)} - \text{gNodeB cable loss (dB)} - \text{Path loss (dB)} \\ & - \text{penetration loss (dB)} - \text{foliage loss (dB)} - \text{body block loss (dB)} \\ & - \text{interference margin (dB)} - \text{rain/ice margin (dB)} - \text{slow fading margin (dB)} \\ & - \text{body block loss (dB)} + \text{UE antenna gain (dB)} \dots\dots\dots(1) \end{aligned}$$

在上例中, 接收信号电平 RXSL=77 dBm, 接收灵敏度(RXS)=-93 时, 链路预算结果为“通过; 通常受限制的是上行链路, 建议计算时下行和上行链路分开预算, 然后考虑最坏链路。

为计算路径损耗, 需使用合适的 5G 传播模型 (5G 使用 3GPP 36.873 定义的 3D 传播模型)。UMa, Umi 和 RMa 模型适用于 2-6GHz 频段, 在 3GPP 38.901 中扩展到 0.5-100GHz)。

如果考虑视距通信 (LOS) 使用 UMa 模型, 路径损耗公式如下:

$$\text{Path loss} = 28.0 + 22 \cdot \log_{10}(d) + 20 \cdot \log_{10}(fc) \dots\dots\dots(2)$$

使用上述公式可以计算出路径损耗; 只需输入“d”--发射机与接收机之间距离 (小区半径) 和“fc”--中心频率; 如果小区半径为小区最远覆盖距离, 则路径损失与最大允许路径损耗(MAPL) 相等。

当获得路径损耗值就可以计算接收信号电平:

$$\text{RSRP} = \text{RS} - \text{Path loss}$$

d--发射机与接收机之间距离 (小区半径)

fc)--中心频率 rs - 小区参考功率

3.3 镜像射线追踪模型的建立

近年, 随着移动通信技术的迅速发展, 使用移动通信网络的用户越来越多, 为了满足不断增长的业务需求, 开始出现了室外微蜂窝、室外微微蜂窝、室内微微蜂窝的移动通信系统, 系统的通信信号从窄带信号变为宽带信号, 发射天线的高度也比较低。传播环境和电波频率的变化导致传统的经验模型不再适用。此时, 就需要其他可靠的预测工具来电波传播的参数, 例如路径损耗或者时延扩展。射线

追踪算法是一种可靠的电波传播参数预测工具。它可以解决宽带信号多径效应引起的码间干扰问题。它可以把收发天线方向图的影响加入到预测计算中。射线追踪的原理是基于 GO 和 GTD、UTD。在离源单位距离处, 波前面被均匀地划分成若干个小单元, 每个单元用一根射线代替其影响, 利用 GO 理论描述射线在传播过程中遇到障碍物发生发射和折射的现象, 利用 GTD 和 UTD 理论描述射线遇到棱边发生绕射的现象。在追踪每条射线之前, 首先必须获得所在地的地理信息, 然后在源处发射若干条射线, 最后开始射线的追踪。每条射线在传播过程中可能会发生多次反射和绕射, 为了确定这些反射面, 就需要进行射线与面的相交测试。如果该射线的电场值低于门限值或者反射次数大于设定的值, 或者射线离开了预测区域, 那么就停止对该射线的追踪, 进行下一条射线的追踪, 直到所有射线追踪结束。当所有射线的追踪完成之后, 就进行射线接收, 然后计算接收点处的接收功率和电磁场强度。所以, 预测区域的地理信息和射线与面的相交测试是射线追踪的两个重要部分。

获取了建筑物数据后, 开始创建镜像射线追踪模型, 镜像理论将反射线和绕射线分别看作是由虚拟源产生的, 因此整个传播环境中的射线就相当是由三类源所产生的, 它们分别是发射源、镜像源 (反射源)、以及绕射源, 城市环境下楼宇密集, 覆盖范围大大缩小, 电波是在建筑物的“峡谷”当中传播, 建筑模型以规则的长方体为主, 电波经过绕射后, 其场强与经过建筑物多次反射和绕射的路径相比, 往往可以忽略, 因此本方案只考虑电波的直射和反射计算。

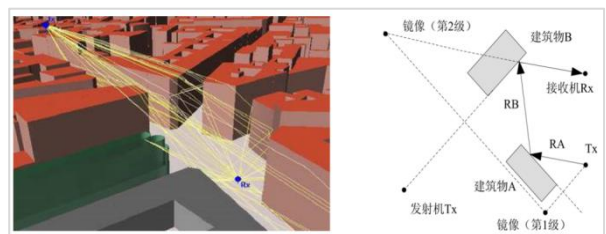


图 10: 射线追踪路径示意图

通过 Python 面向对象的特性, 分解对象模块:

- 基站集合: 基站->小区->射线->端点(发射点与接收点)

- 建筑物集合：建筑物→侧面与顶面→面端点：
- a) 一个规则建筑拆分 4 个侧面和 1 个顶面发射点与建筑有交点，则取与发射点最近的面为反射面进行计算。
- b) 发射点在反射面的镜像点和射线与反射面交点确定反射线的方向，重新以交点为发射点创

- 建新的射线。
- c) 多次反射信号强度必然衰减，射线中加入叠加距离属性，每反射一次，即累计了上一次镜像与交点的距离。
- 创建 5 大类模块（点、线、面、射线、基站）如下：

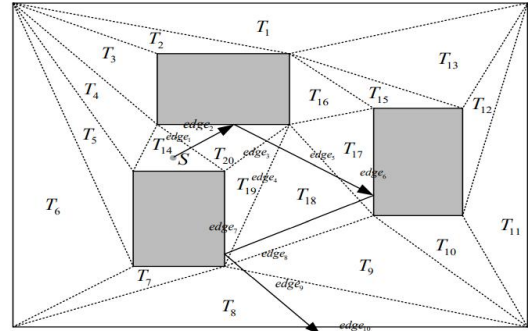
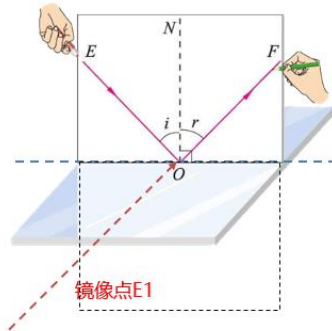
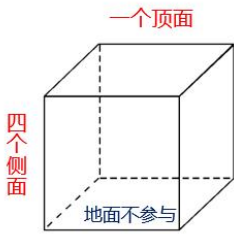


图 11：建筑物模型

图 12：镜像射线反射示意图

图 13：镜像射线在建筑物群的直射和反射示意图

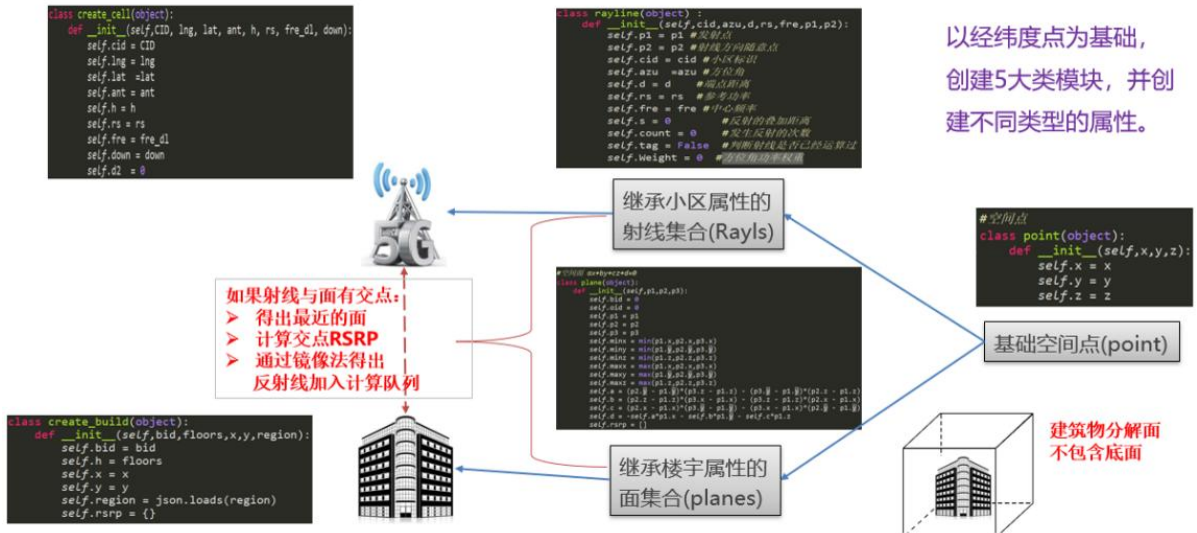


图 14：射线追踪模型创建的 5 类对象

通过流程图研发各功能模块和应用参数信息：

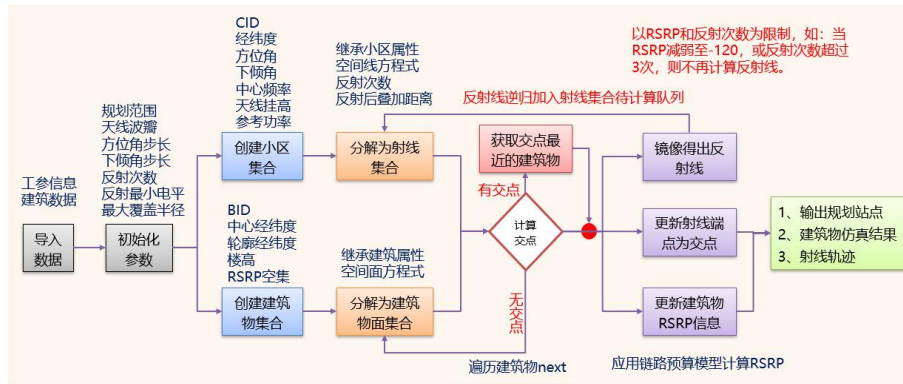


图 15: 射线追踪模型数据导入及处理流程

采用程序流程图方法对射线轨迹过程建模:

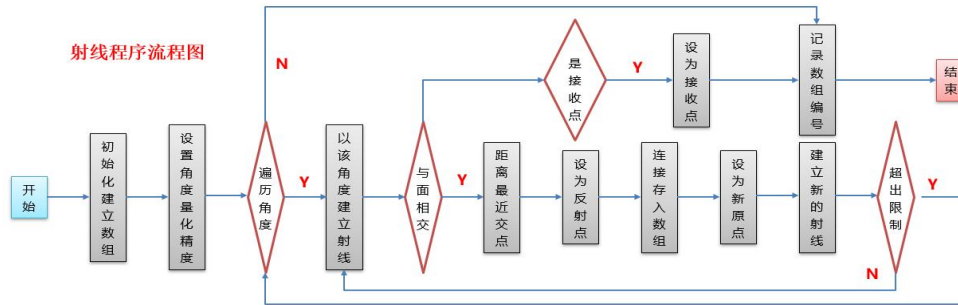


图 16: 射线追踪模型程序节点模块流程图

对过程数据调试:

小区集合: 小区对象属性

射线集合: 射线对象属性同时继承了小区属性

建筑集合: 建筑对象属性

每条射线到达建筑物电平: 精确至楼层的RSRP电平值

图 17: 射线追踪模型过程数据跟踪示意图

模型建立后, 准备相关数据进行测试, 工参表头如下, 其中 ant 为方位角, h 为天线挂高, rs 为参

考功率, down 为天线下角。

CID	lng	lat	ant	h	rs	fre_dl	down
5318231	118.5954	24.8864	110	37	15	1850	10
5318232	118.5954	24.8864	270	37	15	1850	5
5318233	118.5954	24.8864	30	37	13	1850	5
5318651	118.593	24.89551	30	36	15	1850	5
5318652	118.593	24.89551	190	36	14	1850	7
5318653	118.593	24.89551	270	36	15	1850	4
5318791	118.592	24.88794	20	28	12	1850	0
5318792	118.592	24.88794	210	31	15.2	1850	0
5318793	118.592	24.88794	320	31	15	1850	0
5318802	118.6003	24.90269	230	47	13.2	1850	5
5318841	118.5984	24.89094	40	42	14	1850	4

表 3: 射线追踪模型所需工参格式表

建筑物数据如下, x,y 为建筑物中心点, region 为建筑物轮廓经纬度:

BID	Floors	X	Y	region
0	6	118.6874	24.79056	[[118.687486, 24.790634], [118.687449, 24.790478],
1	6	118.607	24.79127	[[118.606889, 24.791654], [118.606954, 24.791566],
2	6	118.6213	24.79159	[[118.621396, 24.791542], [118.621321, 24.791513],
3	6	118.6077	24.7916	[[118.607742, 24.791605], [118.607619, 24.791522],
4	6	118.6073	24.7916	[[118.607393, 24.791542], [118.607356, 24.791513],
5	6	118.6211	24.79174	[[118.621197, 24.791761], [118.62101, 24.791688], [
6	6	118.6229	24.79169	[[118.623041, 24.791642], [118.622944, 24.791594],
7	6	118.5664	24.79173	[[118.566324, 24.791794], [118.566378, 24.791751],
8	6	118.5929	24.79176	[[118.592994, 24.791821], [118.592958, 24.791772],
9	6	118.6139	24.79181	[[118.613961, 24.791825], [118.613854, 24.791728],
10	6	118.6211	24.79186	[[118.621144, 24.791859], [118.620999, 24.79181], [
11	6	118.621	24.79196	[[118.621015, 24.791907], [118.620967, 24.791898],
12	6	118.5956	24.79196	[[118.5957, 24.791928], [118.595577, 24.79187], [118
13	6	118.6065	24.79197	[[118.606589, 24.79198], [118.606476, 24.791902], [

表 4: 射线追踪模型建筑物数据表

以某小区为例, 该小区正大方向受高楼阻挡, 信号只在前面几幢楼宇反射, 后面的低矮建筑受阻无法覆盖到, 实际主覆盖区域为旁瓣覆盖方向:

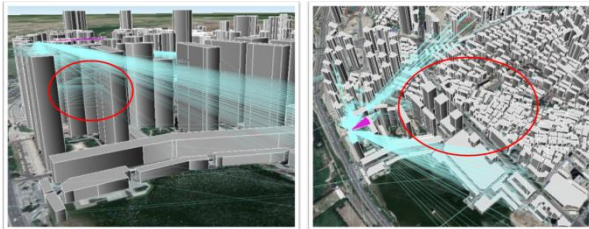


图 18: 射线追踪模型单小区验证

对一小片区域进行射线追踪计算, 生成射线过程如下, 射线追踪法输出的可视化图层, 每条射线均可追溯其源, 每栋建筑物有综合电平值和每个楼层的电平值:

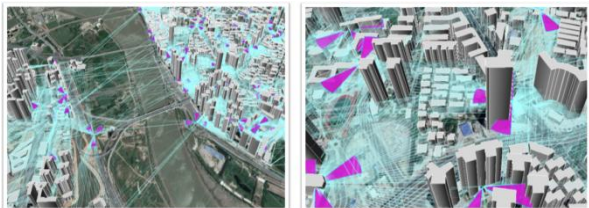


图 19: 射线追踪模型区域验证

3.射线追踪模型的应用实例

3.1 评估建筑的信号覆盖情况

通过 5G 链路预算和电平的计算方法, 应用射线追踪模型, 对建筑物的覆盖进行评估, 并按电平值进行 GIS 化分类, 直观的表达建筑区域覆盖情况:

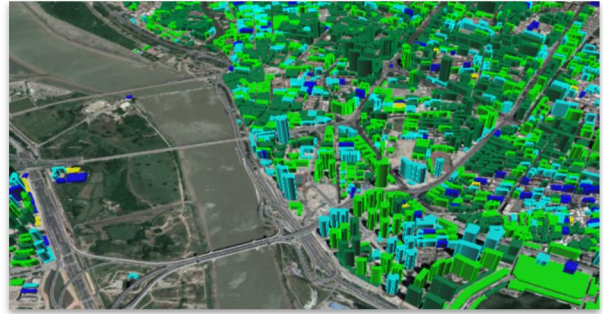


图 20: 射线追踪模型建筑物覆盖评估

同时射线追踪模型还可对平面栅格块进行覆盖计算, 在一个评估区域创建栅格图块, 计算每个栅格的覆盖电平, 进而评估栅格所属场景的覆盖情况。

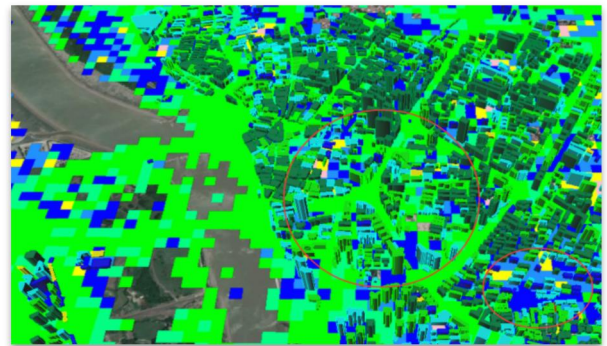


图 21: 射线追踪模型建筑物+栅格地面覆盖评估

3.2 识别建筑区的主覆盖小区

应用本射线追踪模型, 从天线发射的信号以射线表示, 每条射线的属性信息均包含有基站小区的相关信息(CID,小区名, 功率等), 一个建筑物可能接收很多来自不同小区的信号, 因此设置一个门限值, 当电平>-112dbm, 则认为有效覆盖射线, 在这些有效射线中, 电平最大的或取电平值前三个为主覆盖小区, 其它为可能覆盖的小区, 这样即可识别出每栋建筑的覆盖小区。

➤ 根据小区用户数研究用户密度

通过本模型算法，识别建筑物的覆盖小区，提取小区用户接入指标，在 GIS 中渲染用户密度图，通过获取用户密集区域进行网络保障和新业务推广，评估用户分布，指导网络规划。

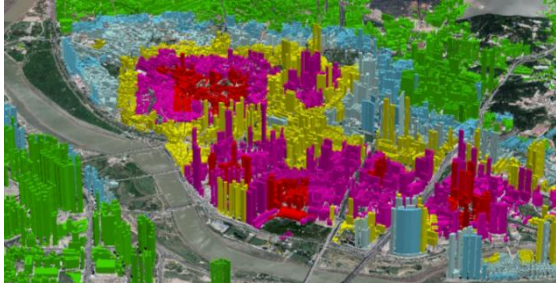


图 22: 利用模型识别的覆盖小区进行用户密度区域分析示意图

- 根据识别的覆盖小区评估建筑级别感知指标（终端侧 RTT、流量等）
通过识别出的覆盖小区，结合用户感知指标，分析用户感知差的区域进行重点优化，提升用户感知度，宣传品牌形象。



图 23: 利用模型识别的覆盖小区进行用户感知区域分析示意图



图 22: 利用模型识别的覆盖小区进行流量区域分析示意图

合理利用射线追踪模型和互联网地理信息，可有效的参与对网络指标的评估，了解地理位置的价值，凸显网络问题，为网络规划和优化提供参考。

4.总结

电波的传播特性参数预测在无线系统规划中，具有非常重要的作用。随着对无线网络通信速度的要求越来越高，无线系统开始往高频、宽带、微小小区发展。无线系统这些新的特点使得传统的经验模型预测传播特性变得不适用。射线追踪技术作为一种电波传播特性预测技术，它能够很好的解决高频、宽带无线系统电波传播特性预测的问题，计算速度快，预测精度高。

参考文献

- [1] 3GPP TR 38.901 V14.3.0 Study on channel model for frequencies from 0.5 to 100 GHz (Release 14) [S].2017
- [2] 朱秀丽, 周治武, 李静, 等. 网络矢量地图瓦片技术研究 [J]. 测绘通报, 2016(11)
- [3] 章谦骅, 章坚武. 一种新的射线追踪方法-测试镜像跟踪法. 微波学报 2014 年 6 月
- [4] 张艳芳, 马力, 邹澎, 程炜. 反向射线跟踪的三维路径搜索方法. 电子设计工程 2011(4)
- [5] 靳海亮, 李留磊, 袁松鹤等. 一种用于三维城市建筑物的可视域分析算法[J]. 测绘通报, 2018(1)
- [6] 移动通信系统终端射线跟踪定位理论与方法 /袁正午编著. 北京: 电子工业出版社. 2007
- [7] 董金梁. 改进射线跟踪法效率的新方法. 微波学报, 2004, 22(6)

一种全方位立体监控告警方法的研究与应用

郭龙缘

(中国电信股份有限公司福建分公司)

摘要: 随着数字化转型升级进程加快,企业 IT 系统架构越来越复杂,业务需求也更加丰富,各业务场景微服务化随应用需求不断拓展,为了保障业务系统的稳定性,近年来各行各业都逐步将业务应用迁移至云服务器,致使云上系统的稳定性受到空前关注。面对海量告警数据以及复杂的应用场景如何快速、准确的做好监控告警成为了重要的研究课题。本文通过借助计算中心超强的运算能力,构建出一套全方位立体监控平台。经过实践证明,该平台可以较好的应用于复杂业务场景下的监控并且告警通知及时率、准确率均满足使用要求。

关键字: 数据处理、并行计算、多场景一体化监控

引言

随着数字经济的发展,数字化转型加速推进,算力作为核心生存力的时代已经到来,2022 年 3 月国家发改委、中央网信办等 4 部门联合印发通知,并规划了 10 个国家数据中心集群,全年启动“东数西算”工程。在这样一个背景下,行业应用、系统运营、网信安全等领域对算力的需求日益增加,产生了较多的最佳实践案例,本文介绍是算力在系统运营领域的应用案例[1-2]。

1.IT 系统监控运营挑战

随着数字化系统规模快速发展,系统的稳定性保障愈发重要,传统的监控手段难以洞察复杂云化系统运转情况,同时大量的监控数据无法快速计算告警信息,导致监控信息不准、不及时,影响客户感知,存在以下几方面挑战[3]。

1.1 监控设备多、用户访问大

当前众多的监控平台面对超大规模监控设备数量以及用户访问量徒增均会出现普遍的问题,一旦运行数据量达到规模级别之后,业务运行软件会逐渐变慢和不稳定,即使增加服务器或内存数量,也无法提高系统的性能[4]。

1.2 分布式云化架构的监控难度空前加大

单体应用的监控难度较低,通过指标、日志可以实现监控和排障,但是系统云化后,特别是应用容器、微服务等技术后,一旦出现问题,无法快速确认问题的发生环节、运行设备[5]。

1.3 基础设施容易成为“背锅侠”

云上网络、主机、PaaS、SaaS、客户端成为一个有机整体。一旦发生故障时,由于缺乏全方位监控手段,难以快速定位真实故障原因,往往由网络和主机来“背锅”。对外发布的原因也经常是“由于网络原因”、“由于性能不足”等。

1.4 故障预警提示慢,难定位

各类网管监控系统遍地开花,B 域、O 域、M 域以及 ITM 域等各业务系统,均建设自己的网管监控系统,并且监控数据不对外开放,无法相互串联分析,更无法对场景化业务建立一体化监控。而导致故障无法预警或者告警不及时,一旦出现故障,更是无法及时定位问题。

2.全方位立体监控研究

全方位立体监控平台通过整合企业内各类 IT 系统业务监控平台,打破现有信息分裂壁垒,构建一套数据、监控、资源、能力集中的监控平台,实现 IAAS、PAAS、SAAS 三层的业务监控指标及采集数据的统一管理。通过打通和融合内部多样化、大

数据量的多元数据，实现对云化架构的深度监控，利用对指标、链路、日志形成观测体系，相互关联，达到以场景为导向的可视化编排目标[6]。最终通过计算中心快速实现采集数据分析、运算，获得预警或者告警信息，及时通知、快速定位故障点，从而满足不同客户运维需求。

2.1 一体化智能可观测

针对云原生架构复杂性，业务故障诊断难的痛点，本文构建全局业务可观测平台，如图 1 所示。全方位立体监控平台通过对云、网、系统、应用、组件的全栈指标采集，将采集源数据统一纳入大数据湖，可通过界面配置，实现端到端监测，保障业务运行稳定。



2.2 监控告警全生命周期流程

建设开放的监控平台，实现系统/组件、容器、云平台等对象的监测采集和告警检测，同时支持将三方系统的性能指标或事件接入到数据存储中心。

借助计算中心可快速实现数据处理，形成各类指标值，最终通过告警中心进行告警归并、屏蔽、汇聚、跟踪、确认以及通知，实现事件处置的闭环管理和统一分析、视图展示。



2.3 采集数据标准化管理

构建全局的 IPS (IAAS、PAAS、SAAS) 三层的指标、告警模型库, 各类监控平台提供基础的监控采集能力, 将纳管的网元和监控指标注册到全方位立体监控平台, 通过 kafka 实时向全方位立体监控平台推送指标告警及指标性能数据[7-8]。

全方位立体监控平台统一对数据进行校验和转换处理, 实现将汇聚分散的 IAAS、PAAS 及 SAAS 三层告警数据进行统一的标准化管理。同时本平台提供丰富多元的数据操作接口, 包括“标准 SQL 接口”、“自有的 API 接口”等, 为上层场景应用提供高质量数据服务, 便于用户进行数据运营分析使用。

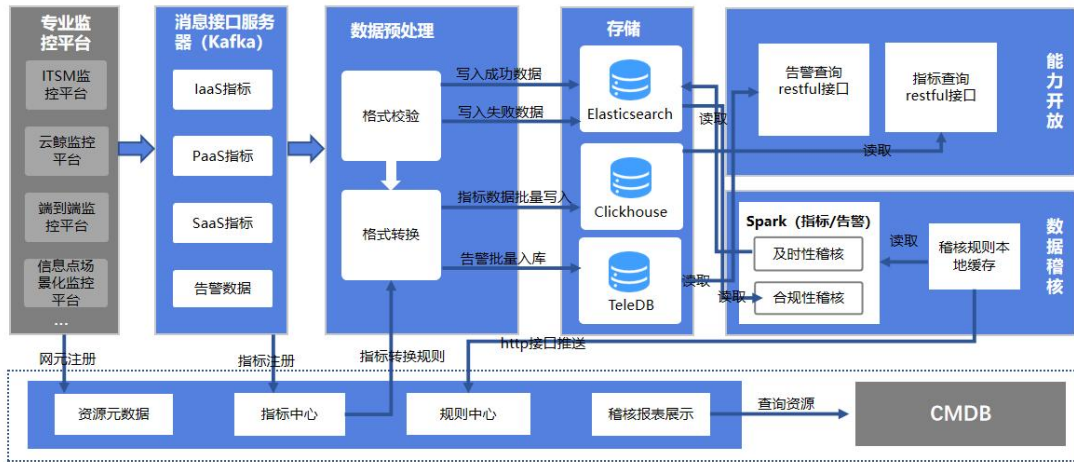


图 3 数据标准化管理流程图

2.4 组合指标管理

针对当前复杂的业务规则、业务场景, 监控告警不再是单一指标即可判断业务是否受到影响, 是否需要告警。为此本平台提供灵活可配置的指标组

合计算能力, 基于数据湖中已有的基础数据, 用户可快速完成业务感知的复合监控指标进行配置, 简化了复合指标的开发流程, 开发周期从天级缩短到分钟级。



图 4 组合指标设计说明图

3.应用结果

该平台已完成建设并应用于多家企业,均取得较好的应用反馈。其中最典型案件是将某企业数个监控平台的 278 个监控指标能力纳管,整理了 119 个监控指标规则,监控规则标准化率达到 83%;完成 3675 个资源的自动监控接入,实现资源监控规则自动化覆盖率达到 70.4%,告警及时率达到 96.1%。全年故障数量年同比下降 50%,故障时长同比下降 60%,IT 类告警在告警板停留时间下降 20%。

IAAS、PAAS、及 SAAS 三层监控运营能力显著提升,助力监控一体化跨域、跨专业的协同异常分析、根因定位、影响分析、助力运维管理一体化实现各专业运营团队之间的协作,提升各方对 IT 运行现状认识的一致性,提升故障效率、定位与解决的效率,助力 IT 运维质量提升的发展。

3.1 统一数据湖 IAAS 的监控能力

目前已经能够汇聚 IT 的主机监控 800 台,并能够依据主机上的 CPU、内存、磁盘利用率、进程等配置监控告警策略,覆盖 IT 上云的全业务基础设施监控。

3.2 统一汇聚数据的 PAAS 的监控能力

统一汇聚已纳管全省云翼组件的监控(ctgcache、telepg、teledb、mq 等),具备实时对云翼组件的监控预警,生产业务应用相结合,具备联动业务告警影响范围的预警和通知。

3.3 跨系统场景化监控能力

在汇聚传统 IaaS、PaaS、SaaS、客户段监控外,加入网络监控,形成云网立体监控体系,可实现跨平台场景化监控,打破枪痛的孤岛式监控。

4.结束语

本文基于“全量 CMDB 资产纳管”数字自智

的运营理念,践行“汇聚、预测、快速定位、快速处理”的监控治理。与传统的烟囱式专业网管不同,本案例数字化转型的思路针对 IAAS 监控、PAAS 监控、SAAS 监控的分散以及分布式云化架构的监控难度,引入“大数据湖”的概念方式,全面汇聚 IT 资产监控运维数据,通过告警数据采集、汇聚、处理、告警策略配置等多环节处理,实现统一、快速、直观、方便的监控能力视图,使得业务、系统、应用、资源的监控情况一目了然,及时发现未监控项和监控不合理项目,补缺补漏,实现 24 小时 360° 无死角监控。

参考文献

- [1] 程宝平,王欣,汪胜等.视联算力网络关键技术研究与应用[J].电信工程技术与标准化,2023,36(01):1-10.DOI:10.13992/j.cnki.tetas.2023.01.001.
- [2] 许琦敏.算法创新算力突围孕育 AI 新范式[N].文汇报,2023-07-08(005).
- [3] 王光宇.监控与告警平台的设计[J].中国科技信息,2018(24):37-38.
- [4] 王骏翔.一种面向多云资源和多云服务的监控告警方法[J].上海船舶运输科学研究所学报,2022,45(02):69-76+82.
- [5] 涂畅.云计算环境下计算集群监控告警系统的设计与实现[D].华中科技大学,2022.DOI:10.27157/d.cnki.ghzku.2022.000064.
- [6] 夏畅.基于 Zabbix 的可视化监控告警平台的研究[J].数据通信,2020(01):32-34.
- [7] 郝鹏海,徐成龙,刘一田.基于 Kafka 和 Kubernetes 的云平台监控告警系统[J].计算机系统应用,2020,29(08):121-126.DOI:10.15888/j.cnki.csa.007611.

一种利用数据行为进行 VoLTE 指标快速动态监测方法

郑银云 杨慰民

(中国移动通信集团福建有限公司)

摘要: 本文根据 VOLTE 指标的数据特点及行为, 构建了 VOLTE 动态指标监测和优化系统, 从时间维度上, 兼顾时间上的大的变化和微小累积变化, 进行 VOLTE 指标的动态预警和监测。同时利用质差排除的方法, 快速找到指标波动的原因, 实现对 VoLTE 指标的动态监测和快速发现问题点, 从而达到快速优化的目的。

关键词: VOLTE 语音; 动态监测; 累积微质差; 质差剔除

引言

VoLTE 的商用实现了语音业务由传统电路域向数据域的转变。VoLTE 是基于 IMS 实现语音控制, 全部承载于 4G、5G 网络之上, 在提供高清语音更短接续的同时也面临着端到端业务实现流程复杂, 涉及用户、终端、无线、传输、EPC、IMS、CS、数通多元协同, 高清语音特点为客户感知提升提出巨大挑战, 确保客户 VoLTE 语音业务感知的关键需要建立 VoLTE 端到端质量管理, 本文主要围绕建立 VoLTE 端到端指标监控体系展开, 展示了利用数据行为进行的 VOLTE 指标快速动态监测方法, 通过指标动态快速监测实现网络对 VoLTE 指标问题监测和优化, 达到保障网络可靠性提高用户感知的目的。

1. 利用数据行为进行 VoLTE 指标快速动态监测方法的优势

现有技术通过简单的数据统计以及门限比对, 可以得到指标的相对静态的情况, 但是由于数据行为随时间在不断变化, 利用传统方法, 存在无法发

现指标的微小变化, 存在未考虑指标从量变到质变的一种情况(细微变化无法瞬间看出劣化, 但是, 时间上的叠加后产生的影响需要考虑, 但是传统方法无法做到)。传统方法在进行指标监测时, 也没有进行数据多维特点的充分分析, 无法做到快速找到可能的 TOP 质差样本, 进行快速的优化。

本文提出的方法从时间维度上同时考虑大的变化和微小量变的影响, 通过统计时间轴上的指标微小量变的叠加, 得到一段时间指标发生的变化, 进而进一步对 VoLTE 指标进行动态的预警和监测。在发现问题的情况下结合多梯度的排除算法, 快速得到质差样本, 进行快速的问题发现和解决。因此, 与采用传统静态监测方法相比如下优点:

更全面: 多个角度共同监测, 不仅进行传统指标的传统监测, 同时动态应用数据行为挖掘潜在的累积的微小质差, 对网络指标异动监测更加全面。

更高效: 巧妙利用数据特点, 通过相应的剔除模型(排除法和归一化), 快速找到对指标劣化影响最大的样本, 从而达到快速处理快速优化指标的目标。

通用性强: 系统不依托与厂家告警, 受设备变

作者简介:

郑银云: 中国移动通信集团福建有限公司网络部网管中心员工

杨慰民: 中国移动通信集团公司首席专家; 中国移动通信集团福建有限公司网络部网管中心副总经理

化影响小,采用 VoLTE DPI 采集的数据,更具客观性。

2. 利用数据行为进行 VoLTE 指标快速动态监测方法的实现

2.1 总体思路

本文提出的方法从时间维度上,不仅考虑大的变化,同时考虑指标微小量变的影响,统计时间轴上的指标微小量变的叠加,即考虑数据行为在时间上的变化。统计一段时间指标发生的变化对 VoLTE 指标进行动态的预警和监测。同时寻找合适的维度,快速找到影响指标的因素:根据地市级维度中的小区维度,所有样本下的指标情况,及排除 top 小区(多个 top 梯度)后的指标情况,实现对 VoLTE 指标的动态监测和快速找到影响指标劣化的小区样本。

该方法的主要步骤为:1. 统计 VoLTE 指标情况,进行快速监测如果指标异常,则预警;2. 统计 VoLTE 指标的前后两天的变化,记录相应的情况 r_1 ,同时统计一个周期内的该 r_1 的变化,当 r_1 达到门限 k_1 ,则触发预警;3. 当发现异常,以小区维度切入,选取指标值 v_1 及相应的计算该指标的分母 v_2 ;4. v_1 和 v_2 经过处理得到综合影响因子 y_1 ,剔除相应 y_1 大的值(具体剔除的样本数量可根据具体情况设置动态门限);5. 对比不同剔除情况下的指标情况和未剔除情况下的指标情况,快速找到可能影响指标的质差样本。

2.2 详细实现思路

2.2.1 数据源

在 VoLTE 采集中利用分光技术实现多接口数据采集,采集接口包括:

- S1U(LTE eNodeB 与 SAE-GW 数据面)
- Sv(LTE MME 与 IMS eMSC 信令面)
- Mw(IMS PCSCF 与 I/S-CSCF 信令面)

主要数据包括:

S1-U: 语音业务数据

Mw: SIP 信令主接口与接续、注册等主要过程相关

- Sv: eSRVCC 切换相关信令

2.2.2 涉及业务过程和指标

用户 VoLTE 语音通话过程,可以描述为 5 个

阶段:附着阶段、注册阶段、接续阶段、保持阶段、通话阶段。该指标动态监测的方法可针对以上所有阶段对应的指标。

2.2.3 详细的处理过程

DPI 采集的相关接口数据作为输入,按照下面流程进行相应的数据处理如图 1 所示:

2.1 统计各指标情况,输出指标记为 kv ;

2.2.1 将 kv 作为参考样本存储起来记为 pv ;

2.2.2 将 kv 与相应的门限 m 进行比较

2.3 如果 kv 异常,则将监测结果输出;

2.4 将 3.1 中的 kv 与前期 3.2.1 中存储的 pv 值进行关联计算对比;

2.5 输出比较结果 r_1 ;

2.6 计算 r_1 的周期性特征;

2.7 判断是否异常;

2.8 如果异常,输出潜在的累积异常;

如果 3.2.2 中结果为异常,进入快速挖掘异常样本的流程:

2.9 选取指标值 v_1 及相应的计算该指标的分母 v_2 (根据实际情况,对整个指标影响的程度不仅体现在指标值上,同时该指标出现次数(基数)也同样重要)

2.10 对 v_1 和 v_2 分别进行排序得到排序的优先级 vp_1 和 vp_2 ;

2.11 考虑要对 vp_1 和 vp_2 进行评估联合使用,需要进行归一化处理得到 vg_1 和 vg_2 ;

2.12 将得到的 vg_1 和 vg_2 进行一定的拟合后得到相应的权重值最后计算得到综合影响因子 y_1 ;

2.13 根据综合影响因子 y_1 ,剔除 y_1 大的样本(具体剔除的样本数量可根据具体情况设置动态门限);

2.14 对比不同剔除情况下的指标情况和未剔除情况下的指标情况,快速找到可能影响指标的质差样本

基于 3.1-3.14 点的算法,完成了基于数据行为分析的 VoLTE 指标快速动态监测

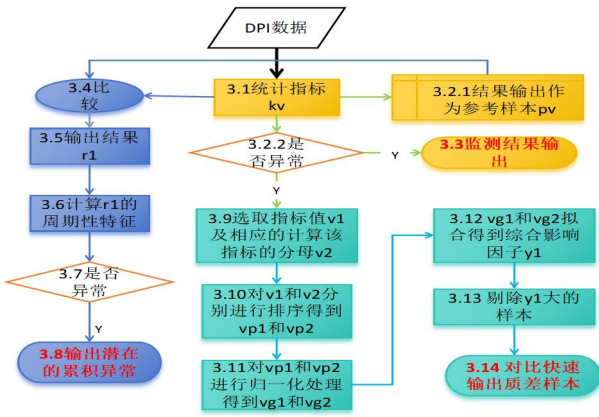


图 1 数据处理详细流程

3.利用数据行为进行 VoLTE 指标快速动态监测方法的应用开发和落地

该成果立足于解决一线生产维护中的问题，解决 VOLTE 指标的变化。随着 VOLTE 业务的出现，

VOLTE 业务质量成为关注的重点，指标监测是必要的手段。该成果从时间维度上，不仅考虑大的变化，同时考虑指标微小量变的影响，统计时间轴上的指标微小量变的叠加，即考虑数据行为在时间上的变化。该成果得到的结果成功落地，目前已经在支撑生产应用。可以快速发现指标异常，发现网络的微小劣化并快速定界后进行处理。

该成果目前按照功能模块宏观展示地市维度指标及相应的周月趋势图，可以查看关键指标及指标变化趋势，发现指标是否异常，同时通过引入联合评估、归一化等算法可以快速的发现引起指标异常的小区，提高排障效率。

该成果中地市级指标主要为端到端接通率、V2V 接续时延、eSRVCC 切换成功率、RTP 上行丢包率、RTP 上行 MOS、RTP 下行丢包率，同时呈现了质差小区数量，通过点击质差小区数，可快速获得质差小区，进行相应的优化。

地市关键性能指标分析

地市	端到端接通率		v-v接续时延		端到端掉话率		esrvcc切换成功率		端到端上行MOS		端到端下行MOS		端到端上行丢包率	
	指标(%)	质差小区数	指标(秒)	质差小区数	指标(%)	质差小区数	指标(%)	质差小区数	指标	质差小区数	指标	质差小区数	指标(%)	质差小区数
全省	93.21	311	2.18	37	0.12	77	97.68	28	3.72	26	3.65	1	0.13	92
福州	93.12	55	2.20	5	0.11	14	97.22	7	3.72	5	3.66	0	0.13	12
厦门	92.89	19	2.22	8	0.17	7	97.93	3	3.72	0	3.66	0	0.15	10
宁德	93.11	30	2.09	0	0.12	1	97.92	5	3.75	2	3.66	0	0.09	5
莆田	93.29	14	2.11	3	0.10	4	97.74	3	3.71	1	3.62	0	0.13	2
泉州	93.41	34	2.15	3	0.10	8	97.53	5	3.73	1	3.67	0	0.12	8
漳州	92.97	42	2.25	4	0.11	10	98.19	3	3.70	9	3.62	0	0.15	25
龙岩	93.53	44	2.20	5	0.11	13	97.62	1	3.72	1	3.65	0	0.11	8
三明	93.58	37	2.16	7	0.13	12	98.06	1	3.65	2	3.59	1	0.12	6
南平	93.49	36	2.15	2	0.11	8	97.54	0	3.68	5	3.62	0	0.13	16

图 2 指标情况和质差小区展示

对于地市级指标的周、半月、月数据进行趋势分析，可方便的展示出指标的波动情况，且通过排

除 TOP 质差小区可直接获得 TOP 质差小区对直接的影响，从而快速锁定质差小区以方便优化处理。

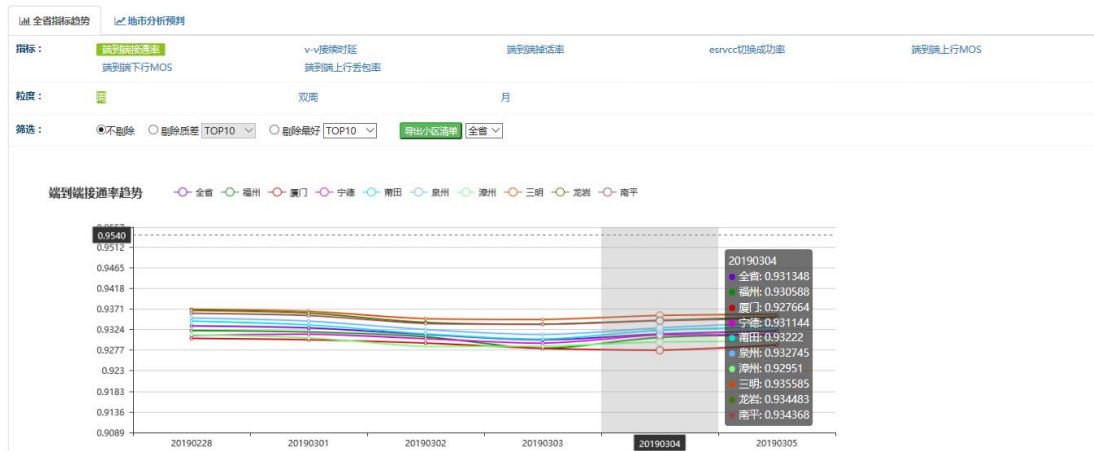


图 3 指标的趋势监测和质差排除展示

虽然对指标的趋势可通过图直观展示，但是对于指标的累积微质差的变化，有时仅趋势图难以看出，通过程序设计利用连续变化趋势，设置动态门

限，来发现微质差的波动，如果有存在异常，则直接进行预警派单，通过工单来管控处理。

预警ID	预警日期	预警内容	预警原因
		近期无VoLTE性能指标预警!	

图 4 预警信息展示

4. 结论

本文设计的 VoLTE 指标快速动态监测方法充分利用已有的 VOLTE DPI 数据资源，利用大数据挖掘思维和动态学习的思路，将 Mw、S1u、Sv 接口的数据按照算法计算出相应的指标，同时利用联合评估、归一化、微质差累积等思路，将该项目的动态指标监测和优化完美呈现。该方法实现的应用对 VOLTE 语音的网络优化起到良好的系统支撑，目前已经在全省推广使用，通过该方法的落地使用后，VOLTE 指标的优化得到明显提升，原 V2V 接续时延在 3s 以上，目前 V2V 时延降低到 2.3s 左右。同时，通过微质差累积的使用，现网成功发现多次的网络隐患。对 45G 语音网络的隐患发现和优化起到良好的支撑作用。

参考文献

- [1] 中国移动通信集团福建有限公司,中国移动通信集团公司.数据动态监测方法、装置及电子设备[P].中国专利: ZL 2019 1 0289978.7,2022.07.01
- [2] 杨慰民. 基于大数据的移动互联网主动运维理论和实践进展[J]. 大数据, 2016(6).
- [3] 庞亮,许新宇. VoLTE 端到端质量优化方法研究 [J]. 2017 年内蒙古通信学会,2017(10).
- [4] 李飞,曹雷雷,付永振.人工智能在 VoLTE 语音质量评测优化中的应用与研究 [J].中国科技纵横 2019 (5)

基于机器学习的基站节能方案研究

黄炳塔 董帝焯 杜丕加 吴晶晶

(中国联合网络通信有限公司泉州市分公司)

摘要: 通信技术的发展日新月异,带来极速上网体验的同时,基站设备的耗电量同步剧增。以现阶段来说,5G 设备的耗电量是 4G 设备的 9 倍左右,基站规模也不断扩大,因此研究节能方案对于运营商势在必行。以基站为例,节能的原理为在空闲窗口采取措施,关闭耗电最大的是射频设备,措施包括载波关断、通道关断、符号关断三项措施,达到节能减排的目的。为使判断空闲窗口更加智能高效,使用 LightGBM 与 Prophet 算法分别应用于节能门限与时间窗口的预测,目的在于寻找到节能开关的 KPI 动态门限与开启时间。

关键词: 耗电量, Lightgbm, Prophet, 节能门限, 时间窗口

1 背景

根据中国联通 LTE 网络组成结构,机房基站设备(包括基站设备、空调、传输和监控等)消耗电费占整网的 83%,其中 42%的消耗来自基站设备。以当前主流的分布式基站为例,3 扇区覆盖的情况下,射频设备能耗占有 80%的基能耗;随扇区数/频谱资源上升,射频设备能耗占比同步上升。射频设备的能耗主要是功放:除去收发信机/基带处理单元/接口板/天线阵子等基础能耗外,功放占据了射频设备 70%以上的能耗。

网络能耗看基站,基站能耗看射频,射频能耗看功放。当前节能技术主要有三个措施,分别是载波级关断技术、通道级关断技术与符号级关断技术,其中符号关断技术是根据低话务时段,无数据发送时,可以通过短时关闭功放(仅保留 CRS 发射)以降低实时功耗。本项目也是基于此考虑,将项目成果应用于符号关断技术。

传统的节能措施,存在如下四个缺点:

1、缺乏灵活能动性:同区域甚至整网参数统一设置;无场景识别,缺乏方案适应性;与真实站点话务量波动匹配度低。

2、数据分析冗长繁琐:传统进行 KPI、话务量等海量数据分析,需要大量外场支撑经验;数据相关性高导致传统分析出错概率高,耗时长,且易出现千人千结论。

3、缺乏可持续演进性:容量、站型或者邻区出现变化时,需要重新规划方案;多层网络协同性差,无法兼容当前网络演进。

4、响应延后无实时性:网络性能实时监控性差,故障处理出现滞后性;网络中难以平衡节能收益与网络性能,部分影响可能导致节能完全关闭。

正是由于此四个缺点,网络更需要一个全智能、可自主策略调整的节能方案。方案功能应包括:

1、对节能时间段智能预测,基于话务历史数据,区分工作日及节假日,准确预测话务的时间分布,并且能够创建多个节能激活时间窗。

2、参数门限智能调节,按小区动态学习最佳节能拐点,如下行平均 PRB 利用率, RRC 最大激活用户数等。

3、可智能迭代,根据历史数据,动态调整各项参数。

2 符号关断功能

2.1 符号挂断功能简介

LTE 系统在实际通信过程中,基站不是任何时候都处于最大流量的状态,所以对于子帧(Sub Frame)中的符号(Symbol),不是任何时刻都充满了有效信息。

未开启符号关断功能,任何时刻 RRU 的功放电源均为打开状态。

开启符号关断功能后，基站调度器根据业务繁忙程度，主动将下行数据调度到指定的符号上，在剩余的无有效信息传输的符号时间，关闭功放电源，达到节能目的。例如，在配置 CP=Normal、CFI=1 时，Port1 通道的 Sub Frame 中 Symbol (0~13) 最大可关闭的集合为 {S1,S2,S3,S5,S6,S8,S9,S10,S12,S13}。

在业务量最高时，所有符号时刻功放均打开，此时与未打开符号关断功能的能耗一致；在业务量最低时，只在必要的业务信号（如广播信号、同步信号等）发送时间打开工作电源，其余时间功放电源关闭，此时最节能。

关闭与开启符号关断功能时，数据与 PA 电源开关状态如图 1 和图 2 所示。

关闭符号关断功能的 PA 电源开关状态时，可以看到根据 UE 业务，数据被随机调度到子帧中的各符号上，此时 PA 电源均为打开。

开启符号关断功能的 PA 电源开关状态时，业务数据被集中调度到某些符号上，在剩下的符号上无数据调度，此时可以关闭 PA 电源实现 DTX（不连续发射），开启符号关断功能后基站处于 DTX 节能状态。

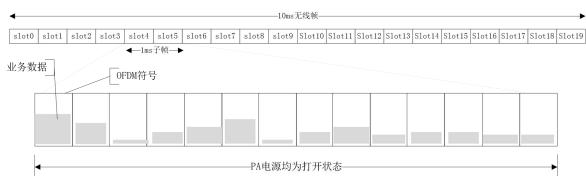


图 1 关闭符号关断功能的 PA 电源开关状态

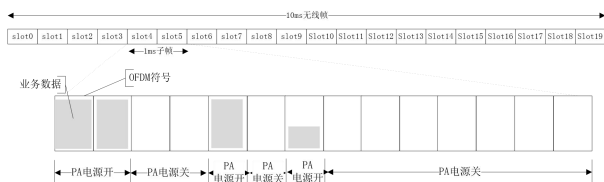


图 2 开启符号关断功能时的 PA 电源开关状态

2.2 符号关断开启流程

如图 3 所示，为符号关断的开启流程。

- 1、OMC 同步 SON 节能策略信息到节能小区所在的 eNodeB，eNodeB 开启节能功能；
- 2、当节能策略配置的节能开始时间到达后，

进入步骤 3，否则继续等待；

- 3、节能开始时间到达后，上报监控，然后进入步骤 4；

- 4、开始进行智能符号节能判决：当检测到 RB 利用率与 rrc 激活用户数小于门限，则进入步骤 5，当检测到 RB 利用率与 rrc 激活用户数大于门限，则进入步骤 6。

- 5、当通知 RRU 关闭子帧，其中关闭个数不能超过最大可关闭子帧个数然后继续周期检测。

- 6、通知 RRU 打开关闭的所有子帧，然后继续周期检测。

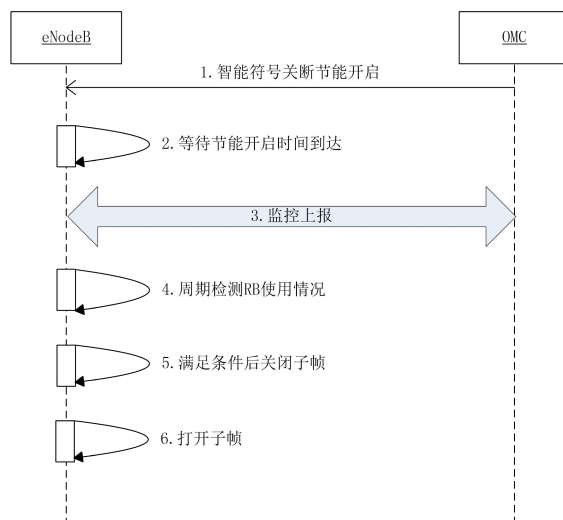


图 3 符号关断开启流程图

2.3 节能时间结束符号关断关闭流程

如图 4 所示，为符号关断的开启流程。

- 1、当策略配置的节能时间结束后，进入步骤 2；
- 2、打开智能符号关断关闭的子帧；向 OMC 上报监控，在网管上呈现小区退出智能符号节能。

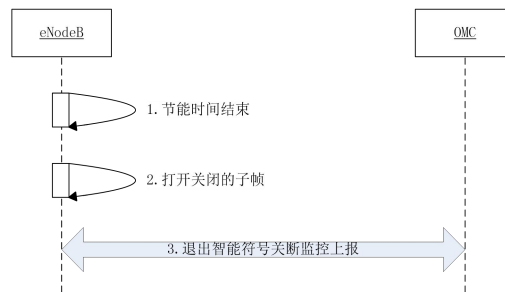


图 4 符号关断关闭流程图

3. 智能节能系统原理

如表 1 所示，传统节能方法设置固定的门限、时间窗口，未结合用户感知。而智能节能方法，基

于历史数据建立数据库，能够建立以小区为粒度的策略控制，并结合用户感知，实现降低 LTE 网络能耗且不影响用户使用感知的目标。

类目	序号	传统节能方法	智能节能系统
节能门限值	1	门限值根据网优人员经验主观设定或者简单分析数据所得	历史数据建立数据库，通过决策树算法学习，并多次验证得到的门限模型，因此更精准
	2	全网小区设置统一门限，未考虑不同场景差别	利用机器学习算法每个小区都有一个最佳节能门限值，更精细化和实时性
	3	门限值固定，除非人工修改	门限值根据历史数据学习动态调节更符合用户和场景的变化规律
	4	未考虑用户感知	用户感知和负荷关联分析，两者兼顾
时间窗口	1	避免对网络的影响，一般设置在凌晨时段 0~4 时	通过机器算法，1 天 24 小时，只要满足条件即可动态调整。
	2	通过网管指标统计来设置时间窗口，指标最小延迟 15 秒，容易出现统计的时间窗口与实际偏差，导致节能时间误差。	通过精准预测节能时间窗口，实时动态调整，提高节能应用精准率。

表 1 节能方法对比

如图 5 所示为智能节能系统原理框图，分两条主线并行，一为基于 Lightgbm 算法，寻找到节能最佳门限，二为基于 Prophet 算法，探索节能最佳时间窗口。最终输出每个小区的节能策略，并自动应用。

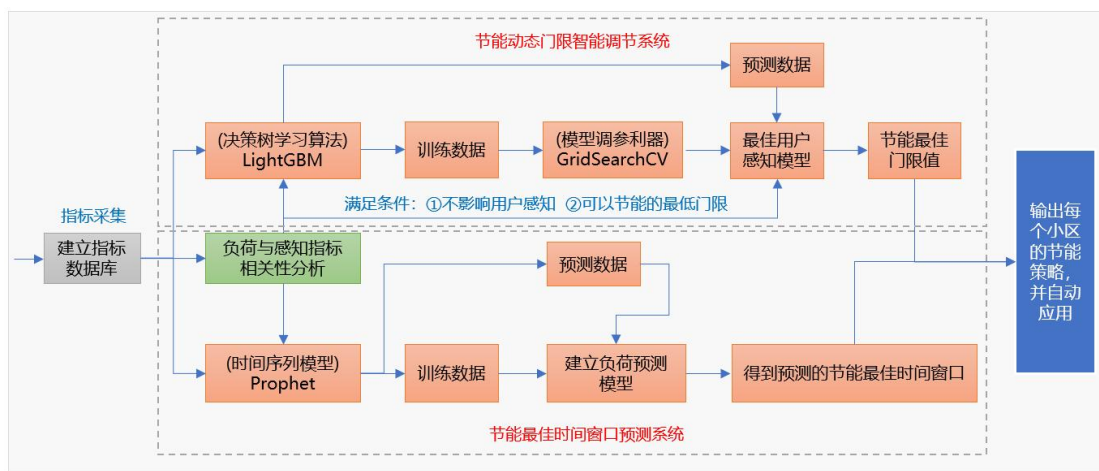


图 5 系统原理框图

4. 动态门限预测

符号关断的两个门限设置为下行 prb 资源利用率与 rrc 最大激活用户数。本次采用 LightGBM 进行回归预测，GridSearchCV 进行调参，通过数据集的训练，得到最佳模型参数。目的为输入所需的优

良 KQI 指标，得到两个门限临界值。

4.1 数据清洗

本次用作训练的数据集日期为 2012 年 3 月 1 日至 2021 年 2 月 28 日。选取基站 ID 为 531885，小区号为 1 的小区作为样本。训练集数据标签如表

2 所示。

数据清洗规则：

- 1、删除有值为零的行，包括所有指标
- 2、删除有值为空的行，包括所有指标
- 3、训练集训练时去掉日期列、唯一索引列、

FDD 下行信道 PRB 资源利用率与用户面最大激活 UE 数

- 4、训练目标为 FDD 下行信道 PRB 资源利用率与用户面最大激活 UE 数最终训练集为 12759 行。

页面响应时长	页面显示成功率	终端侧 RTT	三次握手	初始缓冲时长	视频流媒体有效下载速率	空口下行业务流量 (GB)
web_response_time	web_sussess_time	UE_RTT	three_handshake	Initial_buffer_time	video_speed	flow(G)
304.389164	0.953502	158.413582	36.469972	987.882438	2750.914373	9.28928

表 2 训练集数据标签

说明：时段为小时粒度，由于 15 分钟粒度指标缺失严重，故选取 1 小时粒度。指标中英文对照如表 1 所示，唯一索引的格式为日期-时段-小区号，其中日期为常规格式的日期，因为 KPI 与 KQI 指标来自不同平台，字段不一，故设置唯一索引列，连接 kpi 与 kqi 指标。

本次做预测目标的字段如表 3 所示，需预测两个目标。

用户面最大激活 UE 数	FDD 下行信道 PRB 资源利用率
max_ue	PRB_usage
39	74.42

表 3 预测集数据标签

4.2 参数调优

参数调优程序如图 6 所示，使用 GridSearchCV() 算法对参数进行调优。以预测最大激活用户数为例，读取训练集 data.csv，定义 train_X 与 train_Y 即训练集与训练目标，并使用 train_test_split 函数按 8:2 分配训练集与测试集，逐个对参数传入模型，最终得出最优得分，获取最优参数组合。

```

# 这里输入的是处理之后的数据
train = pd.read_csv('./data.csv', low_memory=False)
train_X = train.drop(['date', 'PRB_usage', 'max_ue'], axis=1)
train_Y = train.loc[:, 'max_ue']
X_train, X_test, y_train, y_test = train_test_split(train_X, train_Y, test_size=0.2, random_state=42)
cv_params = {'max_depth': [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]}
# cv_params = {'num_leaves': [25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39]}
# cv_params = [{'learning_rate': [0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008]}]
# cv_params = {'n_estimators': list(range(1100, 1310, 10))}
# cv_params = {'min_data_in_leaf': [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]}
# cv_params = {'subsample': [0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09]}
# cv_params = {'random_state': [50, 60, 70, 80, 90, 100, 110, 120, 130, 140]}
other_params = {
    'task': 'train',
    'objective': 'regression', # 目标函数
    'max_depth': 9, # 树的深度
    'num_leaves': 31, # 树的最大叶子节点数
    'learning_rate': 0.008, # 学习率
    'n_estimators': 1250, # 迭代次数
    'min_data_in_leaf': 19,
    'subsample': 0.1,
    'colsample_bytree': 1,
    'reg_alpha': 0,
    'reg_lambda': 0,
    'random_state': 50,
}
model = lgbm.LGBMRegressor(**other_params)
optimized_GBM = GridSearchCV(estimator=model, param_grid=cv_params, scoring='r2')
optimized_GBM.fit(X_train, y_train)
means = optimized_GBM.cv_results_['mean_test_score']
params = optimized_GBM.cv_results_['params']
print('每轮迭代运行结果:{}'.format(params, means))
print('参数的最佳取值: {}'.format(optimized_GBM.best_params_))
print('最佳模型得分: {}'.format(optimized_GBM.best_score_))
    
```

图 6 调优程序

调参模型得分结果为：

1、预测 PRB 利用率：

```

'task': 'train',
'objective': 'regression', # 目标函数
'max_depth': 15, # 树的深度
'num_leaves': 33, # 树的最大叶子节点数
'learning_rate': 0.008, # 学习率
'n_estimators': 1270, # 迭代次数
'min_data_in_leaf': 19,
    
```

```
'subsample':0.1,
'colsample_bytree':1,
'reg_alpha':0,
'reg_lambda':0,
'random_state':50,
结果准确率得分为 0.95808, 达到预期目标。
```

2、最大激活 UE 数:

```
'task':'train',
'objective':'regression',# 目标函数
'max_depth':9,#树的深度
'num_leaves':31,#树的最大叶子节点数
'learning_rate':0.008,#学习率
```

```
'n_estimators':1250,#迭代次数
'min_data_in_leaf':19,
'subsample':0.1,
'colsample_bytree':1,
'reg_alpha':0,
'reg_lambda':0,
'random_state':50,
结果准确率得分为 0.91286, 达到预期目标。
```

4.3 性能验证

验证的方法为将 2019 年 8 月 4 日至 9 日的指标做训练, 共 699 行数据。验证结果如表 3 所示, 可见结果虽有出入, 但整体较合理。

prb 利用率绝对误差	样本数	占比	最大激活数绝对误差	样本数	占比
>=10	15	2%	>=10	6	1%
>=9	24	3%	>=9	12	2%
>=8	33	5%	>=8	14	2%
>=7	48	7%	>=7	25	4%
>=6	64	9%	>=6	37	5%
>=5	98	14%	>=5	58	8%
<5	601	86%	<5	641	92%

表 4 验证结果

4.4 预测分析

预测的 KQI 指标一律按照满分标准如表 5

web_response_time	web_sussess_time	UE_RTT	three_handshake	Initial_buffer_time	video_speed
210	0.97	130	35	2000	2000

表 5 KQI 预测集

预测时段包括一整天 24 小时, 将训练集传入可得最终结果如图 7 所示。

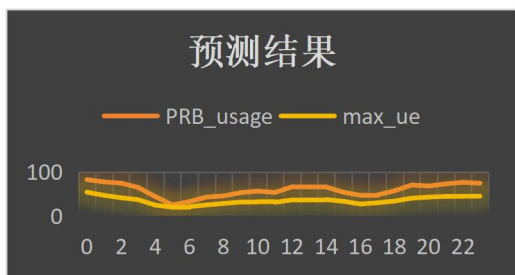


图 7 动态门限预测结果

5.时间窗口预测

Prophet 遵循 sklearn 库建模的应用程序接口。我们创建了一个 Prophet 类的实例, 其中使用了“拟合模型” fit 和“预测” predict 方法。Prophet 的输入量往往是一个包含两列的数据框: ds 和 y。ds 列必须包含日期 (YYYY-MM-DD) 或者是具体的时间点 (YYYY-MM-DDHH:MM:SS)。y 列必须是数值变量, 表示我们去预测的量。

下面实例中使用的是 eNodeBid=531885 每日负荷情况的时间序列数据 (2019/03/25-2020/05/19)。这个数据集具有多季

节周期性、不断变化的增长率和可以拟合特定日期（例如节假日）的情况等 Prophet 适用的性质。

5.1 数据清洗

数据清洗规则如下所示：

1、空值和 0 值用平均值填充

2、把数据取对数后进一步处理。之所以这样做是基于对数函数在其定义域内是单调增函数，取对数后不会改变数据的相对关系，取对数作用主要有：

1) 缩小数据的绝对数值，方便计算。例如，每个数据项的值都很大，许多这样的值进行计算可能对超过常用数据类型的取值范围，这时取对数，就把数值缩小了。

2) 取对数后，可以将乘法计算转换称加法计算。

3) 取对数之后不会改变数据的性质和相关关系，但压缩了变量的尺度，例如 $800/200=4$ ，但 $\log 800/\log 200=1.2616$ ，数据更加平稳，也削弱了模型的共线性、异方差性等。程序如图 8 所示。

```
# Python
import pandas as pd
from fbprophet import Prophet

#显示所有行
pd.set_option('display.max_columns', None)
#显示所有列
pd.set_option('display.max_rows', None)

# 读入数据集
df = pd.read_csv('hour531885.csv')
#df.tail()

# 预测未来3天72小时
predays=72
df1=df.loc[:, ['时间', '负荷']][:-predays]
df1=df1.rename(columns={'时间': 'ds', '负荷': 'y'})
#数据清洗
df1.fillna(df1.mean())
df1['y']=df1['y'].replace(0, df1['y'].mean())
df1['y'] = np.log(df1['y'])
#df1.tail()
#df1['cap'] =5000
```

图 8 数据清洗程序

5.2 建立模型

建立模型采取如下措施进行：

1、需要专门对节假日或者其它的事件进行建模，因此创建一个新的 dataframe，其中包含两列（节

假日 holiday 和日期戳 ds），每行分别记录了每个出现的节假日。这个数据框必须包含所有出现的节假日，不仅是历史数据集中还是待预测的时期中的。如果这些节假日并没有在待预测的时期中被注明，Prophet 也会利用历史数据对它们建模，但预测未来时却不会使用这些模型来预测。

注：也就是说，在待预测的日期里，我们也必须指定所有出现的节假日。

在这个数据基础上再新建两列 lower_window 和 upper_window，从而将节假日的时间扩展成一个区间 [lower_window, upper_window]。

2、季节性是用部分傅立叶和估计的。部分和 (order)中的项数是一个参数，它决定了季节性的变化有多快。默认值 10 通常是合适的，但是当季节性需要适应更高频率的变化时，它们可以增加，曲线将更加的多变。在实例化模型时，可以为每个内置季节性指定傅立叶级数。增加傅立叶项的数量可以使季节性适应更快的变化周期，但也可能导致过度拟合。

3、通过对一个 Prophet 对象进行实例化来拟合模型，任何影响预测过程的设置都将在构造模型时被指定。接下来，就可以使用 fit 方法代入历史数据集来拟合模型。

4、预测过程则需要建立在包含日期 ds 列的数据框基础上。通过使用辅助的方法 Prophet.make_future_dataframe 来将未来的日期扩展指定的天数，得到一个合规的数据框。默认情况下，这样做会自动包含历史数据的日期，因此我们也可以用来查看模型对于历史数据的拟合效果。

5、predict 方法将会对每一行未来 future 日期得到一个预测值（称为 yhat）。如果你传入了历史数据的日期，它将会提供样本的模型拟合值。预测 forecast 创建的对象应当是一个新的数据框，其中包含一系列预测值 yhat，以及成分的分析 and 置信区间。

程序如图 9 所示。

```
holidays = pd.DataFrame({
    'holiday': 'holiday',
    'ds': pd.to_datetime(['2018-02-15', '2018-02-16', '2018-02-17', '2018-02-18', '2018-02-19', '2018-02-20',
        '2018-02-21', '2018-01-02', '2018-04-05', '2018-04-06', '2018-04-07', '2018-04-28',
        '2018-05-01', '2018-06-16', '2018-06-17', '2018-06-18', '2018-08-17', '2018-09-22',
        '2018-09-23', '2018-09-24', '2018-10-01', '2018-10-02', '2018-10-03', '2018-10-04',
        '2018-10-05', '2018-10-06', '2018-10-07', '2018-12-24', '2018-12-25', '2018-12-30',
        '2018-12-31', '2019-01-01', '2018-12-30', '2019-02-04', '2019-02-05', '2019-02-06',
        '2019-02-07', '2019-02-08', '2019-02-09', '2019-02-10', '2019-03-05', '2019-03-06',
        '2019-03-07', '2019-05-01', '2019-05-02', '2019-09-13', '2019-09-14', '2019-09-15',
        '2019-10-04', '2019-10-05', '2019-10-06', '2019-10-01', '2019-10-02', '2019-10-03',
        '2019-10-07', '2019-12-24', '2019-12-25', '2020-01-01', '2020-01-24', '2020-01-25',
        '2020-01-26', '2020-01-27', '2020-01-28', '2020-01-29', '2020-01-30', '2020-01-31',
        '2020-02-01', '2020-03-03', '2020-03-04', '2020-02-01', '2020-04-04', '2020-04-05',
        '2020-04-06', '2020-05-01', '2020-05-02', '2020-05-03', '2020-05-04', '2020-05-05',
        '2020-06-23', '2020-06-26', '2020-06-27', '2020-08-23', '2020-08-24', '2020-10-01',
        '2020-10-03', '2020-10-04', '2020-10-05', '2020-10-06', '2020-10-07', '2020-10-08'])),
    'lower_window': 1,
    'upper_window': 0,
})
# 拟合模型
m = Prophet(yearly_seasonality=10, holidays=holidays, holidays_prior_scale=10)
# 拟合模型并生成预测结果
m.fit(df)
# 生成未来日期数据
future = m.make_future_dataframe(periods=predays, freq='H')
# 生成预测结果
forecast = m.predict(future)
forecasts[['ds', 'yhat', 'yhat_lower', 'yhat_upper']].tail()
```

图 9 模型建立程序

通过 Prophet.plot 方法传入预测得到的数据框，可以对预测的效果进行绘图，如图 10 所示，黑色表示原始的时间序列离散点，深蓝色的线表示使用时间序列来拟合所得到的取值，而浅蓝色的线表示时间序列的一个置信区间，也就是所谓的合理的上界和下界。

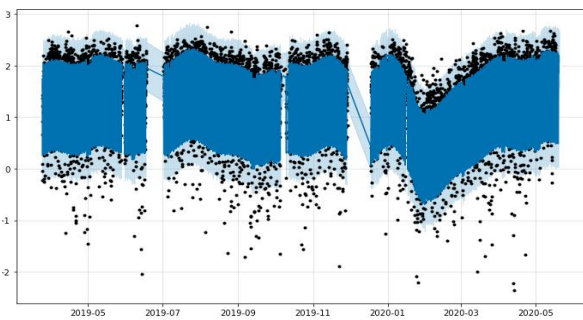


图 10 预测结果效果

5.3 性能验证

Prophet 包含时间序列交叉验证功能，以测量使用历史数据的预测误差。这是通过在历史记录中选择截止点来完成的，对于每一个都只使用该截止点之前的数据来拟合模型。然后，我们将预测值与实际值进行比较。

这个交叉验证过程可以使用 cross_validation 函数自动完成一系列历史截断。我们指定预测水平 (horizon)，然后选择初始训练期 (initial) 的大小和截断之间的间隔 (period)。默认情况下，初始训练期设置为 horizon 的三倍，每半个 horizon 就有一个截断。

cross_validation 的输出是一个 dataframe，在每个模拟预测日期和每个截断日期都有真实值 y 和样本预测值 yhat。特别地，对在 cutoff 和 cutoff +

horizon 之间的每一个观测点都进行了预测。然后，这个 dataframe 可以用来度量 yhat 和 y 的错误。在这里，我们做交叉验证来评估预测 horizon 在 72 小时的性能，从第一次截止时 8880 小时的训练数据开始，然后每 168 小时进行一次预测。如图 11 所示

```
from fbprophet.diagnostics import cross_validation
df_cv = cross_validation(m, initial='8880 hours', period='168 hours', horizon = '72 hours')
df_cv.head()
INFO:fbprophet:Making 7 forecasts with cutoffs between 2020-04-01 23:00:00 and 2020-05-13 23:00:00
```

	ds	yhat	yhat_lower	yhat_upper	y	cutoff
0	2020-04-02 00:00:00	1.941952	1.433295	2.399231	2.124477	2020-04-01 23:00:00
1	2020-04-02 01:00:00	1.689121	1.216427	2.182749	1.544814	2020-04-01 23:00:00
2	2020-04-02 02:00:00	1.286860	0.811299	1.789956	1.697821	2020-04-01 23:00:00
3	2020-04-02 03:00:00	0.828730	0.356237	1.307759	0.690940	2020-04-01 23:00:00
4	2020-04-02 04:00:00	0.494047	-0.004241	0.965676	0.443928	2020-04-01 23:00:00

图 11 cross_validation 输出

performance_metrics 作为离截止点(预测的未来距离)的函数，可用于计算关于预测性能的一些有用统计数据(如与 y 相比时 yhat、yhat_lower 和 yhat_upper)。计算得到的统计信息包括均方误差 (mean squared error, MSE)、均方根误差 (root mean squared error, RMSE)、平均绝对误差 (mean absolute error, MAE)、平均绝对误差 (mean absolute percent error, MAPE) 以及 yhat_lower 和 yhat_upper 估计的覆盖率。这些都是在 df_cv 中通过 horizon (ds - cutoff) 排序后预测的滚动窗口中计算出来的。默认情况下，每个窗口都会包含 10% 的预测，但是可以通过 rolling_window 参数来更改。如图 12 所示。

```
from fbprophet.diagnostics import performance_metrics
df_p = performance_metrics(df_cv)
df_p.head()
INFO:fbprophet:Making 7 forecasts with cutoffs between 2020-04-01 23:00:00 and 2020-05-13 23:00:00
```

	horizon	mse	rmse	mae	mape	mdape	coverage
0	08:00:00	0.242015	0.491950	0.397882	1.283945	0.350681	0.680000
1	09:00:00	0.238245	0.488104	0.394995	1.274424	0.340138	0.677143
2	10:00:00	0.189431	0.435236	0.345172	1.111020	0.240267	0.731429
3	11:00:00	0.150698	0.388198	0.308978	0.943548	0.203220	0.774286
4	12:00:00	0.120497	0.347126	0.275789	0.726779	0.160924	0.831429

图 12 performance_metrics 输出

交叉验证性能指标可以用 plot_cross_validation_metric 可视化，这里显示的是 MSE。点表示 df_cv 中每个预测的均方误差。蓝线表示 MSE。我们可以看到，对于未来三天的预测，误差在 0.25 以内。如图 13 所示。

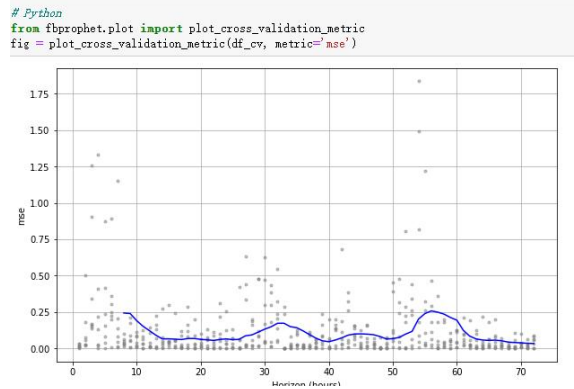


图 13 交叉功能验证性能指标

5.4 结果验证

把未来三天的真实值和预测值做个直观的对

比，可以看出，本模型已经可以很好的预测出未来 3 天每个小时的负荷趋势，因此即可得以小时为粒度的时间窗口。如图 14 所示。

```
import numpy as np
import matplotlib.pyplot as plt
df2=df.loc[:, ['时间', '负荷']] [-preadays:]
df2=df2.rename(columns={'时间': 'ds', '负荷': 'y'})
#数据清理
df2.fillna(df2.mean())
df2['y']=df2['y'].replace(0, df2['y'].mean())
df2['y'] = np.log(df2['y'])
true_data=np.array(df2['y'])
#print(true_data)
fb_pre = np.array(forecast['yhat'].iloc[-preadays:])
#print(fb_pre)
plt.plot(fb_pre, 'r', label='prediction')
plt.plot(true_data, 'b', label='real')
plt.legend(loc='best')
plt.show()
#计算MSE
fb_pre = np.array(forecast['yhat'].iloc[-preadays:])#
MSE = true_data - fb_pre
MSE = MSE*MSE
MSE_loss = sum(MSE)/len(MSE)
print("MSE_loss:",MSE_loss)
```

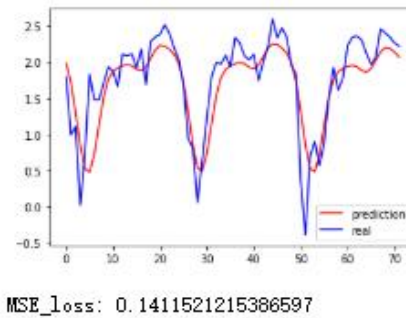


图 14 结果验证

6.结束语

将成果运用于现网中后，主要关注能耗下降与指标变化情况。

2021 年 3 月 15 日开启节能措施做对比验证，

前后一个礼拜的能耗对比如表 6 所示，能耗下降率 = (AI 开启前能耗-AI 开启后能耗) /AI 开启前能耗，可明显看到，能耗下降率在 7%左右，确实起到节能的效果。

未开启节能		开启节能		
时间	功率平均 (瓦)	时间	功率平均 (瓦)	能耗下降率
2021/3/8	152.27	2021/3/16	141.72	6.93%
2021/3/9	152.61	2021/3/17	143.71	5.83%
2021/3/10	154.08	2021/3/18	140.93	8.53%
2021/3/11	154.13	2021/3/19	137.78	10.61%
2021/3/12	154.03	2021/3/20	137.22	10.91%
2021/3/13	152.24	2021/3/21	138.48	9.04%
2021/3/14	149.7	2021/3/22	139.96	6.51%

表 6 能耗对比

在 KPI 指标方面，保持较稳定的趋势，如图 15 所示，在节能措施开启前后，各项指标均保持较为

稳定的趋势，无太大的波动。

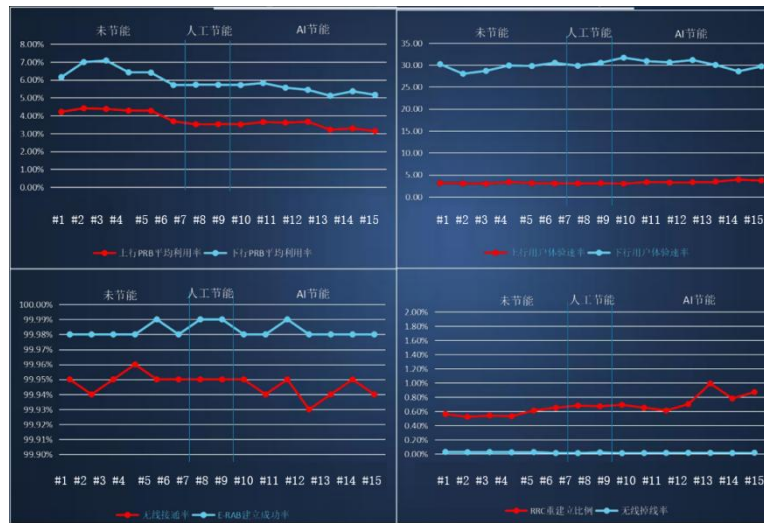


图 15 KPI 趋势

本节能方案立足于保障用户网络使用感知，创新性运用了感知指标作为训练集，预测动态门限与时间窗口。另外，训练集可包含全网小区，意味着可生成一套适用于每个小区的节能控制参数，达到精准节能的目标。对于实际运用或者应用研究有非常良好的示范性作用。

参考文献:

[1] 微笑 sun.[EB/OL]. LightGBM 介绍及参数

调优.2018-07-19-<https://www.cnblogs.com/jiangxinyang/p/9337094.html>

[2] 爱吃串串的瘦子.[EB/OL]. gridSearchCV (网格搜索)的参数、方法及示例.https://blog.csdn.net/weixin_41988628/article/details/83098130

[3] anshuai_aw1.[EB/OL]. 时间序列模型 Prophet 使用详细讲解.https://blog.csdn.net/anshuai_aw1/article/details/83412058

基站工参平台化管理的前瞻性应用方案

黄炳塔 董帝焱 吴晶晶 杜丕加

(中国联合网络通信有限公司泉州市分公司)

摘要: 射频天馈调整是网络优化的基础, 基站工参就是射频天馈调整的依据。随着基站规模日渐扩大, 不光是日常网络优化调整, 4/5G 的共建共享也使得基站工参数据随时变化, 现场优化人员无法随时查阅工参, 网络优化效率受到限制。为了解决基站工参数据能随时查阅, 将其进行平台化管理颇为重要。具体方法为将基站工参数据存储于数据库, 调用微信公众号接口进行访问, 根据查询需求, 随时推送相关工参数据。

关键词: 射频天馈, 平台化, 微信公众号

1. 背景

“开放创新”是中国联通的企业核心价值观之一, 体现到每一个战略的全面深化实施当中。中国联通鼓励创新, 积极推动网络优化向自动化、智能化转型。

射频天馈调整是网络优化的基础, 基站工参就是射频天馈调整的依据。基站工参管理的主要方法目前仍然是定期人工维护, 根据需求分发至工作人员, 保存于本地使用。网络优化现场人员在工作中获取工参的传统方式主要采用查看手提电脑或者电话联系后台人员帮忙查阅, 不仅费时费力, 且不直观。为了解决这一痛点问题, 将基站工参数据存储于数据库, 使用微信公众号平台建立起工作人员移动端与基站工参数据的通道。

2. 环境部署

2.1 itchatmp

itchatmp 是一个开源的微信公众号、企业号接口, 与 itchat 出自同一个团队。它能够快速实现微信公众号交互的自动化, 基于 tornado 框架, 能够轻松满足效率需求。

在 python 环境下可通过 `pip install itchatmp` 命令安装该库。在开发文档中提供了简单的应用案例, 案例中 `itchatmp.update_config(itchatmp.WechatConfig()` 为回

调参数, 指向供消息转发的服务器, 案例功能为将收到的消息重新发回给用户, 如图 1 所示。

```
import itchatmp

itchatmp.update_config(itchatmp.WechatConfig(
    token='yourToken',
    appId = 'yourAppId',
    appSecret = 'yourAppSecret'))

@itchatmp.msg_register(itchatmp.content.TEXT)
def text_reply(msg):
    return msg['Content']

itchatmp.run()
```

图 1 简单应用案例

2.2 微信公众号配置

微信公众号分为服务号、订阅号与企业号, 其中订阅号具体又分为个人订阅号以及组织订阅号, 差别在于权限的不同。本文使用的公众号为个人订阅号, 只需身份证即可免费申请, 与组织订阅号最大的差别在于无权限获取到用户的信息, 考虑到信息安全, 需人工维护注册用户清单。

如图 2 所示, 在微信公众号后台, 路径为开发-基本配置, 可设置及获取一些信息。其中开发者 ID 为微信后台自动配置, 开发者密码与 Token 为

管理员设置，这三项内容需填写到 itchatmp 的回调参数里。IP 白名单需要加入转发服务器的公网 IP，服务器地址即转发服务器的公网 IP。由于是个人订

阅号，消息加解密方式可以选择明文模式，消息加解密钥选择随机生成即可。



图 2 微信后台配置

2.3 腾讯云服务器配置

转发服务器的 ip 必须为外网能直接访问的公网 ip，由于公司设备无法提供公网 ip，故需部署在

云端。本次使用的转发服务器为试用 15 天的产品—腾讯云服务器（Cloud Virtual Machine，CVM），该产品可提供一个可远程登录的虚拟服务器以及一个公网 ip。服务器配置如图 3 所示。



图 3 腾讯云服务器配置

2.4 python 环境

python 的三方库较多，版本不一致可能导致的问题也多，在分享或开源代码时，需要把环境版本同时分享，避免不必要的问题。本次主要使用的库及关联库如表 1 所示。

tornado	6.0.4
urllib3	1.25.8
python	3.7.6

表 1 库名及版本号

库名	版本号
pandas	1.0.1
numpy	1.18.3
itchatmp	0.0.17
pypiwin	32-223
requests	2.22.0
requests_wsgi_adapter	0.4.1
six	1.14

需要注意的是，在使用 itchatmp 0.0.17 时，有发现 itchatmp 库的 components 文件夹 register.py 用法为 from tornado.wsgi import WSGIAdapter，提示报错。经查询，原因为 itchatmp 库使用的 tornado 版本为 4.5.3 才有该库，新版本的无 WSGIAdapter。解决方法安装最新版本的 tornado 6.0.4 后，需要再安装 requests-wsgi-adapter 0.4.1。将原语句:from tornado.wsgi import WSGIAdapter 更改为 from wsgiadapter import WSGIAdapter，问题解决。

3. 运行机制

平台运行的四个必要元素为微信公众号平台、供转发消息的服务器、订阅用户以及程序逻辑，其中逻辑程序运行于服务器上。四个元素之间的关系如图 4 所示。

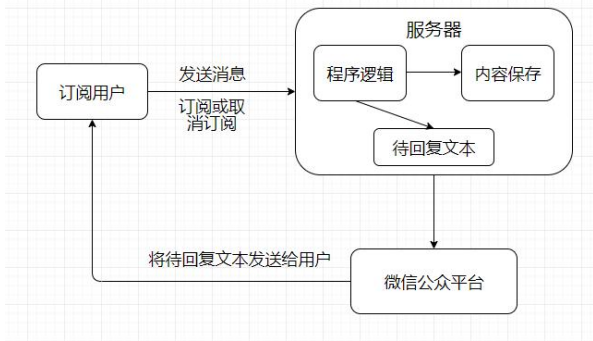


图 4 运行流程图

前文提到，个人订阅号无获取用户信息的权限，只能通过用户与公众号的互动中获取。识别用户的唯一标识即 openid，这也是微信公众号识别每一个订阅者的唯一标识，当微信用户对公众号有关注或取消关注的事件时，程序的处理逻辑如图 5 所示。平台将包含用户 openid 的信息保存为 txt 格式存入服务器。

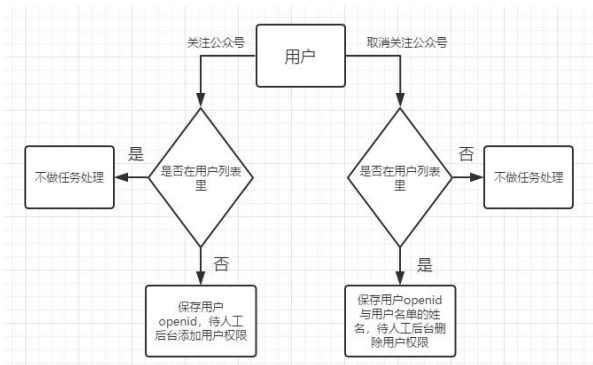


图 5 事件类消息处理流程

当接收用户发送的消息时，可自动判断消息类型，并按地理位置消息与文本消息两个大类进行处理，其中文本消息类型又分为关键字类型与 ID 类型，具体流程如图 6 所示。

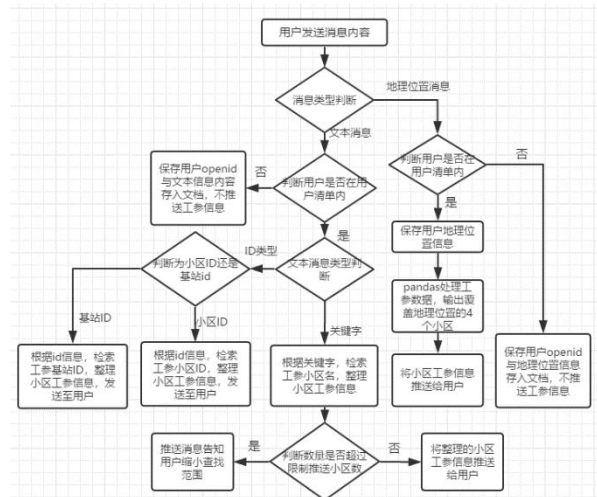


图 6 消息处理流程

4. 程序设计

4.1 工参数据库及用户列表

基站工参数据作为网优工作的一项基础数据，随着天馈调整、参数修改随时发生改变。本项目选取的字段包括站点号、县分、小区号、小区名、PCI、经度、纬度、覆盖类型、街道乡镇、频段、上行频点、下行频点、方向角、机械下倾角、电调下倾角、天线挂高、双工、传输模式、实际发射功率、RS 功率、RRU 型号、入网时间、天线型号、频域带宽 (MHz)，并将机械下倾角与电调下倾角两列合并成“机械下倾角+电调下倾角”格式。如图 7 所示。最终将其存储于 MYSQL 数据库。

```

path=pd.read_excel('LTE基站信息表.xlsx',dtype=str,usecols=['站点号','县分','小区号','小区名','PCI','经度','纬度','覆盖类型','街道乡镇','频段','上行频点','下行频点','方向角','机械下倾角','电调下倾角','天线挂高','双工','传输模式','实际发射功率','RS功率','RRU型号','入网时间','天线型号','频域带宽 (MHz)'])
path['下倾角']=path['机械下倾角']+'_'+path['电调下倾角']
path.to_csv('工参信息.csv',index=False)
    
```

图 7 工参数据

注册用户清单包含两个字段，一个为真实姓名，一个为 openid。如表 2 所示，其中 openid 为每个关注该公众号的用户唯一标识。

姓名	openid
黄**	oZwoLwtcuFXm-TdJMqfklqDsQTFE

表 2 注册用户清单格式

示例	说明
msg['FromUserName']	发送方帐号（一个 OpenID）
msg['Content']	文本消息内容
msg['Location_X']	地理位置纬度
msg['Location_Y']	地理位置经度
msg['Label']	地理位置信息
msg['CreateTime']	消息创建时间（整型）
msg['MsgType']	消息类型，文本为 text、地理位置为 location
event['Event']	事件类型，subscribe(订阅)、unsubscribe(取消订阅)

表 3 接口使用示例

4.3 全局变量与公用程序块定义

程序的全局变量与公用函数用于在需要时被随时调用。本程序设计中：

定义 path 为工参信息变量，格式为 dataframe；

定义 cellid_set 与 enbid_list_set 分别为小区 ID 与站点 ID 的集合，使用集合的原因为当数据量大时，访问速度更快；

定义 town_list 为县分的列表，由于县分数量不多，未采用集合；

定义 keycol_for_msgTEXT 为回复文本类消息的关键字段；

定义 keycol_for_msgLOCATION 为回复地理位置信息的关键字段；

定义 highestnum_for_inputkeywords 为限制回复的输入关键字及 id 的最高小区个数，设置为 14，原因为避免推送消息字数过多被微信公众号限制。

定义 highestnum_for_inputlocation 为限制回复的输入地址信息的最高小区个数。

具体信息如图 8 所示

4.2 微信公众号参数说明

微信公众号接收的消息为 xml 格式，包含着用户的基本信息、接收时间、消息内容等。itchatmp 基于微信公众号开发，高度遵守微信官方的开发文档，在一些接口也保持一致。本次主要介绍文本类消息与地理位置类消息的一些接口案例，以接收用户消息 msg 为例，如表 3 所示，通过这些使用方法，可以在用户与的互动中获取一些信息。

```
path=pd.read_csv('工参信息.csv',dtype=str,usecols=['站点号','小区名','小区号','PCI','下行频点','方向角','下倾角','经度','纬度','覆盖类型','县分'])
cellid_list=np.array(path['小区号']).tolist()#将所有小区id制成列表
cellid_set=set(cellid_list)
enbid_list=np.array(path['站点号']).tolist()#将所有站点id制成集合，访问速度较快
enbid_list_set=set(enbid_list)
town_list=np.array(path['县分']).tolist()#将所有县分名称制成列表，并去重
town_list=list(set(town_list))
keycol_for_msgTEXT=['小区名','小区号','PCI','下行频点','方向角','下倾角']#回复文本消息的字段
keycol_for_msgLOCATION=['小区名','小区号','PCI','下行频点','方向角','下倾角','距离(公里)']#回复地址信息的字段
highestnum_for_inputkeywords=14#限制回复的输入关键字及id的最高小区个数
highestnum_for_inputlocation=4#限制回复的输入地址信息的最高小区个数
```

图 8 全局变量

本程序的公用程序块有两个。第一个为 itchatmp.update_config(itchatmp.WechatConfig()) 函数，这段代码的作用在于填写 itchatmp 的回调参数设置，与微信公众号后台相关联；第二个为 getBasicInfo()函数，该函数的作用为获取对公众号收到的消息进行处理，获取用户的 openid 及消息接收时间。具体程序如图 9 所示。

```
itchatmp.update_config(itchatmp.WechatConfig(
    这几段代码为itchatmp的回调参数设置，与微信公众号后台相关联
    token='huangbt1209',
    appId='wx110cd35e90db5c16',
    appSecret='730ab5a8e1acad05eeb53bf3d74f5882'))
def getBasicInfo(msg):#该函数的作用为对消息进行处理，获取用户的openid，及时间
    date=time.strftime("%Y-%m-%d %H:%M:%S",time.localtime(int(msg['CreateTime'])))#接收消息的时间
    user=msg['FromUserName']#关注用户的OpenId，同一个公众号所有关注者都是唯一的
    return date, user
```

图 9 回调函数与获取用户 openid 与接收消息时间

4.4 文本类消息响应

文本类消息指公众号接收到来自用户的包含数字、中文、英文的消息，本程序将此类消息具体分为 ID 类纯数字消息与关键字混合的消息。

4.4.1 文本类型判断与响应函数

当接收到文本消息时，需要对消息进行判断。for_remsg()函数运用 is.isdigit()判断消息是否是纯数字，再判断是否存在于基站 ID 或者小区 ID 内，以此确定是否为小区 ID 与基站 ID 类消息或者关键字消息。

如果是 ID 类消息，且存在于小区 ID 或者基站 ID 内，则使用 remsg_for_inputID 函数，分别赋予不同的参数，输出推送内容。

如果是关键字类消息，则判断工参文件的小区名的列经筛选关键字后，是否为空。如果不为空，则先判断筛选后的小区是否超过限制，如果为否，则使用 remsg_for_inputkeywords 函数，输出推送内容。

如果消息经判断为无效消息，则推送给用户“无效关键字，请重新输入”的提示。类型判断的具体逻辑如图 10 所示。

```
def for_remsg(msg):#该函数的功能为对文本类消息进行分类判断，包括是否属于基站ID或者小区ID，或者是关键字
if msg.isdigit():
if msg in cellid_set:
remsg=remsg_for_inputID("小区号",msg)
elif msg in enbid_list_set:
remsg=remsg_for_inputID("站号",msg)
else:
remsg="无效ID,请重新输入"
elif not (path['小区号'].loc[path['小区名'].str.contains(msg)].empty):
if len(path['小区号'].loc[path['小区名'].str.contains(msg)])<=
highestnum_for_inputkeywords:
remsg=remsg_for_inputkeywords(msg)
else:
remsg="输出内容篇幅大于公众号字数限制,请缩小查找范围"
else:
remsg="无效关键字,请重新输入"
return remsg
```

图 10 类型判断函数

当接收到文本消息时，itchatmp 有对应的接口对消息进行接收。

@itchatmp.msg_register(itchatmp.content.TEXT)即为 itchatmp 对文本类消息响应的函数，其中 itchatmp.content.TEXT 的意思为在 itchatmp 安装包下的 content.py 内的 TEXT 类。

首先判断发送消息的用户是否存在于注册用户清单。如果存在，则使用 for_remsg()函数处理输出推送内容；如果不存在则将用户发送的消息与 openid 保存为 txt 格式的文件，以待后续将用户添加入注册用户清单，并推送提示“考虑到信息安全且个人订阅号权限限制，请先发送名字，待添加，谢谢”至用户。具体逻辑如图 11 所示。

```
@itchatmp.msg_register(itchatmp.content.TEXT)
#该代码为微信公众号文本消息类响应函数，通过调用之前的函数，以及做一些基本判断
def text_reply(msg):
users=pd.read_excel('users.xlsx',dtype=str)#
#用户列表读取函数内的原因是避免添加用户是重复停止运行程序
users_list=np.array(users['openid']).tolist()
content=msg['Content']
date,user=getBasicInfo(msg)
save_path='./微信机器人/文本信息/未注册用户记录/'#使用OpenID未保存不同用户的消息记录
if msg['FromUserName']in users_list:
remsg_for_remsg(content)
return remsg
else:
if not os.path.exists(save_path):#保存未注册用户回复的信息，方便登记用户
os.makedirs(save_path)
with open(save_path+user+'.txt','w',encoding='utf-8')as file:
file.write(date+"---"+content+user+"\n")
file.close()
else:
with open(save_path+user+'.txt','a',encoding='utf-8')as file:
file.write(date+"---"+content+user+"\n")
file.close()
return "考虑到信息安全且个人订阅号权限限制,请先发送名字,待添加,谢谢"
```

图 11 文本类响应函数

4.4.2 ID 类消息回复

当消息被判断为有效的 ID 类消息，便调用 remsg_for_inputID()函数。该函数的作用在于输出需要推送给用户的信息。具体逻辑为：函数被赋予两个参数：typeid 与 content。typeid 有两种，分别为小区号与基站号，content 即消息内容。

以 typeid 为 123456 的小区号为例，函数的具体实现方法为运用 pandas，在工参信息 dataframe、索引列名为“小区号”里筛选包含小区号 123456 的行，选取前文提到的 keycol_for_msgTEXT 的列，组成一个新的 dataframe—remsg_df，并按行对每个单元格读取，重组为文本信息。具体逻辑如图 12 所示。

```
def remsg_for_inputID(typeid,content):#该函数的功能为输出回复ID类信息的内容，包括ci号enbid
remsg=""
remsglist=[]
remsg_df=path[keycol_for_msgTEXT][path[typeid]==content]
remsg_df.reset_index(inplace=True,drop=True)
cell_num=len(remsg_df['小区名'])
for i in range(cell_num):
remsg1=str(i+1)+","
remsg2=""
for j in keycol_for_msgTEXT:
remsg2+=j+":"+remsg_df[j][i]+";"
remsgtemp=remsg1+remsg2
remsglist.append(remsgtemp)
remsg='\n'.join(remsg for remsg in remsglist)
return remsg
```

图 12 ID 类消息回复输出

4.4.3 关键字类消息回复

与 ID 类消息回复类似，当消息被判断为有效的关键字类消息，便调用 remsg_for_inputkeywords()函数。函数仅有一个参数：content 即消息内容。

函数的具体实现方法为运用 pandas，在工参信息的 dataframe、索引列名为“小区名”里筛选包含关键字的行，选取前文提到的 keycol_for_msgTEXT 的列，组成一个新的 dataframe—remsg_df，并按行对每个单元格读取，重组为文本信息。具体逻辑如图 13 所示。

```
def remsg_for_inputkeywords(content):#该函数的功能为输出回复关键字类信息的内容
    remsg=""
    remsglist=[]
    remsg_df=path.loc[path['小区名'].str.contains(content)]
    remsg_df=remsg_df.head(highestnum_for_inputkeywords)
    remsg_df.reset_index(inplace=True,drop=True)
    cell_num=len(remsg_df['小区名'])
    for i in range(cell_num):
        remsg1=str(i+1)+'\n'
        remsg2=""
        for j in keycol_for_msgTEXT:
            remsg2 += j+' '+remsg_df[j][i]+';'
        remsgtemp=remsg1+remsg2
        remsglist.append(remsgtemp)
    remsg='\n'.join(remsgs for remsgs in remsglist)
    return remsg
```

图 13 关键字类消息回复输出

4.5 位置类消息响应

4.5.1 函数定义

如图 14 所示，该函数的作用为根据两个经纬度，计算两个地点的距离。其中点 (lon1, lat1) 为本位，(lon2, lat2) 为他位。计算结果单位为公里。

```
def haversine(lon1, lat1, lon2, lat2): # 经度1, 纬度1, 经度2, 纬度2 (十进制度数)
    lon1=float(lon1)
    lat1=float(lat1)
    lon2=float(lon2)
    lat2=float(lat2)
    # 将十进制度数转化为弧度
    lon1, lat1, lon2, lat2 = map(radians, [lon1, lat1, lon2, lat2])
    # haversine公式
    dlon = lon2 - lon1
    dlat = lat2 - lat1
    a = sin(dlat/2)**2 + cos(lat1) * cos(lat2) * sin(dlon/2)**2
    c = 2 * asin(sqrt(a))
    r = 6371 # 地球平均半径, 单位为公里
    distance=round(c * r,2)
    return distance#单位为公里
```

图 14 根据两个经纬度计算距离

如图 15 所示，该函数的作用为根据两个经纬度，计算一个点对于另一个点的方位角。其中点 (lon1, lat1) 为本位，(lon2, lat2) 为他位。

```
#根据经纬度计算方位角
def getDegree(lon1, lat1, lon2, lat2):
    lon1=float(lon1)
    lat1=float(lat1)
    lon2=float(lon2)
    lat2=float(lat2)
    radlon1 = radians(lon1)
    radlat1 = radians(lat1)
    radlon2 = radians(lon2)
    radlat2 = radians(lat2)
    dlon = radlon2 - radlon1
    y = sin(dlon) * cos(radlat2)
    x = cos(radlat1) * sin(radlat2) - sin(radlat1) * cos(radlat2) * cos(dlon)
    brng = degrees(atan2(y, x))
    brng = round((brng + 360) % 360 ,1)
    return brng
```

图 15 根据两个经纬度计算方位角

如图 16 所示，定义 remsg_for_inputlocation () 函数，该函数的作用为接收一个类型为 dataframe 的 remsg_df 参数，类似于前文的 remsg_for_inputID 与 remsg_for_inputkeywords 函数，运用 pandas，选取 remsg_df 列索引为前文提到的

keycol_for_msgLOCATION 的列，并按行对每个单元格读取，重组成本文信息。

```
def remsg_for_inputlocation(remsg_df):#对remsg_df的内容做输出，即为要回复的内容
    remsg=""
    remsglist=[]
    for i in range(highestnum_for_inputlocation):
        remsg1=str(i+1)+'\n'
        remsg2=""
        for j in keycol_for_msgLOCATION:
            remsg2 += j+' '+remsg_df[j][i]+';'
        remsgtemp=remsg1+remsg2
        remsglist.append(remsgtemp)
    remsg='\n'.join(remsgs for remsgs in remsglist)
    return remsg
```

图 16 位置类回复消息输出

对于地理位置类消息，消息包含的内容有用户的经度、纬度以及位置标签。定义 output_target() 函数，传入用户经度、用户纬度、用户地理位置标签三个信息。当接收到这三个信息时，需要对工参信息进行整理，整理逻辑如下：

- 1、位置标签含有县分名称，对工参信息 dataframe 的“县分”列进行筛选，缩小范围。
- 2、创建“距离(公里)”列，调用 haversine () 函数，将工参上的经纬度与消息内包含的经纬度传入函数，依次计算距离填入该列。
- 3、创建“absbrng”列，即基站小区覆盖方位角减去用户位置相对于小区的方位角的绝对值。调用 getDegree () 函数，计算用户位置相对于基站小区的方位角，再将基站小区覆盖方位角减去计算用户位置相对于基站小区的方位角，依次取绝对值填入该列。
- 4、当覆盖类型为室分时，方位角绝对值设置为 0，原因为室分的方位角均设置为 360 度，避免计算不准确。
- 5、对“距离(公里)”列进行升序排序，且取小于 5 公里的小区
- 6、对“absbrng”列取小于 50 度的小区，即可得到覆盖用户位置的小区。
- 7、调用 remsg_for_inputlocation () 函数，输出位置类回复消息。

具体逻辑如图 17 所示。

```
def output_target(lonu, latu, label):#输出remsg_df, 并调用remsg_for_inputlocation函数, 作最终输出
for town in town_list:
    if town in label:
        town_used=town
        temp_path[path["经度"]]=town_used
        temp_path["距离(公里)"]=""
        temp_path["absbrng"]=""
        temp_path.reset_index(inplace=True,drop=True)
        times0=len(temp_path["小区名"])
        remsg_df=pd.DataFrame()
        for i in range(0,times0):
            lonc=temp_path["经度"][i]
            latc=temp_path["纬度"][i]
            cover_direction=lonc(temp_path["方向角"][i])
            covertype=temp_path["覆盖类型"][i]
            brng_getDegree(lonc, latc, lonu, latu)
            distance=havensine(lonc, latc, lonu, latu)
            if covertype=="室内":
                temp_path["absbrng"].iloc[i]=0
            else:
                temp_path["absbrng"].iloc[i]=abs(brng_cover_direction)
            temp_path["距离(公里)"].iloc[i]=distance
        temp_path=temp_path.sort_values(by="距离(公里)", ascending=True)
        temp_path=temp_path[temp_path["距离(公里)"]<50]
        temp_path.reset_index(inplace=True,drop=True)
        remsg_df=temp_path.astype("str")
        remsg=remsg_for_inputlocation(remsg_df)
    return remsg
```

图 17 位置类消息-目标小区筛选及调用 remsg_for_inputlocation 函数

4.5.2 消息响应

itchatmp 库使用

@itchatmp.msg_register(itchatmp.content.LOCATION)对地理位置类消息进行响应。如图 18 所示, 本段代码的作用为当收到用户地理位置消息时, 获取用户经纬度及位置标签, 再调用相应的函数输出回复消息, 同时, 对存在于注册用户清单里的用户, 保存地理位置信息; 对不存在于注册用户清单里的用户, 保存地理位置信息及 openid。

```
@itchatmp.msg_register(itchatmp.content.LOCATION)#代码段为保存公众号地理位置类消息, 并调用相应函数, 作最终输出, 并返回消息
def location_msg(msg):
    users=pd.read_excel('users.xlsx',dtype=str)
    users_list=np.array(users['openid']).tolist()
    username_list=np.array(users['姓名']).tolist()
    lon=msg['location']['lon']
    lat=msg['location']['lat']
    data_user=getBasicInfo(msg)
    save_path_Monitoruser='微信机器人/位置信息/未注册用户记录/' # 微信机器人/位置信息/未注册用户记录
    save_path_existuser='微信机器人/位置信息/注册用户记录/' # 微信机器人/位置信息/注册用户记录
    if msg['fromUserName'] in users_list:
        remsg=output_target(lon,lat,label)
        username=username_list[users_list.index(msg['fromUserName'])]
        os.makedirs(save_path_existuser)
        if not os.path.exists(save_path_existuser):
            os.makedirs(save_path_existuser)
            with open(save_path_existuser+username+'.txt','w',encoding='utf-8') as file:
                file.write(date+'---'+label+'---'+用度经度+'lat:'+lon+'lon:'+lat)
                file.close()
    else:
        with open(save_path_Monitoruser+username+'.txt','w',encoding='utf-8') as file:
            file.write(date+'---'+label+'---'+用度经度+'lat:'+lon+'lon:'+lat)
            file.close()
    return remsg
if not os.path.exists(save_path_Monitoruser):
    os.makedirs(save_path_Monitoruser)
with open(save_path_Monitoruser+user+'.txt','w',encoding='utf-8') as file:
    file.write(date+'---'+label+'---'+用度经度+'lat:'+lon+'lon:'+lat)
    file.close()
else:
    with open(save_path_Monitoruser+user+'.txt','w',encoding='utf-8') as file:
        file.write(date+'---'+label+'---'+用度经度+'lat:'+lon+'lon:'+lat)
        file.close()
return remsg
```

图 18 位置类消息响应逻辑

4.6 事件类消息响应

itchatmp 库使用@itchatmp.msg_register(itchatmp.content.EVENT)对事件类消息进行响应。代码如图 19 所示, 该段代码的作用为:

- 1、当用户点击关注公众号且不存在于注册用户单, 保存用户的 openid 与时间, 待人工加入注册用户清单。
- 2、当用户取消关注公众号且存在于注册用户清单, 保存用户的 openid、时间及姓名, 待人工取

消权限。

3、保存的格式均为 txt 格式, 方便查看。

```
@itchatmp.msg_register(itchatmp.content.EVENT)#代码段为保存关注者openid号
def event_reply(event):
    users=pd.read_excel('users.xlsx',dtype=str)
    users_list=np.array(users['openid']).tolist()
    username_list=np.array(users['姓名']).tolist()
    date, user = getBasicInfo(event)
    save_path_subscribe='微信机器人/事件信息/关注/'
    save_path_unsubscribe='微信机器人/事件信息/取消关注/'
    if event['Event']=='subscribe':
        if not event['fromUserName'] in users_list:
            if not os.path.exists(save_path_subscribe):
                os.makedirs(save_path_subscribe)
            with open(save_path_subscribe+user+'.txt','w',encoding='utf-8') as file:
                file.write(date+'---'+user+"\n")
                file.close()
            else:
                with open(save_path_subscribe+user+'.txt','a',encoding='utf-8') as file:
                    file.write(date+'---'+user+"\n")
                    file.close()
    elif event['Event']=='unsubscribe':
        username=username_list[users_list.index(user)]
        if event['fromUserName'] in users_list:
            if not os.path.exists(save_path_unsubscribe):
                os.makedirs(save_path_unsubscribe)
            with open(save_path_unsubscribe+user+'.txt','w',encoding='utf-8') as file:
                file.write(date+'---'+username+user+"\n")
                file.close()
            else:
                with open(save_path_unsubscribe+user+'.txt','a',encoding='utf-8') as file:
                    file.write(date+'---'+username+user+"\n")
                    file.close()
```

图 19 事件类消息响应逻辑

5.功能演示

5.1 文本类消息演示

当关键字查找范围过大, 提示缩小查找范围; 当消息有效, 推送小区基本工参信息; 当 ID 类消息无效, 提示无效 ID; 当消息有效, 推送小区基本工参信息; 当用户未注册, 提示未登记用户, 待添加权限。如图 20 所示。



图 20 文本类消息演示

5.2 位置类消息演示

当接收到地理位置信息,经处理后能推送地理位置周边小区工参信息如图 21 所示。



图 21 位置类消息演示

6. 结束语

基站工参数据平台化管理应用目前小规模应用于日常优化工作中,应用人员仅包括网优工作人员,取得了良好的成效。根据使用经验,还将能应用于领导、县分甚至客户的投诉中,利用定位具体问题小区,提升网优前台的工作效率。在未来的设想中,还需基于工参数据,开发移动端软件,形成类似于区块链的基站工参管理系统,并在此基础上

进行工参更新及纠错功能,这一切还需时间来沉淀,使之更加完善。

参考文献:

[1]简子道.itchatmp | 基于python 微信公众号接口[EB/OL].<https://www.jianshu.com/p/b031379b3a46>,2019-06-21.

[2]东风微鸣.如何编写 Python Web 框架(三)[EB/OL].<https://www.jianshu.com/p/53fcfd5cc3ac>,2019.03.06.

[3]微信官方.微信公众平台开发概述| 微信开放文档[EB/OL].https://developers.weixin.qq.com/doc/offiaccount/Getting_Started/Getting_Started_Guide.html,2020-3-13.

[4]Yunhe_Feng.Python 计算地图上两点经纬度间的距离[EB/OL].https://blog.csdn.net/vernice/article/details/46581361?depth_1-utm_source=distribute.pc_relevant.none-task&utm_source=distribute.pc_relevant.none-task,2015-06-21.

基于 Flink 的网络设备性能实时 监控系统的设计与实现

王振峰 陈志安 陈文 郑银云 杨慰民
(中国移动通信集团福建有限公司)

摘要: 为了解决网络设备性能质量实时监控问题, 缩短用户感知影响时长, 在传统单服务器无法及时处理日益增长的海量数据的前提下, 引入分布式流处理框架 Flink, 实现海量网络数据的实时处理。结合网络 KPI 劣化门限定义, 实现网络设备性能实时监控。本方案用于网络设备性能数据的实时处理与预警监控, 并在此基础上结合短信网关接口、邮箱接口的通知能力, 实现实时性能告警的通知, 进而实现网络设备性能的实时监控目的。

关键词: Flink; 实时; 性能监控

引言

随着网络数据的不断扩大, 网络数据呈现了大数据的 4V 特性: 容量大 (Volume)、速度快 (Velocity)、多样性 (Variety)、价值密度低 (Value)。以福建移动 DPI (Deep Packet Inspection) 数据为例, 每天产生的单据数据量约为 500TB 规模, 使用 Hadoop 集群的 MapReduce 框架的批处理能力执行大数据计算, 数据处理时延超过 2 小时, 能够满足面向业务分析的数据需求。但随着网络运维数智化转型、赋能高质量和高效网络运维的发展诉求, 如何使用大数据来实时支撑网络运维生产成为了当下的痛点。因此研究大数据实时处理成为当下电信运营商网络运维的重点工作。

1. 网络性能监控现状

网络设备性能监控主要使用数据包括网元北向接口性能数据、DPI^[1]数据 2 大类。北向接口性能

数据采集自网络设备的北向接口, 设备商根据 3GPP 协议规范、中国移动 OMC 系统北向接口数据规范生成 XML 文件保存在 SFTP 服务器。DPI 数据使用探针采集自网络设备接口, DPI 系统厂商根据中国移动 5G 上网日志留存系统技术规范生成 CSV 文件保存在 HDFS 文件系统。目前数据处理的主要方法如下:

方法一: 传统采集处理平台

传统的采集处理平台局限于数据处理能力, 适用于小数据量的 ETL (Extract Transform Load) 处理, 网元北向接口性能数据一般使用该方案, 处理时延 0.8 小时。

方法二: Hadoop 系统 MapReduce 处理框架

MapReduce 是 Hadoop 内置的分布式计算组件, 是一种分布式并行编程模型, 适用于大规模数据的并行运算, DPI 数据一般使用该方案, 处理时长 Hadoop 集群规模正相关, 平均数据处理时延 2 小

作者简介:

王振峰: 中国移动通信集团福建有限公司网络部网管中心员工

陈志安: 中国移动通信集团福建有限公司网络部网管中心员工

陈文: 中国移动通信集团福建有限公司网络部网管中心员工

郑银云: 中国移动通信集团福建有限公司网络部网管中心员工

杨慰民: 中国移动通信集团公司首席专家; 中国移动通信集团福建有限公司网络部网管中心副总经理

时。

方案三：自研实时处理程序

自研实时处理程序使用 JAVA 语言开发，能够实现数据的实时处理，适用于数据规模有限的单服务器运算，经过裁剪的 DPI 数据一般使用该方案，该方案以高性能服务器为基础，时延约 30 秒。

2.网络设备性能监控方案的缺点

方案一受限于服务器单机的处理能力，只适用于小数据量的数据处理，同时存在处理时延较长，难以响应监控专业对实时数据的要求。

方案二引入了分布式批处理框架，能够实现分布式计算节点的水平扩展，能够符合大数据的算力要求，但由于 MapReduce 框架采用的是批处理方式，计算过程存在多次落盘，导致计算历时长，难以响应监控专业对实时数据的要求。

方案三采用自研实时处理程序能够满足监控专业对实时数据的要求，但存在对开发人员的技能水平要求高，开发难度大，维护难度较大，计算能力依赖于高性能服务器，数据处理时延与数据量成正比，扩展能力差等问题。随着 5G 流量的不断增

长，在可预见的时间范围内，服务器的计算性能将成为计算瓶颈，该方案将存在数据背压的风险。

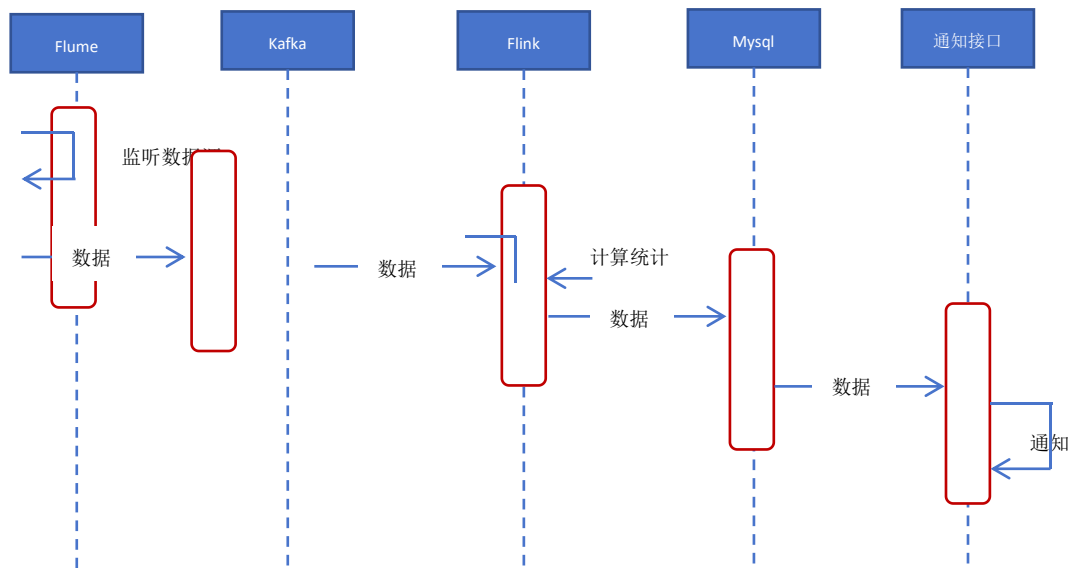
3.基于 Flink 的网络设备性能实时监控系统的设计与实现步骤

本文基于运营商网络采集的海量终端上网信令日志数据，提出了一种基于 Flink 的网络设备性能实时监控系统的设计与实现。该方法适用于数据实时处理场景，最大可支持到 PB 级数据规模，基于 Flink 的 DataStream 编程模型实现数据的流式处理^[2]，使用 Kafka 作为 Flink 的输入源（Source），使用 Mysql 作为 Flink 的输出源（Sink）。

本方法主要步骤如下：

- (1) 使用 Flume 采集 DPI 大数据^[3]；
- (2) 使用 Kafka 承载来自 Flume 采集的数据^[4]；
- (3) 基于 Flink 的 DataStream 编程模型编写业务程序，将结果数据持久化到 Mysql；
- (4) 根据预警规则生成监控预警，并通过短信网关/邮箱通知运维相关人员。

数据流时序图如下：



3.1、数据源

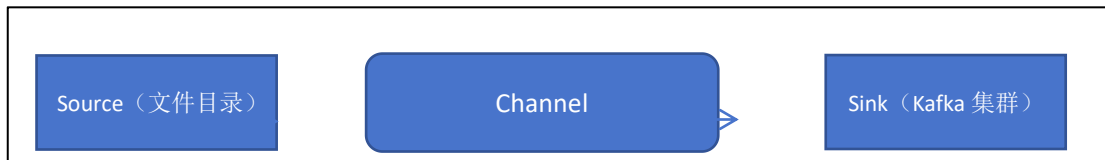
数据源包括 5G DPI 的原始单据，包括

N7/N8/N11/N12/N22/N40 合计 6 个接口数据，涉及 13 个核心性能指标。

接口	接口涉及网元	消息	指标名称
N8	AMF-UDM	Nudm_UECM_Registration	Nudm_UECM_Registration 注册成功次数
			Nudm_UECM_Registration 注册请求次数
			Nudm_UECM_Registration 注册成功率
N8	AMF-UDM	Nudm_SDM_Subscribe	Nudm_SDM_Subscribe 订阅成功次数
			Nudm_SDM_Subscribe 订阅请求次数
			Nudm_SDM_Subscribe 订阅成功率
N8	AMF-UDM	Nudm_SDM_Get	Nudm_SDM_Get 获取成功次数
			Nudm_SDM_Get 获取请求次数
			Nudm_SDM_Get 获取成功率
N11	AMF-SMF	Nsmf_PDUSession_CreateSMContext	PDU 会话建立成功次数
			PDU 会话建立请求次数
			PDU 会话建立成功率
N11	AMF-SMF	Nsmf_PDUSession_UpdateSMContext	PDU 会话更新成功次数
			PDU 会话更新请求次数
			PDU 会话更新成功率
N12	AMF-AUSF	Nausf_UEAuthentication_Authenticate	鉴权成功次数
			鉴权请求次数
			鉴权成功率
N7	SMF-PCF	Npcf_SMPolicyControl_Create	SM 策略关联建立成功次数
			SM 策略关联建立请求次数
			SM 策略关联建立成功率
N7	SMF-PCF	Npcf_SMPolicyControl_Update	SM 策略关联修改成功次数 (SMF 发起)
			SM 策略关联修改请求次数 (SMF 发起)
			SM 策略关联修改成功率 (SMF 发起)
N7	SMF-PCF	Npcf_SMPolicyControl_UpdateNotify	SM 策略关联修改成功次数 (PCF 发起)
			SM 策略关联修改请求次数 (PCF 发起)
			SM 策略关联修改成功率 (PCF 发起)
N22	AMF-NSSF	Nnssf_NSSelection_Get	切片信息获取成功次数
			切片信息获取请求次数
			切片信息获取成功率
N40	SMF-CHF	Nchf_ConvergedCharging_Create	计费建立成功次数
			计费建立请求次数
			计费建立成功率
N40	SMF-CHF	Nchf_ConvergedCharging_Update	计费更新成功次数
			计费更新请求次数
			计费更新成功率
N40	SMF-CHF	Nchf_ConvergedCharging_Notify	计费修改成功次数
			计费修改请求次数
			计费修改成功率

在数据源节点配置 Flume Agent 作为本方案的数据采集系统，实现 DPI 数据的实时采集。

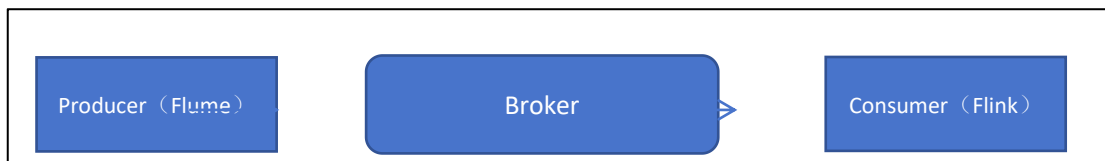
Flume Agent 有三个组件：Source、Channel 和



3.2、Kafka 集群构建

由于数据源节点数据的获取速率是非匀速的，数据量在时间维度存在潮汐效应，故引入 Kafka 集群作为数据缓存中间件，同时 Kafka 具备高吞吐、可持久化、可水平扩展、支持流数据处理等特性，能够适配本方案的实时性要求。

一个典型的 Kafka 集群包含 Producer、Broker、



3.3、Flink 业务程序构建

基于 Flink 的 DataStream 编程模型，构建流数据处理程序，基本运行流程包含如下几个步骤：

(1) 创建流处理执行环境

引入 flink Streaming 库，使用 API 获取流处理环境

```

val
env=StreamExecutionEnvironment.getExecutionEnvironment
    
```

(2) 创建 Kafka 数据源

在执行环境中添加 Kafka 数据源，配置 Kafka 数据源的 Topic。

(3) 无效数据过滤

通过 DataStream 编程模型的 Filter 函数过滤无效单据

(3) 指标字段映射

使用 DataStream 编程模型的 Map 函数，根据指标算法实现字段映射，以 N8 接口的注册成功次数和注册请求次数为例，编程算法如下：

```

val nudm_uecm_registration_SuccNbr =
if(value(13).length > 0 && value(13).toInt==205 &&
value(18).length>0 && value(18).toInt==1) 1 else 0;
    
```

Sink。Source 负责从数据源节点的目录中获取数据，Sink 负责将数据推送至 Kafka 集群。

Consumer，其中 Producer 负责将信息发送到 Broker，Consumer 负责从 Broker 订阅并消费消息。对应到本方案中，Producer 即为第一点介绍的 Flume Agent，Consumer 即为后续介绍的 Flink 程序。

Kafka 中的消息以 Topic 为单位，本方案创建的 Topic 名称为“1helloworld”，Flume 将源数据发送到该 Topic，Flink 从该 Topic 消费数据。

```

val nudm_uecm_registration_AttNbr =
if(value(13).length > 0 && value(13).toInt==205) 1
else 0;
    
```

(4) 窗口分配器

窗口分配器是负责将到来的每条单据分配给一个或多个窗口，本方案使用 DataStream 编程模型的滚动窗口，设置每 60 秒做 1 次窗口计算，编程算法如下：

```

window(TumblingEventTimeWindows.of(Time.seconds(60)))
    
```

(5) 窗口计算函数

确定了窗口分配器后，就要指定窗口计算函数，本方案采用的是 AggregateFunction 增量聚合函数，将第三步中 Map 输出结果作增量累加计算。

(6) 指定数据输出方式

采用 Mysql 作为 Flink 业务程序的输出 Sink，将结果数据持久化到数据库中。

```

dataStream.addSink(new JdbcOutputFormat<>...)
    
```

(7) 触发 Flink 框架执行程序

编写好业务模型后，需要触发 Flink 来执行程序，Flink 不断读取 Kafka 中的消息数据做计算后输出到 Mysql。

env.execute(“Flink with kafka reading”);

3.4、通知接口

通过编写 Python 脚本轮询 Mysql 数据,根据网络设备性能指标劣化门限触发短信/邮箱通知接口。

4.方案结果

根据本方案搭建生产环境,包括 Flume、Kafka 集群、Flink 集群、Mysql,对 6 个接口数据进行计算,统计各接口的 Flink 计算时延如下:

表 1、6 个接口计算结果表

接口	数据量(GB/分钟)	计算平均时长(秒)
N7	1.5	3
N8	1.2	2
N11	1.1	2
N12	1.1	2
N22	0.4	0.5
N40	2.5	5

从上面结果可以看出,使用 Flink 作为计算框架,数据的计算时长均在秒级,能够满足监控专业对数据实时性的要求。

5.结束语

本文提出的基于 Flink 的网络设备性能实时监控系统的设计与实现,使用的均为当前流程的开源框架,包括采集框架 Flume、消息组件 Kafka、流处理框架 Flink。以流处理作为实时程序为着手点,构建从采集、消息传输、消息处理、消息存储、业务触发的全流程实时应用解决方案,平均处理时延为 3 秒,能够满足对网络设备性能实时监控的需求。同时解耦了数据处理、告警规则、告警通知功能,能够灵活根据告警规则自适应调整告警通知发送。

本文提出的方案以及实践经验在行业内具有较强的推广借鉴价值。

参考文献:

- [3] 中国移动通信集团.中国移动统一 DPI 设备技术规范[S].北京
- [4] 吴小芳.基于 Flink 的实时数仓设计及 DPI 业务应用[J].计算机时代,2022(05):56-59+62.DOI:10.16644/j.cnki.cn33-1094/tp.2022.05.01
- [5] 宋其新.基于 Flume 的 A 接口信令采集系统设计及实现[D].北京邮电大学,2016.
- [6] 杨洁.利用 Kafka 技术实现用户实时账单的短信提醒功能[J].数字通信世界,2017(11):211-213.

基于 MEC 打造 5G 智慧园区

陈锋 洪林梦涵 王丽春

(中国联合网络通信有限公司福州市分公司)

摘要: 本文通过 5G 建设,进而利用 5G 新技术,推动生产制造、管理服务的深刻变革,使该产业在实现制造业向网络化、智能化、数字化转型过程中担当技术创新引领作用。初步打造出适合本企业的 5G+“智慧工厂”的新模式。通过升级改造智慧工厂,促进资源配置和使用效率提升,降低企业生产运营成本,增强企业竞争力。福清蓝色产业园在企业智能化、数字化转型方面将部署基于机器视觉的 AGV、AR 远程维护、安全行为监控、基于机器视觉的铝材表面检测等应用。

关键词: 5G 专网;融合切片;ULCL 分流

1. 概述

5G+“智慧工厂”是制造业的新形态,也是未来制造业的新引擎。智慧工厂被认为是 5G 技术的重要应用场景之一,是指通过大数据、物联网等技术加强信息管理和服务,进而制定并安排合理有效的生产计划和策略,它是智能制造的关键载体。本次项目是通过 5G 建设,进而利用 5G 新技术,推动生产制造、管理服务的深刻变革,使中铝瑞闽在实现制造业向网络化、智能化、数字化转型过程中担当技术创新引领作用。初步打造出适合本企业的 5G+“智慧工厂”的新模式。通过升级改造智慧工厂,促进资源配置和使用效率提升,降低企业生产

运营成本,增强企业竞争力。

2. 5G 专网

2.1 5G 专网网络架构

中国联通 5G 行业专网体系 2.0 通过“4+2”创新产品体系,为用户提供“端-网-边-平台应用”4 大基础能力以及“安全和服务”2 大服务底座,满足不同行业差异化和定制化需求。面向各行业提供专属解决方案和产品,高效支撑行业数字化转型升级。客户可根据自身需求定制不同的子产品/能力组合服务,包括终端定制、网络定制、边缘定制、应用平台、安全和服务 6 大能力定制,供客户菜单式选择服务。

作者简介:

陈锋: 高级工程师,现任职于中国联合网络通信有限公司福州市分公司云网中心,云网专家,从事无线网络规划建设优化工作

洪林梦涵: 现任职于中国联合网络通信有限公司福州市分公司云网中心,网络优化工程师,从事无线网络优化工作

王丽春: 现任职于中国联合网络通信有限公司福州市分公司云网中心,从事无线网络维护工作



图 1 5G 专网网络架构

4 大基础能力

网络定制化：基于中国联通 5G 2B 核心网集中一朵云，分布一张网的全程全网优势，提供“基础产品、切片产品、增强产品”三类网络产品，实现一点开户、全国可用，一地创新、全网复制，一点调度、全网开通的定制化服务。

云边协同化：基于专供和共享两种边缘云产品，提供边缘基础能力、增值能力和应用能力三大服务，“云网边”协同，为客户提供一站式“融合、开放、联动、弹性”服务。

终端智能化：基于“网络软件化、软件硬件化、硬件智能化”策略，将 AIoT 能力嵌入终端，提升 5G 行业通用终端智能化水平，打造雁飞系列自主化终端和行业合作伙伴定制化终端。

运营集约化：以集约化 5G 专网运营平台为支撑，实现业务交付线上化、运营管理集约化、服务支撑专业化、终端管理智能化、专网能力丰富化，

为客户提供一点运营、一点订购、一键开通、一站式服务。边缘业务平台也实现全节点及 APP 集中管控、一点创新，全网复制。构建集约化的 5G 切片平台，实现切片业务一点受理、一点签约、全网在线开通。

2 大服务底座

安全体系化：依托中国联通自主可控 5G 主动安全防护体系，由“被动”转向“主动”，向客户提供“终端-网络-运营-应用”端到端 5G 安全能力，实现全方位、全业务和全场景可信赖的保障服务。

服务一体化：以中国联通集约化平台为支撑，为客户提供一点订购、一键开通、一站式服务的体验。围绕“规建维优保”专业服务和可视、可控、可用、可维的“五可”自服务能力，构建一体化交付运维支撑体系，保障专网业务高质量交付、运维。

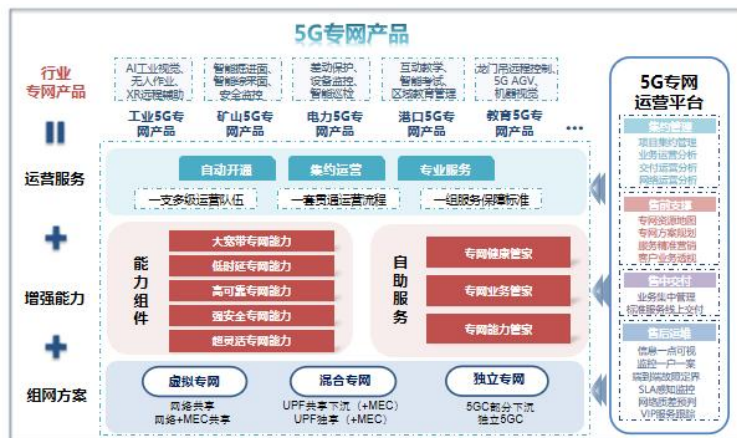


图 2 5G 专网产品

2.2 5G 专网组网形态

5G 独立专网是指利用 5G 组网、切片和边缘计算等技术，采用专有无线基站和核心网一体化设备，为客户构建一张增强带宽、低时延、物理封闭的基础连接网络，实现用户数据与中国联通公众网络数据完全隔离，且不受中国联通公众网络变化影响。该模式下客户网内业务数据及终端/用户行为信息高度保密，企业专网与联通中国联通公众网络端到端完全隔离，不受联通公众网络故障影响。从无线基站、核心网用户面+控制面端到端为客户单独建设，提供物理独享的 5G 专用网络。满足客户大带宽、低时延、高安全、高可靠的数据传输需求。

5G 混合专网是指以 5G 数据分流技术为基础，通过核心网数据面网元 UPF/MEC 的独享，无线基

站和核心网控制网元 AMF/SMF 灵活定制，实现部分物理网元独享的混合组网形式的 5G 专用网络。满足客户大带宽、低时延、数据不出园区的需求。该模式下客户网内业务数据本地卸载，可通过功能定制优化，实现客户生产业务不受联通公众网络故障影响，保障生产安全。

5G 虚拟专网是指基于中国联通 5G 公众网络资源，利用端到端 QoS 或切片技术，为客户提供一张时延和带宽有保障的、与中国联通公众网络普通用户数据隔离的虚拟专有网络。从无线基站、传输到核心网用户面+控制面端到端共享公众网络，通过 QoS 或切片技术，为用户提供具有特定 SLA 保障的逻辑专网。

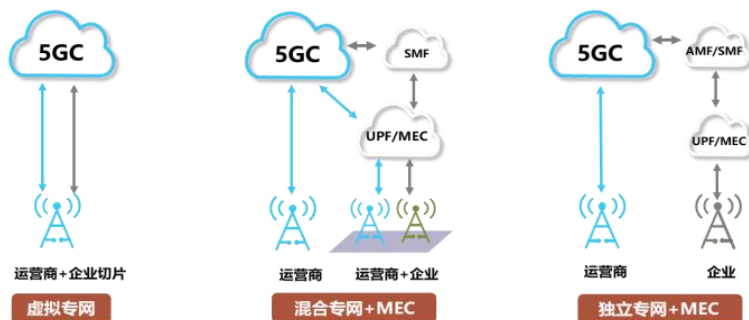


图 3 5G 专网组网形态

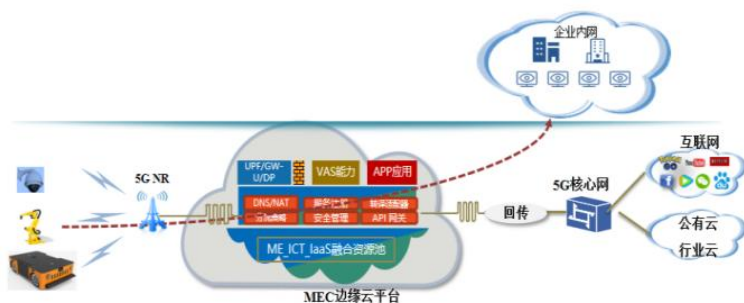


图 4 5G 专网组网拓扑

2.3 5G 专网现场级 ULCL 分流方案

业务场景：企业部署现场式 UPF，企业终端有访问园区和公网需求，进入园区通过 ULCL UPF 分流，内网通过现场级 UPF 访问，公网通过大网 UPF 访问。[1]

建设原则：园区入驻 UPF 使用 UEG 产品建设，且不出口 Internet；使用地市集中部署的 2C 大网

UPF (UDG) 作为主锚点出口 Internet。

业务流：

访问园区业务流向：园区入驻现场级 ULCL/辅锚点 UPF (N6_local) -> 专线 -> 企业园区；

访问公网业务流向：园区入驻现场级 ULCL/辅锚点 UPF (N9) -> 主锚点 UPF (N6_internet) -> 公网；

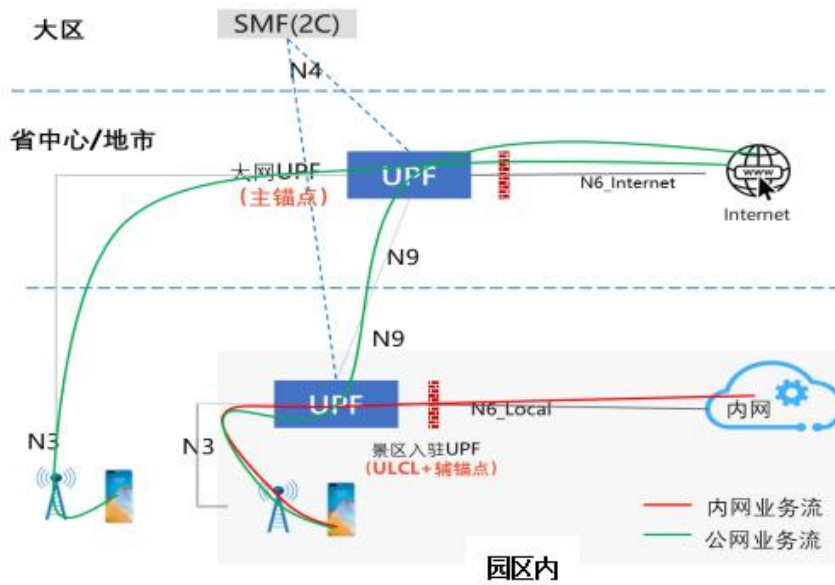


图 5 ULCL 分流示意图

2.4 5G 专网架构

结合瑞闽需求，主要应用于对网络性能，尤其是时延要求高，同时对数据管控有较高要求的工业制造行业，如 AR 远程抢修，5G+机器视觉 AGV、AR 智能点巡检等场景，提供的服务具有业务隔离、高带宽、业务加速、低时延、数据不出场等典型特征。园区 UPF 部署于临近蓝园机房内，达到私有化部署，数据不出园区的业务要求。本次提供的方案建议采用混合专网的方式部署，无线基站及 UPF 为瑞闽独立部署，核心网部分由联通建设。

联通基于最先进的 5G 技术和边缘计算 MEC 平台，在瑞闽蓝园建设部署满足覆盖要求数量的室内外 5G 基站，1 套 MEC 平台，瑞闽蓝园内的行业

终端流量在本地卸载，保障业务高带宽和低时延的同时，基于 MEC 平台能力和资源实现瑞闽蓝园各类型终端的接入管理、系统应用的部署满足机器视觉的 AGV、AR 远程维护、安全行为监控、基于机器视觉的铝材表面检测等场景的落地实施。公众流量通过 5G 核心网疏通，无法访问瑞闽内网。

联通为瑞闽分配专用的 DNN，实现基于瑞闽蓝园园区范围的网络隔离和数据不出园以及特殊场景的用网需求，充分保证生产业务数据的安全性和隔离性。

2.5 边缘云平台

2.5.1 MEC 边缘云定位

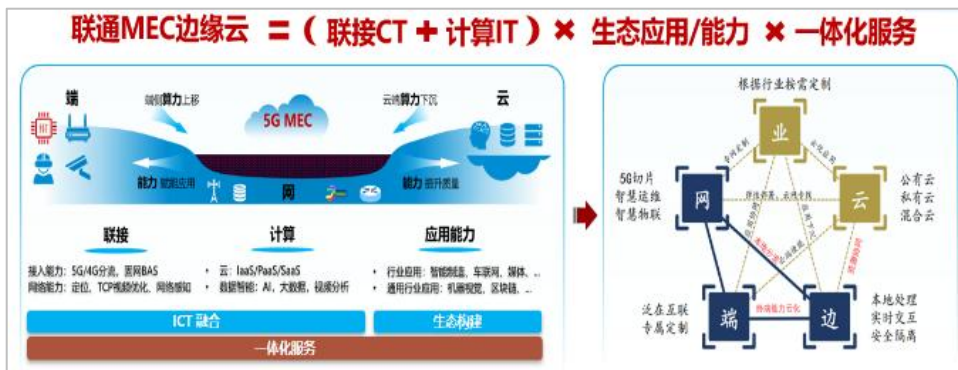


图 6 联通 MEC 边缘云

MEC 平台是网络与业务融合的桥梁，是应对 5G 大带宽、低时延、本地化垂直行业应用的关键。凭借独有的“5G+边缘网络”优势，运营商掌控着 5G MEC 的入场券。中国联通将 MEC 边缘云作为实现 CT+IT+OT 融合的锚点，基于边缘云平台结合网络联接的控制与管理能力，融合分流能力/RNIS 等 CT 能力和云 IT 能力，向应用能力和创新产品进行渗透。[2]

为了更好的满足业务实时响应、一体化交付实施的需求，中国联通集团智能网络中心与政企客户事业群联合成立 MEC BU，全国统筹调度；构建以 5G+MEC 为锚点的“云网一体化”架构，成立 1 个“MEC 业务运营中心”、10 个创新业务孵化基地、100 个行业应用，1000 个边缘节点，全力推进全国 MEC 边缘云节点规划、建设、运维、运营及 MEC 业务拓展的工作。网络与业务有效协同，力出一孔，快速在智能制造、智慧交通、智慧教育、智能安防、智慧医疗、新媒体等领域实现以 MEC

边缘云撬动 5G 项目落地形成收入。

在生态构建方面，中国联通充分发挥“云网边端业”一体化优势，构筑“一点创新，全国复制”的边缘应用商店生态，以“边缘业务平台+行业应用”为中心打通前端市场需求，助力形成高价值行业云，实现行业应用百花齐放。

2.5.2 分布式私有云

联通云分布云是在边缘计算、算力下沉背景下，为满足算力需求多样化，算力需求精细化而生，是融合计算、存储、网络、应用核心能力的分布式开放平台，通过将资源建设在更靠近业务的现场侧，并根据行业应用的要求进行智能调度，满足行业客户在敏捷联接、实时业务、数据优化、应用智能、安全保护等方面的关键需求，提供一站式靠近终端侧客户、覆盖全面的云资源，通过就近计算和处理，优化响应延迟、中心负载和算力成本。帮助数字化服务下沉，有效降低时延和成本。

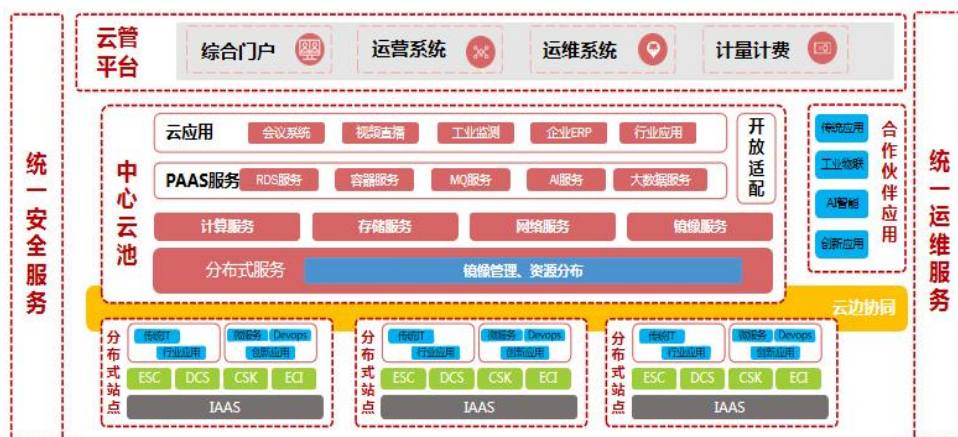


图 7 分布式私有云架构

分布式私有云的价值：

1.一站式全场景服务

分布式云支持中心和分布式站点拥有统一产品能力，支持统一服务，面向不同业务场景提供多类型，多样化的 IAAS、PAAS 服务

2.可持续进化的能力

分布式云开放接口支持多平台多场景集成，同时开放适配，整合第三方应用、PaaS 能力；此外分布式站点借助云边协同，通过“OTA”的方式下发中心云丰富云能力

3.简单直观的运维能力

分布式云的运维管理界面由中心统一呈现，形成形散神聚的运维框架，极大降低运维成本；分布式站点支持“站点自治”，具备一定的自我修复能力，保障业务稳定高效；

4.灵活高效的降本功能

适配多规格通用硬件，避免绑定；采用管控数分离架构，管理基座资源占比小于 15%，可以极大保障业务资源利用率；同时得益于自动化部署能力，可实现小时级快速交付。

应用场景：

对时延敏感的业务，在园区专属站点部署，通

过专线与企业总部的专属站点或私有云形成协同；在企业总部，可管理各分支机构的专属站点，形成资源统一管控和发放、业务统一管理和治理；在需

要通过互联网对外开服时，可动态扩容到联通云骨干云池。

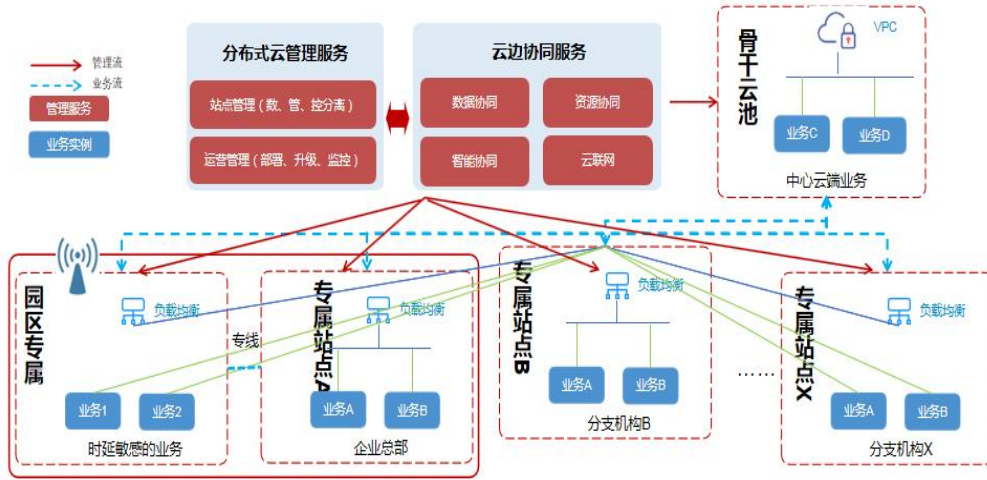


图 8 分布式私有云拓扑

3.5G MEC 前景展望

伴随着 5G 网络的商用，中国联通已建成 MEC 边缘云的运维与运营一体化支撑体系，面向全国提供 MEC 产品展示、业务申请、业务开通、故障监控、服务支撑的商业闭环。

在平台方面，中国联通统筹全局发展，开展 5G 行业登攀计划，依托业务运营中心+孵化基地的架构，项目式+产品式投资的建设模式，双轮驱动边缘业务发展，旨在构建丰富的业务能力，高质量的网络能力和完备的安全能力。支撑全国一体化运营，实现业务一点创新，全国复制，统一管理。

在生态方面，中国联通秉承开放合作的理念，携手产业各界创造一个全新的价值链和充满活力的边缘生态系统，聚焦大视频、VR/AR、工业物联网、车联网等高带宽、低时延等业务，将位置、计

费、QoS 等平台能力开放给第三方，促进 IT 与 CT、OT 融合，深入挖掘边缘网络的潜在盈利，全力构建边缘内容生态圈。

在解决方案方面，针对接入、汇聚、客户机房和户外四种场景，以及共享和专供两种使用模式，中国联通设计了 EdgePOD 一体化边缘解决方案，通过硬件和软件的配合，实现不同场景的快速交付边缘能力。

参考文献：

- [1]严会霞,曲云盛,王刚等.PCF 基于 TA 下发动态规则实现 5G ULCL 分流方案[J].通信与信息技术,2023(02):94-97.
- [1]唐海东.中国联通 MEC 边缘云平台建设分析[J].数字通信世界,2023(06):17-19+40.

智能算力在互联网与信息通信业中的融合发展研究

郑文彬

(福建省邮电工程有限公司)

摘要: 随着信息通信技术的不断发展和互联网的快速普及,智能算力在互联网与信息通信业中的融合发展已成为当前研究的热点之一。智能算力作为一种集成了人工智能、大数据分析和高性能计算等关键技术的创新型计算能力,为互联网与信息通信业带来了巨大的变革和机遇。通过对智能算力在互联网与信息通信业中的融合发展进行研究,为互联网与信息通信业的发展提供重要的理论指导和实践参考,深入探讨智能算力在该领域的应用前景,推动智能化技术与信息通信业的深度融合,实现智慧城市的可持续发展。

关键词: 智能算力; 信息通信; 行业发展分析

引言

随着人工智能产业的快速发展,通用计算集群正向智能计算集群转变,智能算力为智能化网络架构的发展提供了可能。同时运营商在 5G 时代面临着角色转变,需要建设新一代算力网络基础设施,重视边缘算力布局,发展云边协同计算新模式,并且持续关注智能网络优化和智慧城市建设两个方面。在智能网络优化方面,智能算力可应用于数据中心的智能化、云计算与边缘计算、大数据处理和分析以及信息通信网络的智能化等领域,以提升网络性能和用户体验。在智慧城市建设方面,智能算力的应用可以实现城市各项基础设施的智能化管理和优化,为居民提供更便捷、高效的生活方式。

1. 研究背景与研究意义

互联网与信息通信业已成为现代社会中不可或缺的重要组成部分,对经济、社会和个人生活产生了深远的影响。随着科技的不断进步和人工智能的快速发展,智能算力逐渐崭露头角,并正在引领互联网与信息通信业的融合发展。因此,深入研究智能算力在这一领域中的应用和发展趋势具有重要的背景和意义。

互联网与信息通信业的快速发展需要更强大的计算能力来支持其各种应用场景。传统的计算资

源往往无法满足日益增长的数据处理和分析需求。而智能算力,借助机器学习和深度学习等技术,可以实现对大规模数据的高效处理和智能分析,从而提供更精确、更快速的决策支持和服务;智能算力在互联网与信息通信业的融合中具有广泛的应用前景。例如,在智能网络优化方面,智能算力可以应用于数据中心的智能化管理,优化云计算和边缘计算的协同工作,以及提供强大的大数据处理和分析能力。随着人工智能技术的迅猛发展和全球竞争的加剧,智能算力已成为各国和企业之间竞争的重要战略要素。通过深入研究智能算力在互联网与信息通信业中的融合发展,可以帮助各国和企业把握新的发展机遇,制定相应的战略规划提升核心竞争力,实现可持续发展。

2. 智能算力的定义与发展

2.1 智能算力的发展历程

智能算力的发展历程可以追溯到计算机科学的早期阶段。自 20 世纪 50 年代以来,计算机技术一直在不断演进和创新,从最初的大型机到个人计算机的普及,再到今天的云计算和人工智能,智能算力经历了多个阶段的发展。早期的计算机是庞大而笨重的机器,主要用于执行数值计算任务。在这个阶段,智能算力的概念还未被提出,计算机主要

通过编程和算法来实现任务的自动化。然而，随着计算机体积的逐渐缩小和处理能力的提高，人们开始思考如何赋予计算机更智能的能力。

随着人工智能的兴起，智能算力进入了新的阶段。20 世纪 50 年代和 60 年代，人工智能的先驱们开始研究如何使计算机具备像人类一样的智能和学习能力。这促使了智能算力核心技术的发展，包括机器学习、模式识别和推理等。然而，由于当时计算机的处理能力和数据存储能力有限，智能算力的应用仍受到限制。20 世纪 80 年代和 90 年代，智能算力进入了新的爆发期。计算机的处理速度大幅提升，存储容量不断增加，同时出现了更加高效的算法和数据处理技术。这些进步推动了智能算力在各个领域的应用，包括图像识别、语音识别、自然语言处理等。智能算力开始在计算机视觉、语音识别和智能助理等领域展现出惊人的能力。^[1]

近年来随着大数据和云计算的兴起，智能算力进入了新的阶段。大数据的产生和存储需求对计算能力提出了更高的要求，云计算技术的发展使得分布式计算和资源共享成为可能。智能算力通过与云计算的结合，使得大规模数据的处理和分析变得更加高效和智能化。此外，深度学习等新兴技术的发展也进一步推动了智能算力的前进，使其在人工智能领域取得了重大突破。

可以预见的，未来智能算力将继续迎来新的发展机遇和挑战。随着人工智能技术的不断突破和创新，智能算力有望实现更高的处理能力和智能化水平。同时，随着物联网和边缘计算的兴起，智能算力还将在更广泛的领域发挥作用，如智慧城市、智能交通等。智能算力的发展将为社会带来更多便利和智能化的应用，推动人类社会向更加智慧的方向迈进。

2.2 智能算力的核心技术概述

智能算力作为实现智能化和自动化的关键要素，依赖于一系列核心技术的支持。这些技术涵盖了多个领域，包括机器学习、深度学习、自然语言处理、计算机视觉等。下面将对智能算力的核心技术进行概述。

机器学习是智能算力的基石之一，机器学习通过训练模型和算法来使计算机能够从数据中学习和提取规律，从而实现对新数据的预测和决策。常

见的机器学习算法包括线性回归、决策树、支持向量机等。通过机器学习，智能算力可以从大量数据中挖掘出有价值的信息，为决策提供支持；深度学习是机器学习的一个分支，通过建立深层神经网络模型来实现对数据的学习和抽象。深度学习模型由多层神经元组成，可以自动学习和提取数据的高级特征，对图像、语音、自然语言等进行识别和理解。深度学习技术的突破使得智能算力在计算机视觉、语音识别和自然语言处理等领域取得了显著的进展。

自然语言处理是指让计算机能够理解和处理人类语言的技术。自然语言处理包括语言识别、语义理解、语言生成等任务。通过自然语言处理，智能算力可以实现与人进行自然交流，包括语音助手、智能翻译、文本摘要等应用。自然语言处理技术的发展为智能算力在人机交互和信息处理方面带来了巨大的潜力。此外计算机视觉也是智能算力的重要技术之一，计算机视觉旨在使计算机能够理解和分析图像和视频数据，实现目标检测、图像识别、人脸识别等任务。通过计算机视觉，智能算力可以从图像和视频中获取丰富的信息，应用于安防监控、医学影像分析、自动驾驶等领域。

3. 智能算力在互联网与信息通信业融合中的应用：智能网络优化与智慧城市建设

智能算力在互联网与信息通信业的融合中发挥着重要作用，通过数据中心的智能化、云计算与边缘计算、大数据处理和分析、信息通信网络的智能化等应用，智能算力能够提高网络的性能、效率和安全性，推动互联网与信息通信业的智能化转型，促进智慧城市的建设和可持续发展。数据中心是互联网与信息通信业中重要的基础设施，承载着大量的数据处理和存储任务。通过智能算力的应用，数据中心可以实现自动化管理和智能优化。智能算力可以通过机器学习和深度学习技术，对数据中心的运行状态、负载情况进行实时监测和分析，从而实现资源的智能分配和负载均衡，提高数据中心的性能和效率。云计算作为一种高效的计算模式，^[2]已经在互联网与信息通信业中得到广泛应用。通过智能算力的支持，云计算可以实现更高级别的自动化和智能化。智能算力可以实现对云计算资源的智能调度和管理，根据不同应用的需求进行动态分配，

提高资源利用率和性能。同时，边缘计算作为云计算的延伸，将计算和存储能力推向网络边缘，通过智能算力的支持，可以实现边缘设备的智能化管理和协同计算，提高数据处理和响应的速度。

随着互联网和信息通信技术的发展，大量的数据被生成和积累，如何高效地处理和分析这些数据成为重要课题。智能算力通过机器学习和深度学习技术，可以实现对大数据的智能处理和分析。通过智能算力的支持，可以从海量数据中提取有价值的信息，发现潜在的关联和模式，为决策提供有力的支持和预测。信息通信网络是互联网与信息通信业的基础设施，智能算力的应用可以实现网络的智能化管理和优化。通过智能算力的支持，可以对网络的实时性、带宽和负载进行智能监测和调整，提高网络的稳定性和性能，同时智能算力还可以实现网络安全的智能防护和入侵检测，保护网络和用户的安全。

4. 互联网与信息通信业的发展趋势

4.1 近年信息通信技术的发展概述与未来发展预测

4.1.1 人工智能产业发展促使通用计算集群向智能计算集群迈进

人工智能产业的迅猛发展对计算能力提出了更高的要求，这推动了通用计算集群向智能计算集群的转变。传统的通用计算集群主要面向一般的计算任务，如数据处理、存储和传输等。然而，随着

人工智能技术的快速发展和广泛应用，对于处理大规模数据和复杂模型的需求越来越迫切，传统的通用计算集群已经无法满足这些要求。在人工智能产业的推动下，智能计算集群逐渐崭露头角。智能计算集群是针对人工智能任务而设计的高性能计算集群，它具备更强大的计算能力和智能化的特点。首先，智能计算集群拥有更高效的计算硬件，如 GPU（图形处理器）和 TPU（张量处理器），这些硬件可以并行处理大规模的数据和模型，提高计算速度和效率。其次，智能计算集群配备了更先进的软件和算法，如深度学习框架和分布式计算技术，可以实现对复杂模型的训练和推理，从而提供更智能化的计算服务。

4.1.2 智能算力为智能化网络架构的发展提供了可能

智能算力的快速发展为智能化网络架构的实现提供了广阔的可能性。传统的网络架构主要侧重于数据的传输和交换，而智能化网络架构则强调对数据的智能处理和决策能力。智能算力通过其强大的计算能力和智能化的特点，为智能化网络架构的发展打开了新的大门。智能算力通过机器学习、深度学习和自然语言处理等技术，能够从大规模的数据中提取有价值的信息和模式，实现数据的智能分析和决策，这使得智能化网络架构能够更好地理解和响应用户的需求，提供个性化的服务和优化的决策。同时智能算力的快速发展也对芯片提出了更高的要求。下表 1 为智能算力所需的主流芯片性能对比。

系统开发	芯片架构	算力	架构	速度	部署位置
软件—硬件	CPU	低	低	低	—
软件—硬件	GPU	中	中	中	云、边
硬件—软件	FPGA	高	高	高	云、边、终
硬件—软件	ASIC	高	高	高	云、边、终

表 1 智能算力所需的主流芯片性能对比

4.2 运营商的发展需要建设新一代算力网络基础建设

4.2.1 5G 时代运营商的角色转变与 NWDAF 网络架构

运营商在 5G 时代面临着角色转变，需要建设

新一代算力网络基础设施。传统上运营商主要负责提供通信服务和基础设施的建设与运维，然而随着 5G 技术的快速发展和应用的普及，运营商的角色逐渐演变为更加综合和智能化的服务提供者。NWDAF (Network Data Analytics Function)，是 5G

网络架构中的一个重要组成部分，其主要任务是分析和处理来自各个网络功能（NF）的数据，并提供分析结果给其他网络功能使用。NWDAF 是实现 5G 网络全面智能化的关键，其不仅可以从海量的网络数据中提取有价值的信息，而且还可以用这些信息来优化网络的性能和服务质量。NWDAF 在新一代算力网络基础设施建设中，提供了对网络性能的深入理解，驱动了更智能和高效的网络管理，同时也支持了更加定制化的服务。^[1]

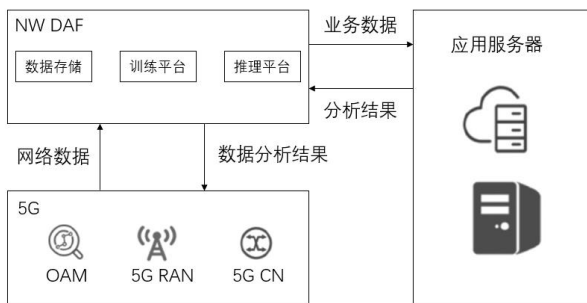


图 2 NWDAF 的逻辑架构

4.2.2 重视边缘算力布局，发展云边协同计算新模式

运营商在发展过程中需要重视边缘算力布局，以实现新一代算力网络基础设施的建设。边缘算力是指将计算能力推向网络边缘，使数据的处理和分析更加快速和高效。通过发展边缘算力，运营商可以实现云边协同计算的新模式，提供更优质的服务和更高效的网络体验。重视边缘算力布局是因为边缘计算的需求日益增长。传统的云计算模式存在一定的局限性，例如数据传输时延较高、网络拥塞等问题。^[4]而边缘计算通过将计算能力移到离用户更近的地方，可以实现数据的近距离处理和快速响应。随着物联网设备的快速发展和应用，对边缘计算的需求将进一步增加，运营商需要建设新一代算力网络基础设施来满足这一需求。

发展云边协同计算新模式有助于提升网络性

能和用户体验，云边协同计算将云计算和边缘计算相结合，充分发挥两者的优势。云计算提供强大的计算和存储能力，适合处理大规模数据和复杂模型；而边缘计算具备低时延和快速响应的特点，适用于实时数据处理和决策。通过云边协同计算，运营商可以在云端和边缘节点之间灵活调度计算任务，使数据能够在最优的位置进行处理，提升网络性能和用户体验。

5. 总结

随着人工智能产业的发展和 5G 技术的普及，运营商面临着新的发展机遇和挑战。为了实现新一代算力网络基础设施的建设，运营商需要重视边缘算力布局，并发展云边协同计算的新模式。边缘算力的布局可以满足边缘计算的需求，实现数据的快速处理和响应，提高网络性能和用户体验。运营商在建设新一代算力网络基础设施时，应重视边缘算力布局和发展云边协同计算模式，以满足日益增长的数据处理需求，提升网络性能，提供优质的服务，并确保数据安全和用户隐私的保护。这将为运营商带来新的发展机遇，并推动整个行业向智能化和创新发展的方向迈进。

参考文献:

- [1] 颜媛媛. 2030 年人工智能算力将增长 500 倍? 算力竞争时代开启 [N]. 电脑报, 2023-05-01(008).
- [2] 陈红丽. 算法算力在人工智能领域的发展与应用 [J]. 中国安防, 2023, (05): 30-33.
- [3] 高文. 高文院士: 人工智能和智能算力助力数字经济高效发展 [J]. 高科技与产业化, 2023, 29(04): 12-13.
- [4] 方正梁. 智能计算 建设算力网络的突破口 [N]. 人民邮电, 2023-04-25(004).

双谱估计方法计算精度分析

吴文兵

(福州外语外贸学院)

摘要: 为了评价双谱估计方法的精度,采用余弦函数的信号进行估计。由于余弦信号的三阶矩谱和三阶累积量谱都为零,所以无论哪种方法估计出的双谱其不为零的部分即为该估计方法的误差。通过对直接法、间接法和 AR 参数法三种估计方法的比较,得出各种估计方法产生误差的大小,并通过改变各种估计方法中不同参数的取值,观察参数变动的情况下其双谱误差数值的改变,从而为各种双谱估计方法优劣评估以及选用提供一种判断依据。

关键词: 高阶矩, 双谱估计, 误差, 精度

引言

高阶累积量可以自动抑制高斯背景噪声的影响,正因为这样,高阶累积量这一统计量已日益受到人们的重视并已成为信号处理中一种非常有用的工具。由三阶累积量得出的双谱包含了信号的非对称、非线性信息,可以用来描述非线性相位耦合,尤其是二次相位耦合,已经在故障诊断中得到了广泛应用^[1-4]。

谱估计的现代史是从 Tuckey 于 1949 年的突破开始的。Blackman-Tuckey 给出了用 Wiener 相关法从采样数据序列得到功率谱估计的实现方法——BT 法。

1965 年 FFT 的出现产生了周期图法;BT 法与周期图法的致命弱点是频率分辨率的限制。为了克服这一缺点,1967 年 Burg 受到在地震应用研究中线性预测方法的启发,导出了最大熵谱估计法;E.Parzen 于 1968 年正式提出了 AR 谱估计方法;此后十几年来发展了许多高分辨的谱估计方法,称为现代谱估计方法,而 BT 法与周期图法称为传统谱估计法。

无论哪种方法,每一种谱估计技术都可以认为是一种模型法。具体地说,就是根据对过程的先验知识建立一个近似实际过程的模型,其次,利用观察数据或自相关函数来估计假设的模型参数,最后

作谱估计。

常用的模型有:周期图与 BT(正弦谐波总和模型)、AR、MA、ARMA、Prony、最大似然等。各种估计之间的性能变化,就是由于假设的模型与实际过程拟合的好坏不同引起的。虽然不同模型可以产生类似的结果,但是有些模型所需要的参量可能相对少一些。

实际中,一般利用有限长度数据估计一过程的高阶累积量谱或高阶矩谱。一般有两种主要方法:

- (1) 常规(Fourier 型)方法
- (2) 参数方法:AR、MA、ARMA 等方法

本文讨论常规方法以及 AR 参数法的双谱估计效果。常规方法可分成下列两类:

- (1) 间接法:按照式(6)或式(7)的定义进行估计;
- (2) 直接法:按照式(8)的定义进行估计。

由于各种估计方法都不可避免存在一定误差,本文将利用对余弦信号的双谱估计对这些方法产生的误差进行定性分析。

1. 高阶累积量

设 $\{x(n)\}$ 为零均值 k 阶平稳随机过程,则该过程的 k 阶累积量 $c_k(\tau_1, \tau_2, \dots, \tau_{k-1})$ 定义为随机变量 $\{x(n), x(n+\tau_1), \dots, x(n+\tau_{k-1})\}$ 的 k 阶联合累积量,即

$$c_{kx}(\tau_1, \tau_2, \dots, \tau_{k-1}) = \text{cum}\{x(n), x(n+\tau_1), x(n+\tau_2), \dots, x(n+\tau_{k-1})\} \quad (1)$$

而该过程的 k 阶矩 $m_{kx}(\tau_1, \tau_2, \dots, \tau_{k-1})$ 则定义为随机变量 $\{x(n), x(n+\tau_1), \dots, x(n+\tau_{k-1})\}$ 的 k 阶联合矩, 即

$$m_{kx}(\tau_1, \tau_2, \dots, \tau_{k-1}) = \text{mom}\{x(n), x(n+\tau_1), \dots, x(n+\tau_{k-1})\} \quad (2)$$

这里, $\text{mom}()$ 表示联合矩, 三阶累积量为:

$$c_{3x}(\tau_1, \tau_2) = E\{x(n)x(n+\tau_1)x(n+\tau_2)\} \quad (3)$$

k 阶累积量谱定义为 k 阶累积量的 $k-1$ 维 Fourier 变换, 即

$$C_{kx}(\omega_1, \dots, \omega_{k-1}) = \sum_{\tau_1=-\infty}^{\infty} \dots \sum_{\tau_{k-1}=-\infty}^{\infty} c_{kx}(\tau_1, \dots, \tau_{k-1}) \exp[-j \sum_{i=1}^{k-1} \omega_i \tau_i] \quad (4)$$

k 阶矩谱定义为 k 阶矩的 $k-1$ 维 Fourier 变换, 即

$$M_{kx}(\omega_1, \dots, \omega_{k-1}) = \sum_{\tau_1=-\infty}^{\infty} \dots \sum_{\tau_{k-1}=-\infty}^{\infty} m_{kx}(\tau_1, \dots, \tau_{k-1}) \exp[-j \sum_{i=1}^{k-1} \omega_i \tau_i] \quad (5)$$

最常用的高阶谱是三阶谱(又称双谱), 估计式如下:

$$B_c(\omega_1, \omega_2) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} c_{3x}(\tau_1, \tau_2) e^{-j(\omega_1 \tau_1 + \omega_2 \tau_2)} \quad (6)$$

$$B_m(\omega_1, \omega_2) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} m_{3x}(\tau_1, \tau_2) e^{-j(\omega_1 \tau_1 + \omega_2 \tau_2)} \quad (7)$$

对于周期性功率信号, 一般估计其高阶矩谱, 令

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-i \frac{2\pi}{N} nk} \quad 0 \leq k \leq N-1$$

则由式(7)可得其双谱可按式计算:

$$B_x(k_1, k_2) = \frac{1}{N} X(k_1) X(k_2) X^*(k_1 + k_2)$$

2. 双谱估计方法

a: AR 参数法

假设 $y_1(t)$ 是系统实际输出信号 $y(t)$ 经过去噪后的信号, 系统输出的随机振动信号是由均值等于零的非高斯的白噪声 $a(t)$ 造成的, 建立 AR 模型:

$$y_1(t) + \sum_{i=1}^p \psi_i y_1(t-i) = a(t), (t=1, 2, \dots, N)$$

式中的 $\psi_i (i=1, 2, \dots, p)$ 为自回归系数, p 为自回归模型的阶数。AR 双谱表达式

$$B^{AR}(\omega_1, \omega_2) = \gamma_{a3} H(\omega_1) H(\omega_2) H^*(\omega_1 + \omega_2)$$

$H(\omega)$ 为传递函数,

$$H(\omega) = \frac{1}{1 + \sum_{i=1}^p \psi_i e^{-j i \omega}}$$

对 $y_1(t)$ 运用 AR 模型的参数法估计出模型系数 $\psi = [\beta]$, AR 双谱幅值表达式

$$|B^{AR}(\omega_1, \omega_2)| = \frac{|\gamma_{a3}|}{\left| 1 + \sum_{i=1}^p \beta_i e^{-j i (\omega_1 + \omega_2)} \right| \prod_{k=1}^2 \left| 1 + \sum_{i=1}^p \beta_i e^{-j i \omega_k} \right|}$$

要估计出模型系数, 必须先对模型进行定阶。奇异值分解(简称 SVD)是一种模型定阶方法,

归一化 Frobenius 范数定阶法是一种基于奇异值分解的定阶法。归一化 Frobenius 范数定义为:

$$\mu(p) = \frac{[\lambda_1^2 + \lambda_2^2 + \dots + \lambda_p^2]^{1/2}}{[\lambda_1^2 + \lambda_2^2 + \dots + \lambda_n^2]^{1/2}}, p=1, 2, \dots, n. \quad (9)$$

式中的分母就是 Frobenius 范数。显然, $\mu(p) \leq 1$ 。

令 $\alpha \approx 1$ 是一个接近 1 的门限值, 当 $\mu(p) \geq \alpha$ 时,

认为前 p 个奇异值占主导地位, 即选择满足此条件的最小 p 值作为矩阵 A 有效秩(即 AR 阶数)的估计结果。典型地, 常选择 $\alpha = 0.995$, 即 $\mu(p) \geq 0.995$

时所对应的最小 p 值即为 AR 模型的最佳阶数。

b: 直接法

对于确定性信号, 双谱估计直接法步骤为:

步骤 1. 将数据分为 K 段, 每段含 M 个样点, 即 $N=K \cdot M$, 且每段数据去均值。如果为确定性信

号,则仅用一个记录(N=M)。如果为了获得 FFT 算法适合的长度,可对每段数据补零。

步骤 2. 假设 $\{x_i(q)\}, q=0,1,\dots, M-1$ 是第 i 段数据, 计算 DFT 系数

$$X^i(\lambda) = \sum_{n=0}^{N-1} x^i(q) e^{-i\frac{2\pi}{M}q\lambda} \quad 0 \leq \lambda \leq M-1, 1 \leq i \leq K$$

步骤 3. 最后 K 段数据进行平均得到给定数据的双谱

$$B_x(\lambda_1, \lambda_2) = \frac{1}{K} X(\lambda_1) X(\lambda_2) X^*(\lambda_1 + \lambda_2)$$

c: 间接法

对于确定性信号, 双谱估计间接法步骤为:

步骤 1. 将数据分 K 段, 每段 M 个样点, 即 $N=K \cdot M$

步骤 2. 每段数据去均值;

步骤 3. 假设 $\{x_i(q)\}, q=0,1,\dots, M-1$ 是第 i 段数据, 计算三阶矩

$$M_3(\tau_1, \tau_2) = \frac{1}{M} X^i(q) X^i(q + \tau_1) X^*(q + \tau_2)$$

步骤 4. 按照式 (7) 进行双谱估计。

3. 余弦信号的双谱

设有余弦信号 $x(n)=\cos \omega t$, 由三阶累积量计算公式

$$c_{3x}(\tau_1, \tau_2) = E\{x(n)x(n+\tau_1)x(n+\tau_2)\}$$

可知, 余弦信号的三阶累积量为 0, 所以无论按照式

(7) 还是式 (8), 其计算出的双谱都为 0. 参见文献 (5) 和文献 (6)。

4. 余弦信号双谱估计误差分析

由上面分析可知, 余弦信号无论是求其三阶累积量谱还是三阶矩谱都为零。现根据上述三种估计方法进行分析。由于已知余弦信号的双谱为零, 所以无论哪种方法估计出的余弦信号双谱, 只要出现不为零的部分, 都可以视为该种估计方法产生的误差。

首先由 AR 参数法进行估计。

本文取余弦信号 $y(t)=\cos(2\pi f_0 t)$, 在 f_0 分别为 1HZ, 5HZ 和 50HZ, 采样频率为 1000HZ、采样时间

为 1 秒的情况下, 分别采用 AR 参数法、直接法和间接法进行双谱估计, 分别示于图 1、图 2 和图 3。从上述图像可以看出, 在 f_0 为 1HZ 时, 估计出的双谱非零部分分布范围最小, 但数值较大。因此, 在采样频率一定的情况下, 随着信号频率的增加, 无论何种估计方法, 其估计出的双谱误差其分布范围都是增加的, 但数值却是减小的趋势。再在保持信号频率为 1HZ 不改变的情况下, 采样频率分别采用 100HZ、500HZ 和 2000HZ 时, 三种方法估计出的双谱分别图示于图 4、图 5 和图 6。从图中可以看出, 在信号频率固定的情况下, 当采样频率从 100HZ 变为 500HZ 的时候, 参数法和直接法的估计结果都呈现出较大的变化, 就是估计精度提高得比较多, 而间接法则提高得比较少。再从 500HZ 变化到 2000HZ 的时候, 三种方法估计的结果改变得都很少。图 7 和图 8 表示的是在信号频率固定为 1HZ、采样频率固定为 1000HZ 的情况下, 通过改变直接法和间接法中步骤一的数据分段数目得到的结果。图中显示, 随着数据分段数从 64 段变到 16 段, 两种估计方法得到的结果, 其精度都得到了较大的提高。对比图 2-a 和图 3-a, 两者表示的都是信号频率为 1HZ, 数据分段数目等于 8 段的结果, 再对照图 7-a 和图 8-a, 其数据分段数目等于 4 段, 其两种方法估计出的结果差别已经不大。在 AR 参数法中, 再通过改变式 (9) 归一化 Frobenius 范数值, 得出的双谱如下所示: 图 (9) 和图 (10) 的图像都是在信号频率为 50HZ, 采样频率为 1000HZ 的情况下得出的。将图 (9) 和图 (10) 与图 1-c 对比, 图 (9) 和图 (10) 的 Frobenius 范数值 $\mu(p)$ 分别等于 0.8 和 0.999, 图 1-c 的范数值等于 0.995, 通过对比可以看出, 在 AR 参数法中, Frobenius 范数值的改变也会导致估计精度的改变, 当 Frobenius 范数值从 0.8 改变到 0.995 的时候, 其估计精度变化较大, 因为图 (9) 与图 1-c 相比, 双谱非零值的分布范围明显更广, 而当 Frobenius 范数值从 0.995 改变到 0.999 的时候, 则变化不明显, 这也间接证明了前述的 Frobenius 范数值 $\mu(p)$ 以 0.995 作为阈值的合理性。从上述分析结果可以看出, 相对来说, 参数法的估计效果要略好于直接法和间接法。

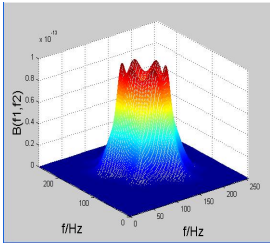


图 (9)

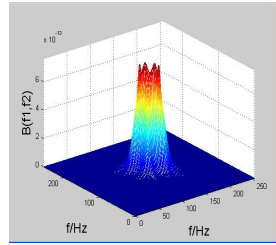


图 (10)

5.小结

本文通过对双谱理论值为零的余弦信号采取三种方法进行双谱估计,分析了各种估计方法导致的误差,并通过改变参数,分析不同的估计方法在不同情况下误差的变化,从而为各种估计方法的精度提供一种判断标准。

参考文献

[1] 张贤达. 现代信号处理[M]. 清华大学出版

社,2002,263-281.

[2] 吴文兵,黄宜坚. 基于双相干谱的减压阀故障诊断. 计算机测量与控制,2011(10)

[3] 邵忍平,黄欣娜,刘宏昱. 基于高阶累积量的齿轮系统故障检测与诊断[J]. 机械工程学报,2008,44(6): 161-168.

[4] Dusan Kocur, Radoslav Stanko. Order bispectrum: a new tool for reciprocated machine condition monitoring [J]. Mechanical Systems & Signal Processing, 2000, 14(6): 871-890.

[5] 张严,王树勋. 非线性相位耦合的切片谱分析方法[J]. 电子学报,1998,26(10):104-109.

[6] YANG Jiangtian, ZHOU Lihua, CHEN Jiayi. Higher order spectra and its application in machinery fault diagnosis. Transactions of tianjin university Vol. 5 No. 2 1999,5(2):190-194.

新时期算力基础设施发展的思考

曹 曦

(中国电信股份有限公司福建分公司)

摘 要：随着数字经济的发展，算力成为支撑大国科技竞争和产业创新的关键要素。在 ChatGPT 掀起的 AIGC 浪潮吸引社会各界对于算力发展强烈关注的时代大背景之下，本文尝试从算力基础设施的概念内涵及发展面临的挑战分析入手，提出新时期算力基础设施发展的思考与建议，为行业及我省算力基础设施高质量发展提供借鉴参考。

关键词：算力，算力基础设施，数据，大模型

抬头刷脸轻松地进入住宅小区，戴上“眼镜”“亲身”遍游壮美河山，质检机器在产线上核验成品替代人工既高效又准确还能 24 小时在线……，这看得见的一幕幕是令人称奇的数字化应用，而看不见的是支撑这些应用实现的强大算力！

算力，是数字经济时代的核心生产力，也是衡量国家综合竞争力的重要指标之一。作为算力重要载体的算力基础设施，肩负着新时期全面贯彻新发展理念，为加速经济社会数字化转型和打造国际竞争新优势提供更加强劲、普适、普惠和绿色低碳算力的时代重任，使命光荣、意义重大！

1. 算力基础设施的概念与内涵

算力基础设施是数字经济时代，集信息计算力、数据存储力、网络运载力于一体的新型信息基础设施。

《数字中国建设整体布局规划》做出“引导通用数据中心、超算中心、智能计算中心、边缘数据中心等合理梯次布局”的战略部署，国家发改委等部门印发《关于加快构建全国一体化大数据中心协同创新体系的指导意见》规划了打造“数网”、“数纽”、“数链”、“数脑”、“数盾”五位一体的顶层设计布局，为新时期算力基础设施发展指明了方向，明确了路径。

2. 新时期算力基础设施发展面临的挑战

(一) 期待更加强劲的算力

随着 5G、移动互联网、物联网等信息技术的

广泛应用，以及 ChatGPT 等大模型训练、推理需求的增多，数据量暴涨和算力需求激增已然成为事实。数据显示，ChatGPT 总算力消耗约为 3640PF-days，需要八个算力为 500P 的数据中心才能支撑运行。IDC 最新发布的 Global DataSphere 2023 显示，中国数据量规模将从 2022 的 23.88ZB 增长至 2027 年的 76.6ZB，复合年均增长率（CAGR）达到 26.3%，居全球第一。



图 1 全球及中国数据量规模增长示意图

海量规模数据为 ChatGPT 等 AI 大模型训练推理提供了丰富的“养料”，推动 AI 大模型加速成熟，同时也进一步加剧全球数据量暴涨和推高算力需求，据预测，2030 年 AI 爆发带来的算力需求将增长 500 倍！显然，巨量、急速攀升的算力需求仅仅依靠当前孤岛、点状式分布的数据中心已是难堪大任，新形势新发展的需求，不单单是算力布局、

算力规模、算力结构面临严峻挑战，而且对于算力基础设施的算力通达、算力响应、存储容量、存储类型等核心能力以及云网边端协同、算存运融合等都提出了更多、更高、更具挑战性的要求。

（二）期待更加经济适用的算力

席卷全球的“百模大战”使得英伟达 GPU 成了抢手货，A100 GPU 单价高达 15 万元人民币，即便是面向中国的阉割版本 A800，单价也高达 9.5 万元，还常缺货！以 GPT-3 为例，支撑其开发和训练就需要 3 万多片 A100，一次模型训练的成本在 200-1200 万美元，初始投入成本就达 8 亿美元，训练期间电费开支大约每日 5 万美元，训练时长达数周到数月。显然即便能顺利买得到足够数量的 A100，也只有少数头部企业才具备“玩得起”的实力，大模型的“烧钱”是行业的共识，成为横亘在科技企业、尤其是初创企业面前的“大拦路虎”！但大模型最终要走向开源化、商业化、生态化落地，而且随着通用大模型面向行业及大众化等应用场景增多，蕴藏着的商机和演绎出的盈利模式将吸引更多企业涌入，激发出更加丰富多样的新应用、新业态、新模式，同时也将进一步提升市场对于更泛在普及、更经济适用且安全稳定的算力供给的迫切期待，要求算力基础设施加快前瞻性演进和创新变革，响应市场呼声丰富算力供给。

（三）期待更加绿色低碳的算力

数据中心是备受诟病的耗电大户，支撑大模型的 AI 服务器则尤为更甚，单张英伟达 A100 GPU 卡功耗为 300-400W，H100 GPU 卡为 500-750W，以万张 A100 或 H100 为起步的 AI 算力中心的耗电量可想而知！根据斯坦福人工智能研究所（HAI）发布的《2023 年人工智能指数报告》，OpenAI 的 GPT-3 仅单次训练的耗电量就高达 1287 兆瓦时，而且 AI 训练的电力需求每 3-4 个月就翻一倍！面向大众开放一个月的 ChatGPT 仅电量消耗就高达 150 万美元。随着参数量更大、性能更优的 GPT-4 及后续升级版本的陆续推出，必然带来更加令人恐怖的电力消耗！显然，我国的算力基础设施发展不能再重走 OpenAI 这样“高算力、高耗能”的路子，算力基础设施的发展应当在满足市场需求的同时，逐步提升太阳能、风能等绿电能源的使用比例，坚定地朝着绿色化、低碳化方向演进发展。

3.新时期算力基础设施发展的思考

未来，算力基础设施应具备在云、边、端之间按需分配、灵活调度计算、存储和网络资源的整体统筹能力，加快云边端协同、算存运融合的一体化算力基础设施建设与发展成为打通数字基础设施大动脉，夯实经济社会高质量发展数字底座的必然选择。结合新时期的挑战与发展，提出以下三个方面的建议。

（一）从关注规模增长向关注性能质量提升转变，提供优质充沛算力

我国算力规模一直保持高速增长，截至 2022 年底，算力总规模已达 180EFLOPS，近五年平均增速超过 30%，算力规模排名全球第二。为适应新时期发展需求，行业在关注规模增长的同时，需要朝着更加关注性能质量提升，实现高质量发展方向演进，重点聚焦以下三个方面：

一是加快算力资源池化。长期以来，算力碎片化、上架率低、忙闲不均等问题困扰行业也影响算力价值的提升，要从根本上解决这些问题就需要从顶层设计入手，加强算力基础设施的整体规划布局，加快算力资源朝着供需统筹协调有序、综合算效最优方向发展，即以省为单位推动分布在省内的政府建设的数据中心、基础电信运营商建设的数据中心以及社会资本建设的各种类型大中小数据中心之间全光互联，加快算力并网，发挥算力集群的规模化、集约化效应。

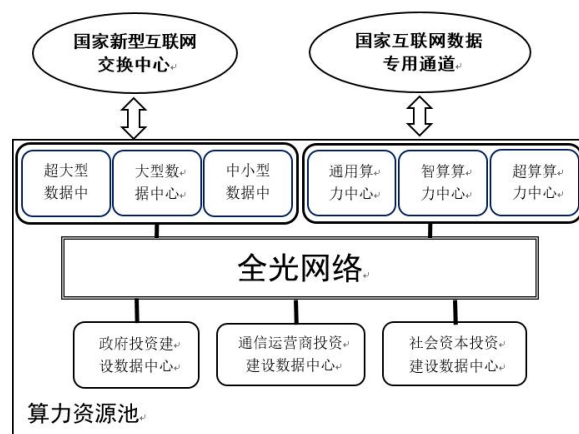


图 2 算力资源池化示意图

算力池化主要有三方面的优势，一是有助于打破算力孤岛，实现集约规模发展。只有不同地域分

布的数据中心节点之间实现物理上联通才有可能形成数据要素的高效流通,发挥出数据要素的巨大潜在价值。二是有助于算力资源的统筹协调,实现需求与供给之间最佳匹配,提升整体算效和算力资源的整体利用率。同时也促进社会闲散算力资源的有效利用,避免重复建设与资源浪费。三是有助于算网一体整体布局规划和优化提升服务能力,通过国家新型互联网交换中心或国际互联网专用数据通道升级算力资源池到省内产业集聚区以及到国家算力枢纽节点或海外算力节点的通信链路性能,有条件的区域甚至可以实现一跳入算,进一步改善算力应用体验。

二是优先发展智算算力。根据 2022 中国算力大会发布的数据,无论是中国还是全球,通用算力依然占“大头”,占比高达 78%,而作为 ChatGPT 类大模型重要支撑的智算算力占比仅为 21%,显然这样的算力结构难以支撑“百模大战”的训练推理和各类大模型行业化应用需求,因此算力结构亟需从当前以通用 CPU 密集型为主向以 GPU、FPGA、ASIC 等专用芯片为主的并行计算密集型转型,将发展智能计算作为优先方向,加快智算算力为主的算力基础设施建设,为大规模人工智能发展和 AI 大模型应用落地提供坚实支撑,促进人工智能赋能千行百业。

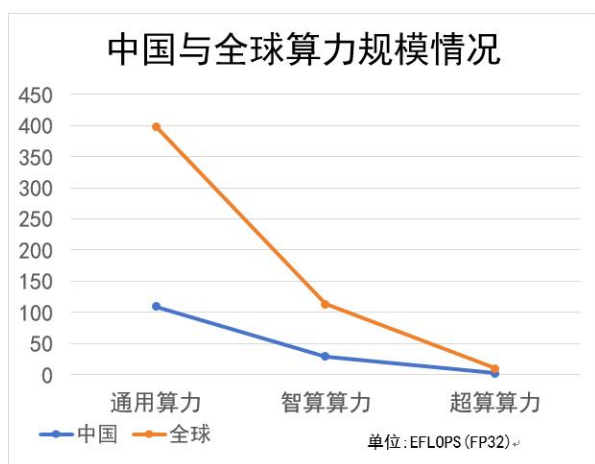


图 3 中国与全球算力规模情况示意图

三是提升数据中心机架密度。通常,数据中心一般采取增加场地空间、扩大机架规模和增加服务器数量方式来应对日益增长的算力需求,但这样的部署方案会增加数据中心运营成本,且限制了部分

场地受限数据中心的发展,因此,部署高密度服务器等提升机架密度的升级方案应运而生,高密度服务器可在非常小的物理空间里面集成更多的处理器和输入/输出扩展能力,显著提升数据中心单位面积算力性能,同时由于机架密度的提升又带来散热难题,因此往往需同步采用新型液冷技术,相较于传统的风冷技术,液冷具有降低能耗、增强系统性能稳定以及支持更高性能处理器的优势。

(二) 从独立并行发展向体系协同融合发展转变,推动算力泛在普及

由互联网、金融、教育、交通等行业领衔,越来越多的算力行业应用正在落地深耕,产业对于可以随时随地使用且更加经济实惠、方便易用的算力服务需求日益增多,要求算力基础设施具备更加泛在化、普惠化和普适化的算力供给能力,推动算力基础设施向着类似水、电一样随用随取、唾手可得的公共社会化服务设施演进。为此,需要重点关注以下三个方面:

一是推动算力泛在化。算网融合是实现算力泛在化的重要基础,长期以来,算力和网络是两个各自独立发展的垂直领域,要实现融合首先必须先从算网协同入手,建立完善云资源接入和一体化调度机制,使得多元异构、海量泛在的算力基础设施,基于全光网络连接成为一体化算网技术服务体系,推动算力与通信网络在形态、协议等方面逐步趋向融合,逐渐形成一体化信息基础设施,最终迈向真正意义的算网融合。算网融合是一个长期的、渐进式的融合迭代过程,除了涉及算力、网络、安全、应用等标准、技术方面的亟待突破以及产业链支持外,还会面临着算力度量、标识等方面的难题,需要行业及产业各方共同努力。

二是推进算力大众化。核心在于算网一体化服务,算网一体化服务是推动算力从面向大企业、专属行业扩展到面向包括中小企业、初创小微企业在内提供“一点接入、即取即用”的社会级公共服务,降低算力应用门槛的关键环节。基于算网一体化服务,对外以简单、便捷、易用的方式受理社会各界订单式、任务式算力服务需求,对内依托算网一体化调度平台体系将不同类型(通用算力、智算算力、超算算力、边缘算力等)、不同归属主体(政府、通信运营商、民企等)以及不同区域(沿海与山区,东部与西部等)的算力资源统一纳管和统一调度,

极大地提升了算力资源的整体统筹和全局配置能力，结合算力服务需求自动匹配最优资源，按照统一的标准进行全局编排、调度、封装和提供端到端一致性质量有保障的算力服务。

三是探索算力普惠化。全球人工智能发展浪潮推动算力逐渐改变长期以来仅作为计算工具的基本属性和角色定位，朝着推动产业结构升级和打造经济社会数字化创新引擎方向迈进，然而，重金难求的英伟达 GPU、高不可攀的 AI 大模型训练成本却难倒了众多“英雄汉”，大模型之热的同时引发了行业及市场对经济实惠算力服务需求的迫切渴望。众所周知，成本是决定服务价格的关键决定性因素之一，算力成本主要取决于数据中心运营成本，而用电成本占据了数据中心运营总成本的 60-70% 之多，因此降低数据中心用电成本是推动算力服务价格下降、实现算力普惠的关键所在！当前，随着光伏模组市场价格大幅下调以及太阳能、风能、核电等清洁能源装机并网规模增大，福建等一些省份出现电力供给盈余增多，急需电网提升消纳能力才能有效化解。与此同时，国家也在加快构建以新能源为主体的新型电力系统，并且通过数据中心政府采购标准中增设“可再生能源最低使用率”等年度递进式量化指标等要求，引导行业加快绿色数据中心的建设，推动风电等绿色清洁能源为数据中心赋能，有望大幅降低数据中心用电成本，使得算力如用水、用电一样大众化、普惠化的时代不再是遥不可及，更不会“高”攀不起！

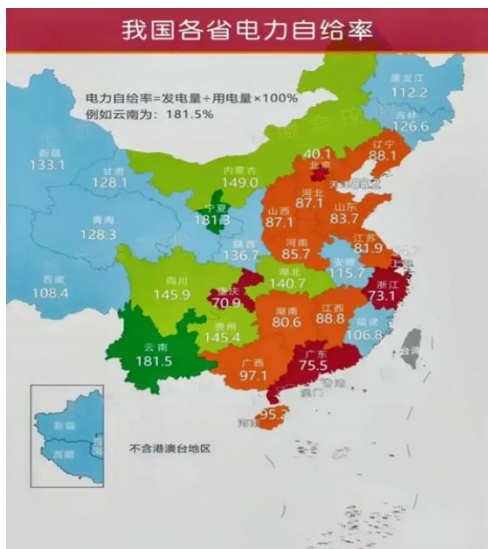


图 4 各省电力自给率示意图

(三) 从高能耗向绿色低碳转变，发展绿色算力

近期，中央深改委会议审议通过《关于深化电力体制改革加快构建新型电力系统的指导意见》，强调要加快构建清洁低碳、经济高效、供需协同的新型电力系统，为绿电赋能数据中心加快算力普适化普惠化进程，反哺经济社会高质量发展提供了最新指引。位于海峡西岸的福建，具有绿电赋能数据中心发展的巨大潜力和独有的优势。

据国际可再生能源署 (IRENA) 统计，近些年来，光伏发电成本保持持续下降趋势，光伏度电成本已低于风电、天然气，成为最具有竞争力的电力供给方案。根据中国光伏行业协会的监测数据，从 2023 年下半年开始光伏组件等价格大幅下跌，带动光伏 LCOE (平准化发电成本) 大幅走低，年均光照大于 1800 小时的宁夏、西藏、新疆等西部省份，到 2025 年左右光伏发电成本可能降到 0.1 元/度。而年均光照时长为 1200-1300 小时的广东等东部省份，2024-2025 年后光伏发电成本降为 0.25 元/度会是大概率事件。根据福建省气候公报数据，2022 年全省年均日照时数为 1829.2 小时，2021 年为 1935.8 小时，即便参照广东 0.25 元/度这样的发电成本，也仅相当于当前福建数据中心用电单价的 40%，就已经能为规模庞大的现网数据中心节省出相当可观的年用电成本了。



图 5 光伏发电成本趋势图

风能在福建有更大的发展与应用潜力，两岸高耸的武夷山脉和阿里山脉形成的“狭管效应”，使得来自大陆或者大洋的大范围气流经过海峡就如同从风箱里吹出来一样，更加集中且风力增强。强劲的海风，再叠加常年的冬夏季风、海陆风和热带气旋等的影响，使得福建沿海风能资源优势非常突出，全年 7 级以上大风的天数可达 100 多天，平潭

岛、南日岛、东山岛等沿海岛屿都是全国风力最强劲的海岛。更关键的是海上风电发电价格也在持续跳水,2022 年 7 月福州马祖岛和连江两个外海风电场竞配电价已降到 0.204 元/千瓦时、0.193 元/千瓦时,不但低于福州当地的燃煤发电基准价(0.3932 元/千瓦时),甚至比全国最低的新疆煤电上网标杆电价(0.25 元/千瓦时)还更低。随着十四五时期更多的太阳能发电、海上风能以及漳州等地核电机组陆续投产,不但将进一步增大福建的电力供给富余量,还将面临更大的电网消纳能力掣肘的尴尬!那么,作为长期稳定运行的耗电大户,快速发展的数据中心无疑为海上风电消纳提供了一个全新的解决思路,推动数据中心企业与海上风电等绿电生产企业在“源网荷储一体化”等方面形成更紧密的合作,既符合国家“双碳”战略和深化电力体制改革的政策导向,也是福建算力基础设施绿色化、低碳化发展大有可为的探索路径。将福建丰富的绿电资源优势转换为更加强劲澎湃、触手可及、简便易

用、经济实惠的算力引擎动能,驱动经济社会高质量发展跃迁变革。

参考文献:

- [1]福建省气候公报(2022 年).福建省气候公报(2021 年).[R]中国天气网福建站,2023-04-11.2022-03-15.
- [2]《中国综合算力指数(2022 年)》.[R]2022 中国算力大会,2022 年 7 月.
- [3]《中国算力白皮书(2022 年)》.[R]2022 中国算力大会,2022 年 7 月.
- [4]《中国存力白皮书(2022 年)》.[R]2022 中国算力大会,2022 年 7 月.
- [5]《中国运力白皮书(2022 年)》.[R]2022 中国算力大会,2022 年 7 月.

基于全过程工程咨询模式的数据中心建设项目探索与应用

高扬明

(福建省邮电规划设计院有限公司)

摘要:近年来,数据中心作为新基建的代表发展较快。同时,数据中心工程建设特点不同于普通的建筑工程,必然需要一种与之相匹配的项目管理模式。本文通过对数据中心项目建设特点、难点、各阶段资质要求等进行分析,探讨全过程工程咨询管理模式在数据中心建设项目中的应用,以为同类项目建设提供参考。

关键词:数据中心 建设投资 全过程工程咨询

前言

2020年3月4日中共中央政治局常务委员会召开会议,会议指出要加快5G网络、数据中心等新型基础设施建设进度。数据中心项目作为新基建的重要组成部分,进入了大规模发展阶段。数据中心项目是一个具有建设投资及运营成本高、建设周期长、多专业交叉及综合与分阶段建设实施等特点的复杂系统工程。同时随着我国建设工程领域越来越规范化和建设单位服务需求越来越精细化,近两年全过程工程咨询作为国家鼓励以及推行的新型服务模式在房屋建筑和市政基础设施领域建设中逐渐得到推广应用。本文主要探讨全过程工程咨询管理模式在数据中心建设项目中的应用和实践,以为同类项目建设提供参考。

1. 数据中心项目建设特点

(1) 建设投资及运营成本高

数据中心项目的建设投资主要包括土地购置、建筑工程、市电引入、机电工程、室外工程、通信工程等,运营成本主要包括维护成本、运营水电物

业等。数据中心本体投资(主要包括建筑、机电、室外工程、通信设备)基本上差距不大,土地购置和市电引入投资由于项目选址差异而不同,一般而言5KW数据中心按可安装1000个机架建设投资约2亿元,折合单机架投资在20万元左右。

(2) 建设周期长

数据中心建设项目从前期的建议书、可行性研究报告、勘察、设计(包括建筑设计和工艺设计)、招标、手续办理等工作,到开工建设、竣工验收并交付使用,建设周期一般在24个月左右。

(3) 分阶段建设实施

数据中心的项目建设分为三个阶段,项目决策阶段、工程实施阶段和项目运营阶段。项目决策阶段主要包括项目立项、编制可行性研究报告等,工程实施阶段主要包括数据中心机房的设计、施工以及设备的安装等,项目运营阶段主要包括设备机柜的安装、设备调试和试运行、试验等。

(4) 多专业交叉及综合

数据中心建设项目涉及总平布置、建筑结构、暖通空调、给水排水、供配电系统、消防系统、智

作者简介:

高扬明,毕业于福州大学土木工程专业,学士学位,中级工程师,注册监理工程师、注册咨询工程师、中国通信学会会员,长期从事通信建筑以及通信基站配套方面的设计、研究工作,目前任职于福建省邮电规划设计院有限公司。

能化及安防系统、通信及电源系统等，各个专业系统之间既相互独立又相互关联。前期工程与后期工程的设计执行的技术标准不同，前期工程和后期工程又存在较多的相互制约因素，因此反映出数据中心建设项目的复杂性。

2. 数据中心项目建设难点

通过分析数据中心项目投资构成，数据中心主楼建筑工程、机电安装工程、室外土建配套工程的投资基本相差不大，主要受单机架功率和机架数量的影响。土地购置费用和外市电引入投资因各个数据中心的选址不同而有所区别，因此数据中心的选址除了要满足国标《数据中心设计规范》GB50174—2017 的选址要求以外，还要综合考虑土地成本和外市电投资对数据中心项目投资回收期的影响。项目咨询单位应重视外市电专业前期对场地的实地考察工作，从技术经济角度进行多方案比选，优中选优，最终形成一个业主认可的可研方案。

传统模式下数据中心项目各参建单位业务定位和利益诉求存在差别，很难对全局进行把控，实现项目利益最大化。建设单位人员通常技术力量不足，在项目具体决策方面较为依赖各技术服务单位。监理单位由于自身业务定位和自身欠缺设计专业知识，一般对设计意图领会不足，难以提出合理化建议以及对设计单位进行管理。一期工程的建筑设计单位很难兼顾考虑二期工程的通信工艺设计，通常不会主动对其衔接部分进行优化。在二期的设备安装工程中，影响一期建筑工程设计方案的技术因素较多，对技术专业性强要求高，是工程的建设难点。因此一期的建筑工程设计和二期的设备工艺设计工作具有一定的风险因素，极易造成工程返工、进度拖后以及工程质量不达标等问题。^[4]

3. 数据中心项目各阶段资质要求

数据中心建设项目一般分为一期的建筑工程及市电工程和二期的通信设备电源安装及调试两个阶段。从参与建设的技术服务单位来看，主要包括勘察设计单位、监理单位、施工承包单位这三家单位。从一期的建设内容来看，其主要包括数据中

心主楼建筑工程(含机电工程)及外市电配套工程，大型的数据中心有的还配属专用变电站。根据国家有关资质标准，设计单位应具备建筑行业(建筑工程)及以上设计资质，涉及专用变电站的还应具备电力行业(变电工程、送电工程)专业资质及以上设计资质；施工承包单位应具备适合对应规模的相应等级建筑工程承包资质、机电工程承包资质，涉及专用变电站的还应具备电力工程承包资质；监理单位应具备工程监理房屋建筑工程专业及以上资质、工程监理机电安装工程专业及以上资质和工程监理电力工程专业及以上资质。

数据中心建设项目二期的建设内容主要包括通信设备、通信电源的安装与调试，有的项目也把智能化系统划归为二期进行建设。二期的设计内容包含了通信设备工艺设计、通信管线设计，设计单位需具备电子通信广电行业(通信工程)设计资质；施工承包单位需具备相应等级的通信工程施工总承包资质；监理单位则需具备工程监理通信工程专业及以上资质。^[5]

4. 全过程工程咨询模式特点

全过程工程咨询是指对建设项目全生命周期提供组织、管理、经济和技术等各有关方面的工程咨询服务，包括项目的全过程工程项目管理以及投资咨询、勘察、设计、造价咨询、招标代理、监理、运行维护咨询以及 BIM 咨询等专业咨询服务全过程工程咨询服务可采用多种组织方式，由业主授权一家单位负责或牵头，为项目决策至运营持续提供局部或整体解决方案以及管理服务。全过程工程咨询服务贯穿决策、建设、运营这三个阶段的几乎所有环节，在项目实施过程中，全过程工程咨询服务能够突破不同阶段之间的阻碍，促使不同阶段、环节的建设内容相互穿插、关联。因此对于数据中心建设项目，全过程工程咨询可以为建设单位提供集投资决策、设计、造价、监理、项目管理等一体化的、全方位的、跨阶段的综合管理服务。全过程工程咨询服务商可以从项目决策阶段开始直到项目运营阶段，也可以从项目决策阶段开始到项目实施阶段，详见下图。

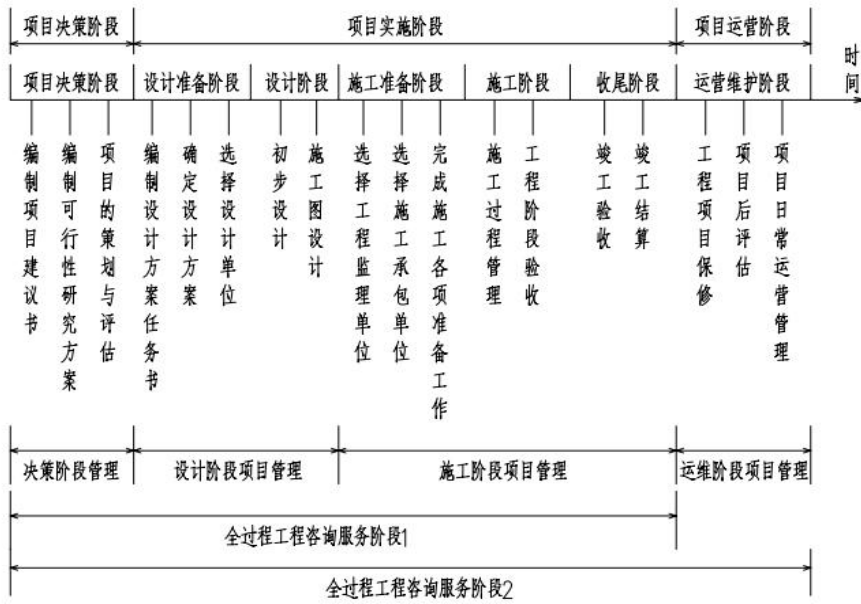


图 1 两种全过程工程咨询服务阶段的服务范围和内容

5. 全过程工程咨询模式对数据中心建设项目的优化

(1) 对参与服务单位的优化

由上述数据中心项目各阶段资质的要求可知，若采用传统模式，将有数量较多的服务单位共同参与完成项目，主要包括：前期咨询单位、勘察设计单位、工程监理单位、施工承包单位、招标代理单位、工程造价咨询单位等，这样不利于业主方的建设管理。鉴于目前市场设计及监理行业面临行业整

合，企业优胜劣汰加剧，大量公司由原来的专业化服务模式向多元化服务模式转型。采用全过程工程咨询管理模式可大幅简化业主方的合同关系，充分发挥全过程工程咨询服务单位的主导作用，提高工程项目建设水平。由此，采用全过程工程咨询模式与传统模式与项目各个阶段建设内容所需资质的要求汇总进行对比如下表所示：

不同服务模式下数据中心项目建设各阶段建设内容所需资质要求汇总对比表

建设项目阶段	建设内容	服务发包模式	所需设计资质	所需监理资质	所需施工资质
数据中心项目一期	建筑工程	传统模式	建筑行业（建筑工程）及以上设计资质	工程监理房屋建筑工程专业及以上资质	相应等级建筑工程承包资质
	机电工程			工程监理机电安装工程专业及以上资质	机电工程承包资质
	外市电工程（含专用变电站）			工程监理电力工程专业及以上资质	电力工程承包资质
数据中心项目二期	通信设备及电源工程		电子通信广电行业（通信工程）设计资质	工程监理通信工程专业及以上资质	相应等级的通信工程施工总承包资质

数据中心项目可研及一~二期	项目前期及以上所有内容	全过程工程咨询模式	同时具备电子通信广电行业(通信工程)甲级、建筑行业(建筑工程)、电力行业(变电工程、送电工程)设计资质或工程设计综合甲级资质	同时具备工程监理房屋建筑工程专业、机电安装工程专业、电力工程专业、通信工程专业及以上资质或工程监理综合资质	可根据情况选择具备相应多资质的施工总承包单位
			实施阶段设计与监理可组成联合体		

(2) 对项目建设投资的优化

数据中心项目建设投资中土地购置费用、建筑工程费以及外市电工程费均跟前期项目选址强相关，对于全过程工程咨询单位来说，基本上可将成本管控思维前置，从项目决策阶段贯穿到投产运营阶段。以建设一个 A 级数据中心为例，根据《数据中心设计规范》GB50174—2017 的相关规定：从火车站、飞机场到达数据中心的交通道路不应少于 2 条道路^[1]。其目的是为了防止当数据中心发生故障时，外界支援不会因为道路堵塞影响救援。另外对于供电电源要求应由双重电源供电(一主一备)，其目的是为了满足《数据中心设计规范》中 A 级机房对可靠电力供给的要求。^[2]此外《数据中心设计规范》中还规定新建 A 级数据中心的抗震设防类别不应低于乙类，而 YD 5054—2010《通信建筑抗震设防分类标准》又提出对于抗震设防类别为重点设防类(乙类)的建筑，应按高于本地区抗震设防烈度一度的要求加强其抗震措施。根据相关的项目经验数据，建筑工程的造价将因为抗震措施提高增加 20%~30%。由上述几点可知，对于数据中心的建设地区和位置，需要综合考虑项目地址的交通便利性、外市电引入便利性以及该区域的抗震设防烈度等因素。在项目前期可研阶段，全过程工程咨询单位可在多方案比选时统筹考虑，有利于在后期的设计

阶段进行限额设计以及投资控制。

6. 结语

数据中心建设项目不但在建设内容上复杂多样，并且在建设各个阶段参建服务单位所需资质要求也较为多样化，这为建设单位发承包的合同关系设定带来困难，也为业主方的项目投资控制带来一定的困难。因此在我国工程建设领域积极推行全过程工程咨询服务这一大背景下，在数据中心建设项目中，基于全过程工程咨询模式，探讨如何将项目建设各个阶段、各个环节贯穿起来，为建设单位提供全方位、跨阶段的综合管理服务显得尤为重要。此外，全过程工程咨询单位介入阶段越前置，对工程投资的影响越大。而作为咨询设计单位，也应加强人才队伍建设，使得设计企业相较于其他企业来说，在提供全过程工程咨询服务时具备更夯实的专业支撑。

参考文献

[1]中华人民共和国国家标准《数据中心设计规范》GB50174—2017
 [2]钟景华.《数据中心设计规范》GB50174—2017 解读—数据中心分级与选址.工程建设标准化.2017 年第 12 期