

基于数据中台的科技型产业园区能效管理平台研究与应用

崔一澜*, 孙成

(上海仪电人工智能创新院有限公司, 上海市 徐汇区 200232)

Research and Application on Energy Efficiency Management Platform of High-tech Industrial Parks Based on Data Middle Platform

CUI Yilan*, SUN Cheng

(Shanghai INESA Artificial Intelligence Innovation Institute Co., Ltd., Xuhui District, Shanghai 200232, China)

Abstract: The construction of the energy efficiency management platform for industrial parks mainly adopts the chimney architecture with high development cost, limited data sharing and repeated development of the same functional code. In view of the above problems, this paper discusses possible ways to support the fine management of energy use in industrial parks by introducing the concept of middle platform, so as to provide a unified collection and storage of multi-source and heterogeneous data, enable data sharing and provide front-end business with reusable and easily buildable model applications. Taking one typical high-tech industrial park located at Xuhui district, Shanghai as a case study, this paper designs and implements an energy efficiency management platform based on data middle platform, which supports the accumulation and reuse of general data, general feature operators, and general machine learning models, and can efficiently drive the rapid construction of various types of applications, including power abnormality detection and power demand forecasting etc. Such architecture provides a strong data service support for industrial parks to achieve the goal of energy saving, emission reduction and green development.

Keywords: high-tech industrial park; refined energy efficiency management; power demand forecasting; data middle platform

摘要: 当前产业园区能效管理平台建设过程主要采用“烟囱式”的体系架构, 存在建设成本高、重复开发、技术能力难以积累, 数据不共享、可扩展性差等弊端。为此, 以科技型产业园区低碳转型为切入点, 通过引入数据中台概念对园区多源异构大数据进行汇聚和存储, 为前端业务提供可共享复用、可快速构建的数据应用服务, 探索智慧园区能效精细

化管理方式。以上海市某科技型产业园区为实证, 基于数据中台构建能效管理平台, 依托平台积累的数据、特征算子及模型资产, 驱动用电异常检测、用电时序预测等数据应用的快速构建, 为园区实现节能减排、绿色发展提供坚实的数据服务支撑。

关键词: 科技型产业园区; 能效精细化管理; 能耗功率预测; 数据中台

0 引言

目前, 随着国家“科技驱动创新发展”战略的实施, 中国的科技型产业园区呈现较快增长趋势, 已逐渐成为国民经济不可或缺的组成部分。科技型产业园区聚集科技企业、科技人才和金融资本等生产要素, 随着其高速发展, 势必会成为能源消耗和碳排放的密集区域; 而园区面对大量种类繁多的科技企业, 也往往存在监管乏力的现象, 导致园区碳排放问题不断加剧^[1]。发展低碳经济已成为中国实现经济可持续发展的必由之路, 科技型产业园区是国家调整产业结构、优化产业升级的重要载体和手段, 推动和实现其“低碳排放”颇具紧迫性与必要性。

电力是科技型产业园区能源的核心。产业园区用电规模大, 能源智慧高效管理起着至关重要的作用。然而, 在低碳转型的大背景下, 园区想实现降低能耗和碳排放的目标却束手无策。传统园区能耗数据未有效互联互通, “数据孤岛”的存在制约了能耗数据的充分挖掘和利用。对于部分信息化和数字化已初见成效的园区, 随着智慧园区信息化建设的不断深入, 信息系统已全面覆盖园区, 积累了大量宝贵的数据资源, 并为园区各项活动的有效运营提供有力支撑。但

基金项目: 2019年度上海市人工智能创新发展专项支持项目(基于微软开源技术的人工智能资源调度管理平台)。

也存在一些突出的问题: ① 园区的业务系统多是以各自为中心的“烟囱式”建设模式, 跨业务场景间存在流程和数据壁垒; ② 旧有信息系统未建立对园区管理中数据应用需求的有效支撑, 数据分析应用数量多且散, 低水平重复、实用价值低的现象普遍存在; ③ 数据资产和分析能力未能得到有效积累并形成服务, 复用的程度较低, 新业务系统开发往往需要重新搭建, 导致构建系统的建设周期长、成本高、灵活性差^[2]; ④ 园区多维数据价值未被充分挖掘, 园区运营过程中产生的诸如园区楼宇入驻率、停车数据、园区活动等数据, 可以从另一些侧面反映园区内企业、人员等的用电情况, 合理利用园区多维数据以及园区相关外部数据进行融合分析, 将对园区运营的降本增效形成有效的增值服务。

近年, 对于园区能源管控以及低碳发展的研究已经成为众多学者关注的热点, 国外开展低碳产业园区以及园区节能监管体系的研究较早, 而中国则较晚, 其研究领域主要集中在产业园区低碳发展策略和路径研究^[3-4]、产业园区能耗及碳排放的测度与评估^[5-6]、园区数字能源平台的构建^[7]、分项能耗计算和能耗数据分析等方面。总体来看, 中国仍处于积极探索低碳产业园区建设阶段, 已有研究以概念设计、战略规划居多, 方案的指导性和可操作性仍显不足^[3]。与此同时, 在物联网、云计算、大数据和人工智能等新一代信息技术支撑下, 越来越多的研究开始关注园区智慧能源管理平台设计^[3, 7]。现有平台多集中于数据汇聚、存储、处理和描述性分析能力构建, 未建立数据打通融合以及共享能力, 平台可扩展性较差。此外, 现有研究多集中突出表现传统重工业减排重要性, 少有对碳排放量已进入绝对量排放较少的科技型产业园区进行针对性分析^[5]。随着科技型产业园区低碳转型需求的不断明确和多样化, 为高效响应和支撑未来持续增长的多样化数据挖掘和数据分析需求, 亟需实现数据资产化、分析能力服务化。

阿里巴巴于2015年在IT行业首次启动中台战略, 其核心是建立企业核心资源和共性技术的可重复使用能力, 避免重复构建、提高开发效率。数据中台是融合多条线业务系统数据的信息共享和服务发布平台^[8]。数据中台从后台业务系统采集数据, 进行数据汇聚与存储、数据管理与治理、数据处理与计算、数据服务和应用, 从根本上打破数据生产、存储、分析、服务、流通的技术壁垒, 形成数据共享和服务发布^[9]。数据中台使得企业可以针对特定问题, 快速匹配所需

的能力及资源, 进而更敏捷响应快速变化的业务场景^[8, 10]。目前, 数据中台的思想已被众多互联网公司所采纳, 指导他们搭建企业级数据平台, 以沉淀企业级共享数据、模型和服务^[11]; 同时, 部分信息化程度较高的传统领域也开始探索数据中台的应用^[2]。数据中台上述技术特点有望解决园区“数据孤岛”及“烟囱式”应用重复开发问题。此外, 借助大数据分析和人工智能算法, 挖掘多源异构数据价值, 复用特征算子和分析模型, 可提高算法迭代更新速度, 从而推动创新应用系统的快速构建。

综上, 本文以目前科技型产业园区电力能效管理现状及挑战为立足点, 结合物联网、大数据、中台技术、人工智能等先进信息通信技术, 探索园区用能的精细化、智能化管理方式, 并开展实证研究。基于数据中台构建智慧园区能效管理平台, 形成可扩展的服务共享体系, 利用中台积累的数据资产进行能耗评估和预测, 驱动包括用电异常检测、用电时序预测等各种典型应用的快速构建, 有望大幅提升园区对区域内能源的计量、监测、精细化管理与调控水平, 并可为中国包括工业园区、校园和商业住宅小区等在内的各类园区智慧能效管理提供借鉴, 实现智慧、绿色的发展目标。

1 数据中台关键技术

1.1 多源异构大数据的汇聚和存储

数据中台首先要具备强大的数据汇聚与存储能力, 以完成原始数据的积累。“数据孤岛”产生的核心原因在于, 传统信息系统通常是围绕特定的应用场景或业务设计开发, 数据属于独立运行的组织: 在物理上, 数据是独立存储和维护的; 在逻辑上, 数据是松散没有连接的。数据中台本质上是一种自下而上的数据孤岛解决方案, 其优先考虑数据的全局特征, 制定数据规范、定义数据标准; 基于统一的数据认知, 采集内部各业务系统数据, 交换获取外部来源数据, 最终以操作型数据存储、数据仓库的形式存储数据, 从而实现上层应用、共享的数据服务与底层数据源之间的集成^[10]。

数据汇聚需支持多种类型数据的共同读取和分析, 包括关系型数据库、非关系型数据库, 以及来自文件系统等对象型数据。数据仓库技术(extract-transform-load, ETL)是系统间数据汇聚的主要手段, 传统的ETL过程是自顶而下、需求引发的, 过程较为

繁琐,且数据在转换处理后可能会出现缺失。因此,对于难以在数据汇聚初期定义好数据格式的数据来源,可采取“先汇聚、后治理”的方式,也即当前逐渐趋于流行的ELT模式。经汇聚后,数据通常具备结构化、半结构化、非结构化等多样的存储特征,但在逻辑上仍是分散、孤立、没有统一标准的。

数据存储在实践中多基于数据类型、使用场景等划分成源数据区、共享数据区、分析数据区三类存储区域。其中,源数据区存储从内部业务系统和外部数据源获取的源数据,并满足归档、保管、查询需求;共享数据区存储经过统一治理、融合关联后的规范资产型数据,面向共享发布;分析数据区存储经过数据挖掘、统计分析的数据结果,如用于建模的主题型数据集、特征算子库和模型库。

1.2 多源异构大数据的管理和共享

中台内的数据是稳定、标准化、价值密度高、面向共享服务的资产型数据,将资产型数据封装成数据服务,以接口的方式开放给前台应用系统,从而释放数据资产价值,实现数据能力复用。数据中台的这个优势得益于中台数据管理治理和共享服务能力的建设。

1.2.1 数据管理治理

元数据作为顶层抽象控制层对中台内的数据进行全方位的管控,是数据共享服务的基础。元数据管理通过数据分类与编码、数据字典等形成数据的一致理解 and 统一坐标参照,使数据更方便检索、使用或管理,同时也统一了数据交换、存储和应用的口径^[12]。在此基础上,通过梳理核心数据资源,建立数据标准和规范,将原始分散、重复、低质量的数据治理成格式、类型、单位统一,编码和逻辑一致的高质量数据集,通过数据抽取、清洗、转换、融合、加载等流程,最终形成标准统一、逻辑一致的定制化数据资产^[13]。

此外,需基于数据全生命周期开展数据质量、数据安全的管理。数据质量通过数据清洗、数据验证等技术手段控制数据的使用、存储和传输质量,保证数据的正确性、完整性、真实性、时效性和可靠性。数据安全通常结合信息安全的技术手段,如数据脱敏、数据加密、权限控制等,保障数据资产在使用和共享交换过程中的安全。

1.2.2 数据共享服务

通过构建标准的数据服务体系,统一共享口径,可以使沉淀在中台上的数据集和数据服务流动并被使用,体现了数据中台以服务的方式赋能业务系统的能

力。数据中台和应用系统间通过共享服务形成数据闭环,从而促进应用系统和数据中台数据资产的不断迭代升级。

数据共享服务可以封装开放主题式数据集进行数据交换,也可以封装数据计算逻辑形成数据分析服务,如数据查询、在线数据分析、算法模型等,帮助快速构建可视化的数据应用。数据共享通过应用程序接口服务、基于权限的数据库接口和文件等技术手段实现安全可控的访问。此外,必须明确定义和规范数据交换标准,包括数据交换内容、格式、传输方式,以及各类实体间数据接口的标准化等方面。

1.3 数据应用系统的快速构建

数据中台通过计算和处理能力建设数据资产,并通过算法模型实现数据价值的深度挖掘。数据共享服务是数据中台能力的出口,通过直接使用中台沉淀的数据能力,实现数据应用系统的快速构建。

具体而言,数据中台通过离线的计算能力,整合抽取共性的数据集,并预处理得到通用的特征算子库。利用平台的建模分析能力进行模型算法迭代和复用。模型的训练服务通常基于工作流引擎,逐步完成下载文件、数据预处理、特征工程、模型训练和模型迭代等过程。新的数据分析应用可以使用通用的数据集和特征算子,减少数据预处理过程中的重复编码,并直接调用可复用的模型算法进行迁移学习,减少模型训练的成本,实现快速灵活构建新的数据应用系统^[14]。

实际中,需根据数据应用场景特点选择构建数据中台的计算和处理能力,包括离线计算、在线计算、在线查询以及统计分析、数据挖掘及机器学习等建模分析能力。离线计算适用于吞吐量大、延时高的数据处理场景,如离线数仓加工、大规模数据清洗。在线计算适用于实效性要求较高的实时或流式处理场景,如实时监控告警。在线查询主要指对于数据的条件筛选过滤查询过程,如数据库检索场景等。为满足智能应用需求,需构建数据建模分析能力,包括算法开发及模型训练、评估、预测等。

2 基于数据中台的智慧园区能效管理平台

智慧园区能效管理平台基于中台的思想抽取前台共性需求,整合后台可复用功能以及标准化数据资源,并共享数据服务来驱动包括数据监测、能耗预测、异常检测、画像分析等数据分析及人工智能应用

的快速构建。通过沉淀共性的数据资源、通用的特征算子、可复用的模型服务, 避免重复开发功能性代码和模型, 提高模型利用率, 缩短模型迭代周期和开发成本, 快速响应园区管理需求。园区开发新的管理应用可以像搭积木一样复用中台的能力, 也可重复多次算法迭代升级分析模型。

本研究选取某科技型产业园区开展基于数据中台的智慧园区能效管理平台实证研究。该园区基本具备数据中台行业应用的共性条件: ①具有一定规模的信息化建设基础; ②预期会增量的数据分析应用; ③业务之间存在可复用模块。

2.1 产业园区基本情况

该园区位于上海市徐汇区, 总面积0.21 km², 目前已入驻300多家企业, 其中, 科技型企业占比达70%以上, 年总耗电量6×10⁶ kWh。园区在信息化建设过程中, 已对各用电区域的电表进行了远传化, 形成了用电基础数据的采集能力。同时, 园区运营过程中也产生了其他诸多数据, 如电梯、门禁、照明、水表、楼宇、企业、园区活动数据等, 传感器设备达1332个。目前, 园区逐渐对上述设施建立了数据采集系统, 并且构建了物业运营管理平台, 对各系统的数据进行汇集和管理。

2.2 园区能效管理现状

园区的信息化和数字化已经初见成效, 但是, “烟囱式应用”的问题仍然存在。物联网设备采集系统和园区综合管理信息系统大量建设, 且相互之间存在重叠和冗余。数据存储在不同类型、版本的数据存储组件上, 存在逻辑、规范不一致的情况; 此外, 数据与旧有系统紧耦合, 读取和使用过程受限。针对不断增加的数据分析需求, 开发工作需要从头开始, 效率上受限于旧有系统之间的契合程度, 分析和挖掘很难建立在海量数据上, 未体现数据资产的规模价值。

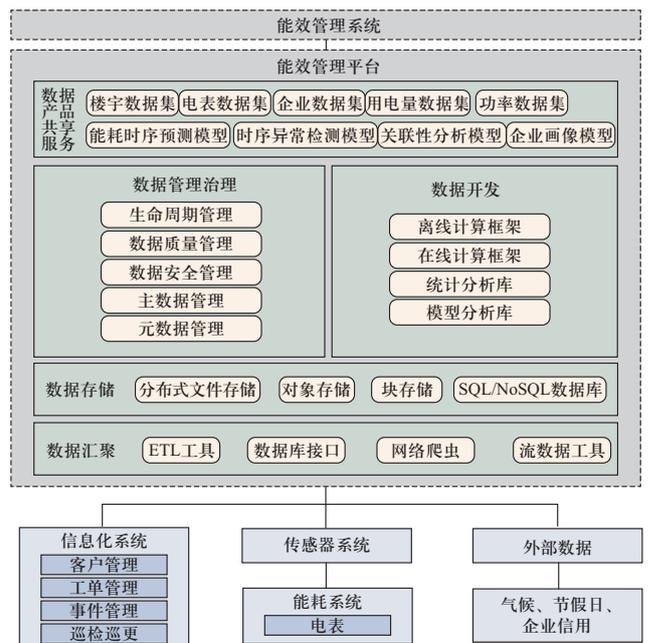
园区在数据打通融合以及数据价值分析挖掘上仍存在较大改进空间, 尤其对于能耗数据, 在收集和存储之上并没有形成能效管理层次的数据价值。企业、楼宇级别的能耗效能、运行情况监管, 是园区精细化管理的重要一环, 可以有效地辅助园区运营降本增效。一方面, 可以通过园区用电历史数据进行园区用电量和契约用电负荷的建模和预测, 预测契约用电负荷的峰值时间, 配合储能放电降低用电成本, 同时合理进行园区用电负载管理和风险管理。另一方面, 通

过对园区不同区域、系统、单位的用电量分析, 可以形成园区在能效方面的成本明晰化、收益可控化、运营精细化, 提升园区运营效率。

2.3 能效管理平台建设

在园区已有物联网设备采集系统和园区管理信息系统的基础上, 研发建设智慧园区能效管理平台。平台汇总园区能耗、停车、外部气候、租户企业信息等多源数据, 利用大数据组件的分布式处理计算能力, 对数据进行统一、整合、关联, 加工产生资产型数据, 并以服务的形式共享开放, 实现主要用能实体的能效在线监测、统计分析及预测调控, 为园区精细化能源监管与分析提供数据支持。

智慧园区能效管理平台的整体架构如图1所示。该平台基于云平台构建底层技术框架, 根据数据采集、存储、计算和服务的需求选择组件。数据中台能力包括数据汇聚、数据存储、数据管理治理、数据开发和数据共享服务, 最终实现前台能效管理智能化应用, 进行数据可视化和决策支持。



1) 数据汇聚。该层使用数据库接口、网络爬虫和Flume等流式数据采集工具汇聚多源多维数据, 包括能耗数据、物业租户管理数据、物业设备管理数据等内部数据, 以及从天气预报、企查查等网站获取的天

气、企业信用等外部数据。根据各业务系统提供的访问方式进行数据汇聚,其中:关系型数据库存储的业务数据采用ETL方式获取;对象型数据通过接口访问服务;外部网络数据则使用商用的爬虫工具获取。

2) 数据存储。该层采用基于HDFS分布式存储系统和MongoDB的非关系型数据库的数据存储方案。该方案支持文件管理和文件存储的解耦,可以充分发挥不同分布式组件的优势,从而更好地适配多源异构的数据文件,并且易于扩展。数据存储层包含三种类型的数据:一是源数据,存储园区管理系统、设备采集系统汇聚的原始数据;二是共享数据,存储经过统一治理、融合关联后的规范资产型数据,面向共享发布;三是用于建模分析的主题型数据集、特征算子库和模型。

3) 数据管理治理。该层依据园区管理业务对数据的需求,统筹管理园区各层级的主数据。园区数据的来源均为物联网和大型信息系统,各类数据的统筹管理需以元数据管理为基础。首先,对汇聚后的数据进行转换,按照层级、关联、分类多级匹配等方式,完成一体化的规范化整合,消除不同来源数据间的关联障碍。其次,建立主题型数据集和共享数据的标准数据集,形成三大类数据,包括源数据、融合关联后的统一规范数据、统计分析类数据。最后,建设元数据管理层,根据数据分析、数据服务、数据应用的需求扩展优化数据,形成新的数据资产,并纳入元数据管理范畴。

4) 数据开发。该层完成数据关联处理和建模分析任务,具体包括三类功能需求。一是统一管理计算资源用于数据处理和模型训练,以解决应用系统分散面临的资源不足或资源浪费问题。二是实现数据资产化并为数据应用系统的搭建提供数据特征处理和模型训练能力。通过大数据计算框架对数据进行进一步的综合统计分析,形成的分析结果同样汇入元数据管理范畴,进行中台资产数据的补充、数据价值的提升。三是采用离线训练和在线分析相结合的方式实现数据处理。针对批处理场景,使用基于Spark分布式计算框架,对批数据进行离线的训练建模;针对流处理场景,如在线接入流数据实现每日更新分析结果,使用基于Spark Streaming的流式计算框架进行数据的清洗、转换并加载入分布式存储平台。

5) 数据共享服务。该层实现数据资产共享交换以及模型分析能力复用,为前台数据应用系统提供共享的在线模型服务,包括能耗时序预测、时序异常检

测、关联性分析、企业画像标签等,结合可视化展示界面快速构建数据应用系统。

2.4 平台应用效果评估

本研究搭建的智慧园区能效管理平台于2020年11月起在案例园区部分区域进行试点建设。基于云+端的架构构建智慧园区数据中台对数据资产进行治理。利用中台的元数据、主数据、数据生命周期管理等能力进行规范化管理的同时,利用大数据分析组件的计算能力产生更多支持高层业务分析场景的数据。此外,利用中台的共享能力以服务形式发布数据,使数据在数据中台和应用系统之间形成闭环,实现了数据资源的有效利用和价值迭代增长。在上述基础上,基于统一标准的数据模型及共享服务接口,利用数据中台对设备采集数据、园区业务数据等的交互查询、处理计算能力,设计多模态能耗数据应用。

截至2021年8月,该平台累计采集能耗数据、物业用户管理数据、物业设备管理数据、气候与企业信用等外部数据2.3亿条,提供能耗数据服务调用次数达万次。围绕用电精细化管理、削峰填谷、企业租户管理等方面,该平台能够按照能耗感知、智能分析、异常检测及预测的步骤支撑园区能效管理决策。

3 能效管理智能化应用

能效管理智能化应用通过服务调用共享接口,基于中台沉淀的数据资源和模型资产,上线数据应用服务并提供统计分析、可视化展示、推理预测等能力,从而辅助园区运营决策。

本研究中,针对园区包括能耗、停车、园区管理活动在内的多源多维数据,进行融合分析、关联分析、模式分析。针对各类数据之间的关联关系,对不同层级用电实体(园区、楼宇、企业)的用电情况进行建模,提供用电异常检测和用电预测的数据服务,为园区提供用能精细化管理能力,并辅助园区运营管理工作降本增效。

研究中利用工作流引擎,根据工作流文件的描述配置数据预处理和模型训练的流程。通过基于权限的数据库接口服务获取中台沉淀的用电主题数据集,包括楼宇基本数据、楼宇电表关联数据、企业电表关联数据、楼宇及企业用电数据,以及加工好的如用电量同比环比、用电量众数、偏度峰度、用电量变化趋势等表征数据分布的特征算子。使用在公开数据集上经

过验证的异常检测、时序预测基准模型进行迁移学习, 在用电数据上快速迭代训练, 完成最终算法模型的生产, 并通过应用程序接口服务开放共享形成用电异常检测和用电预测的数据服务, 从而实现数据应用的快速搭建。

3.1 园区能耗异常检测

物联网系统采集的能耗数据大多以时间序列方式呈现, 通过合适的异常检测算法, 可以对采集设备的异常状态、用能主体的异常使用加以检测和标记, 从而起到监控告警的作用。目前常用的异常检测技术大致分为两大类: 基于统计分布的多维度数据异常检测与基于时间序列建模的单维度异常检测。

其中, 基于统计分布的多维度异常检测技术包括基于分类、基于近邻、基于聚类等。基于分类的异常检测技术严重依赖标记数据且难以识别新异常类型, 而基于聚类的异常检测技术在面对高维数据时无法给出有意义的类别。相较之下, K最邻近(K-nearest neighbor, KNN)算法使用数据点与其第K近的邻居间距离作为异常得分, 可以依据少量标记数据实现分类, 但其难以检测密度变化较大的数据集^[15]。HBOS(histogram-based outlier score)^[16]是一种经典的基于直方图的非参数异常检测方法, 其使用每个数据点相对周围数据点的概率密度大小作为异常得分, 可以高效处理大数据集。此外, 以孤立森林(isolation forest, IForest)^[17]为代表的集成异常检测技术在实践之中广受使用, 能有效处理高维数据和海量数据。以主成分分析(principal components analysis, PCA)为代表的谱异常检测技术通过将高维数据投影到更低维的空间中, 实现高维数据异常点的有效检测^[18]。

然而, 上述基于统计分布的异常检测技术难以完整的把握数据在时间维度上的变化特征和模式。Prophet^[19]是Facebook在2017年开源的基于时间序列建模的算法, 该算法基于趋势项、周期项对时间序列进行分解和拟合, 可以用于进行时序异常检测。使用该算法得到对于不同时间点的连续置信区间, 对于超出置信区间的所有时间点, 判定为时间序列分布的异常值。然而, 此类基于时间序列建模的异常检测技术往往缺少对多维度序列关联性和联合分布特点的把握。

综上, 本研究融合应用两类异常检测技术, 包括IForest、PCA、HBOS、KNN、Prophet等5种异常检测技术, 形成一个能够对多维度时间序列进行异常检测的算法模型(详见附录A)。基于上述算法, 研究

中针对园区、楼宇、企业的用电量和契约用电负荷(maximum demand, MD)进行无监督异常检测, 对能耗数据在历史变化中的统计分布异常(见图2)以及时间变化异常(见图3)进行综合提取, 经过参数调优, 实现了AUC(area under curve)值达到0.9的无监督异常检测效果。

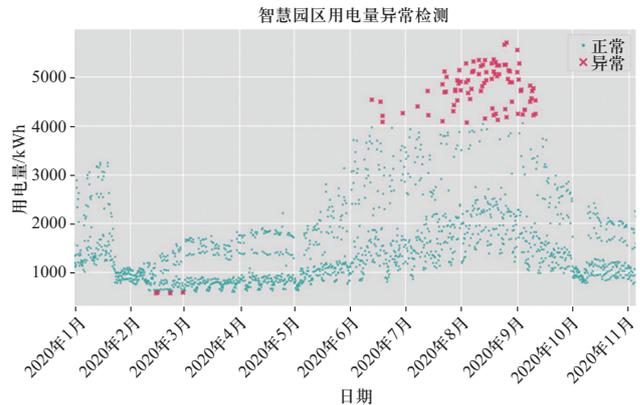


图2 用电量统计分布异常检测

Fig. 2 Abnormal detection of power consumption in statistical distribution

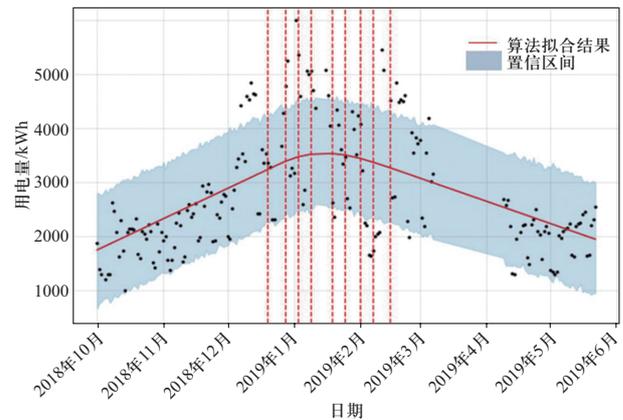


图3 用电量时序异常检测

Fig. 3 Abnormal detection of power consumption in timing distribution

上述用电量异常检测结果可以帮助园区展开高效的设备管理工作, 自动识别异常情况并触发报警和人工巡检, 极大缩短远传电表等设备的故障修复时间。此外, 异常检测结果可用于辅助预测性维护, 即通过大量数据总结出相关异常的特点和模型, 在此基础上配合多维度时序数据预测算法, 形成未来一段时间是否会出现特定异常情况的预判。进一步的, 结合图表等可视化展示方式可将算法识别出的异常结果反馈给行业专家进行修正, 借助专家校验进一步优化异常检测算法效果。

3.2 园区能耗功率预测

在当前两部制电价政策中，园区用电成本包括基于峰值功率测算的基础电费和基于实际用电量的分时电费。如何精确的预测和控制峰值功率，成为影响园区用电基础电费成本的关键环节。通过智慧园区能耗功率精细化预测的优化系统，能够实现园区每日能耗高峰功率及其发生时段的精准预测，从而指导园区精细控制能耗和储能设备的使用，达到降本增效的作用。

相比于传统机器学习模型，长短期记忆网络(long short-term memory, LSTM)^[20]由于其具有天然的序列建模优势，可以自动学习序列隐含信息以及复杂高阶特征，减少了对特征工程的依赖。此外，对LSTM进行模型集成可以有效增加模型的准确性和鲁棒性，从而提高预测效果。

因而，本研究使用堆叠LSTM模型，结合特征工程自动化提取工具tsfresh^①构建时序预测算法（详见附录B）。使用包括气候、节假日等属性的多维度时间序列数据，进行多维度联合建模，预测园区、楼宇及企业的用电量和契约用电负荷MD。

研究发现，用电数据与气候数据、节假日时间数据呈现强相关，融合气候、节假日等多维度数据，可有效提升园区能耗功率预测准确度（见图4）。经过参数调优，最终园区用电量预测模型平均绝对百分比误差（mean absolute percentage error, MAPE）评价指标小于4%（见图5），契约用电负荷预测模型MAPE平均指标小于12%（见图6），达到园区管理需求。

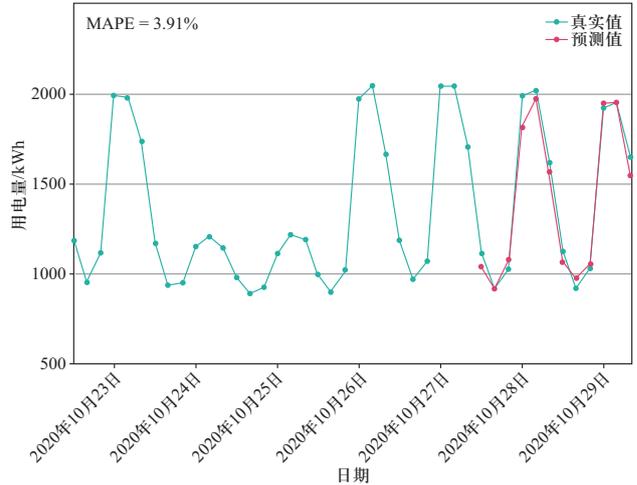


图5 用电量预测

Fig. 5 Power consumption forecast

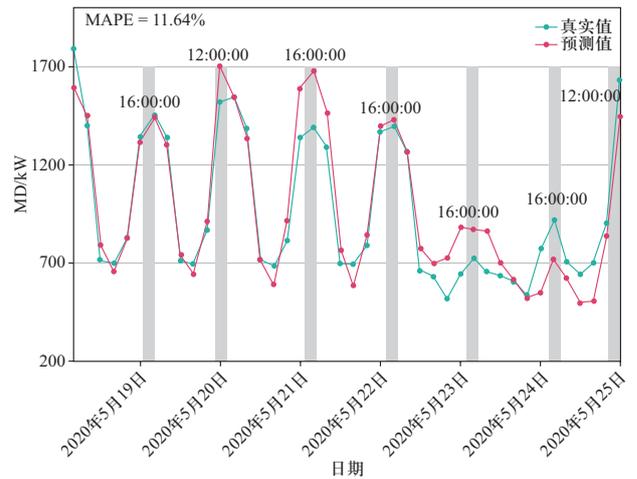


图6 契约用电负荷预测

Fig. 6 Power maximum demand forecast

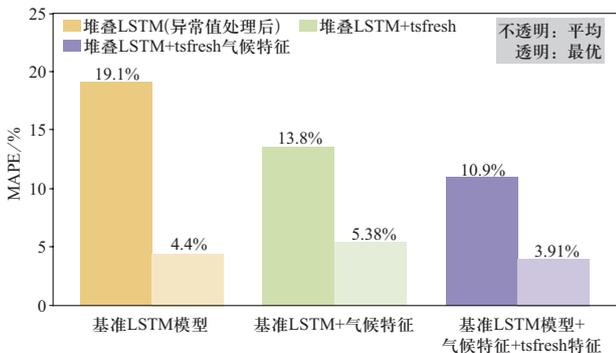


图4 不同模型多点预测用电量结果

Fig. 4 Multi-point prediction of power consumption results with different models

① tsfresh: <https://tsfresh.readthedocs.io/en/latest/>.

4 结论及展望

本研究立足科技型产业园区“低碳转型”的重要意义和迫切需求，从园区能效精细化管理中面临的多源异构数据汇聚、多源异构数据管理治理、多源异构数据价值挖掘等现实挑战出发，提出了基于数据中台的智慧园区能效管理平台架构设计。数据中台的技术特点可以有效解决智慧园区能效管理的问题和需求。首先，使用分布式存储技术存储采集和交换的多源异构数据，解决传统园区能耗数据隔离的问题；其次，利用中台的数据管理治理能力沉淀数据资产、特征算子和模型资产，解决数据处理、模型开发等代码重复开发、复用度低、周期长的问题；最后，借助大数据分析和人工智能算法构建数据服务，快速实现面向园

区管理应用场景的决策和智能创新。研究选取某科技型产业园区开展实证研究, 通过数据中台支撑建立能效精细化管理应用, 形成园区、楼宇、企业级别用电管理。通过平台沉淀的数据资产、特征算子及算法模型, 快速构建园区不同层级的用电实体的用电模式分析应用, 通过异常检测、能耗预测等人工智能应用完成设备故障预警、园区用电功率优化使用、入驻租户管理等智慧园区综合管理业务, 配合园区储能设备进行削峰填谷, 最终实现园区能源运营的降本增效。

在下一步的研究中, 一方面将在现有的用电模型基础上融合更多维度数据, 诸如园区人流量数据、经济活动数据等, 以提升能耗异常检测及功率预测算法的效果和性能。并探索数据驱动与专家经验驱动结合的方式, 即专家对数据洞察结果进行评判, 给出部分结果是否正确的标记, 进而利用基于少量标记数据的半监督算法优化迭代异常检测模型。另一方面, 将在辅助决策自动化方面进行更多的理论研究及开发, 如配合储能装置的综合能源管理调度算法的研发, 在实践中不断丰富和发展面向园区能效管理的中台技术, 探索出智慧园区“数据汇聚—融合分析—辅助运营管理”的有效模式, 为园区低碳转型提供强有力的支撑。

参考文献

- [1] 史竹琴, 朱先奇, 苏妮娜. 科技型产业园区低碳转型的动态演化研究: 基于利益相关者视角[J]. 软科学, 2020, 34(6): 64-71.
SHI Zhuqin, ZHU Xianqi, SU Nina. Dynamic evolution of low carbon transition in high-technology industrial parks—based on stakeholder perspective[J]. Soft Science, 2020, 34(6): 64-71(in Chinese).
- [2] LYU G X, LIU P, LU Y M, et al. A data middle platform architecture based on microservice serving power grid business[C]//2021 8th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2021 7th IEEE International Conference on Edge Computing and Scalable Cloud. Washington, DC, USA. IEEE: 219-224.
- [3] 颜培霞. 中国低碳产业园区的研究进展与未来展望[J]. 生态经济, 2019, 35(5): 26-30.
YAN Peixia. The research progress and future development of low-carbon industrial park in China[J]. Ecological Economy, 2019, 35(5): 26-30(in Chinese).
- [4] 刘磊, 杨文海. 中国碳中和产业园区建设探析[J]. 城乡建设, 2021(9): 38-40.
- [5] 徐涵, 温日琨, 金力豪. 产业园区碳排放核算、回归与趋势: 以杭州青山湖科技城为实证[J]. 绿色科技, 2019(20): 210-215.
XU Han, WEN Rikun, JIN Lihao. Carbon emission accounting, regression and trend in industrial park: empirical study from Qingshan Lake high-tech city in Hangzhou[J]. Journal of Green Science and Technology, 2019(20): 210-215(in Chinese).
- [6] 杨培志, 韩春洋. 基于LEAP模型的长沙市某产业园区长期能源需求量与碳排放分析[C]//中国环境科学学会2019年科学技术年会——环境工程技术创新与应用分论坛论文集(一). 西安, 2019: 306-310.
- [7] 张鑫龙, 陈启新, 许丽华, 等. 智慧能源管理系统在产业园区的应用[J]. 建筑科技, 2020, 4(2): 89-91.
ZHANG Xinlong, CHEN Qixin, XU Lihua, et al. Application of intelligent energy management system in industrial park[J]. Building Technology, 2020, 4(2): 89-91(in Chinese).
- [8] 赵增涛, 罗勇, 梁成辉. 电力企业中台云化构建及大数据分析研究[J]. 水电与抽水蓄能, 2020, 6(3): 50-55.
ZHAO Zengtao, LUO Yong, LIANG Chenghui. Research on middle platform construction and large data analysis of power enterprises based on Aliyun[J]. Hydropower and Pumped Storage, 2020, 6(3): 50-55(in Chinese).
- [9] 崔恒志, 王翀, 吴健. 基于数据中台的数据资产管理体系[J]. 计算机系统应用, 2021, 30(3): 33-42.
CUI Hengzhi, WANG Chong, WU Jian. Data asset management system based on data center[J]. Computer Systems & Applications, 2021, 30(3): 33-42(in Chinese).
- [10] ZHANG C Z, HOU L. Data middle platform construction: the strategy and practice of National Bureau of Statistics of China[J]. Statistical Journal of the IAOS, 2020, 36(4): 979-986.
- [11] LU Z H, WANG N N, WU J, et al. IoTDeM: an IoT Big Data-oriented MapReduce performance prediction extended model in multiple edge clouds[J]. Journal of Parallel and Distributed Computing, 2018, 118: 316-327.
- [12] 江疆, 黄剑文, 杨秋勇. 基于元数据的数据资源管理机制的研究[J]. 国网技术学院学报, 2016, 19(2): 33-36.
JIANG Jiang, HUANG Jianwen, YANG Qiuyong. The study of data resource management mechanism based on metadata[J]. Journal of State Grid Technology College, 2016, 19(2): 33-36(in Chinese).
- [13] QIAN H S, XU L M, ZHOU C, et al. Research on construction and key technology of water conservancy data middle platform[J]. IOP Conference Series: Earth and Environmental Science, 2021, 768(1): 012112.
- [14] ALAHAKOON D, YU X H. Smart electricity meter data intelligence for future energy systems: a survey[J]. IEEE Transactions on Industrial Informatics, 2016, 12(1): 425-436.
- [15] LIAO Y H, VEMURI V R. Use of K-nearest neighbor classifier for intrusion detection[J]. Computers & Security, 2002, 21(5): 439-448.
- [16] GOLDSTEIN M, DENGEL A. Histogram-based outlier score (HBOS): a fast unsupervised anomaly detection algorithm

[J/OL]. Semantic Scholar, 2012[2021-10-14].[https://www.semanticscholar.org/paper/Histogram-based-Outlier-Score-\(HBOS\)%3A-A-fast-Goldstein-Dengel/405bde43709582b0026c0fd6f0afe2c3c57f792e](https://www.semanticscholar.org/paper/Histogram-based-Outlier-Score-(HBOS)%3A-A-fast-Goldstein-Dengel/405bde43709582b0026c0fd6f0afe2c3c57f792e).

- [17] LIU F T, TING K M, ZHOU Z H. Isolation-based anomaly detection[J]. ACM Transactions on Knowledge Discovery from Data, 2012, 6(1): 1-39.
- [18] RINGBERG H, SOULE A, REXFORD J, et al. Sensitivity of PCA for traffic anomaly detection[C]//Proceedings of the 2007

ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems - SIGMETRICS '07. June 12-16, 2007. San Diego, California, USA. New York: ACM Press, 2007.

- [19] TAYLOR S J, LETHAM B. Forecasting at scale[J]. The American Statistician, 2018, 72(1): 37-45.
- [20] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

附录A 园区能耗异常检测算法

园区能耗异常检测算法流程如图A1所示。首先，针对能耗数据的统计分布进行异常建模。统计分布异常的基本假设是，正常的数据点会出现在随机模型发生概率高的区域，而如果数据点出现在随机模型发生概率低的区域则会被认为是异常点。基于这个理念，研究中融合使用孤立森林（isolation forest, IForest）、主成分分析（principal components analysis, PCA）、HBOS（histogram-based outlier score）、K最邻近（K-nearest neighbor, KNN）4种异常检测技术对能耗数据进行统计分布的异常检测；对每个时间点所有的异常打分进行叠加，超过阈值的点将被判定为异常点。其次，使用时间序列模型Prophet在时间维度进行异常检测。最终，叠加统计异常与时序异常的综合结果进行能耗数据异常检测。

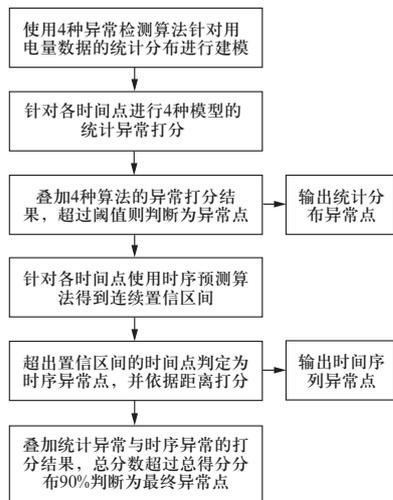


图 A1 异常检测算法流程
Fig. A1 Pipeline of anomaly detection algorithm

附录B 园区能耗功率预测算法

园区能耗功率预测算法流程如图B1所示。首先，基于多源多维数据的关联分析结果进行特征初步筛选；其次，结合特征工程自动化工具tsfresh完成特征选取；最后，使用LSTM基准模型不断迭代，进行每日用电量、峰值契约用电负荷MD及其出现峰值时段的预测。

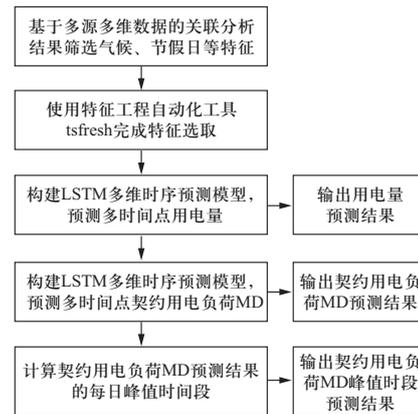


图 B1 时序预测算法流程
Fig. B1 Pipeline of time series forecast algorithm

收稿日期：2021-11-29；修回日期：2022-03-18。



崔一澜

作者简介：

崔一澜（1992），女，博士，研究方向为复杂性科学与环境经济复合系统模拟应用研究、机器学习、公共数据共享等。通信作者，E-mail: cuiyl@shaic.com。

孙成（1992），女，硕士，研究方向为数据治理与中台技术，E-mail: 18605669908@163.com。

（责任编辑 张鹏）