

基于 LLM-PINN 的建筑楼宇空调聚合 灵活调控研究

陈胜^{1*}, 赵宇¹, 惠红勋², 苏慧玲³, 卫志农¹

- (1. 河海大学电气与动力工程学院, 江苏省 南京市 211100;
2. 智慧城市物联网全国重点实验室 (澳门大学电机及电脑工程系), 澳门 氹仔 999078;
3. 国网江苏省电力有限公司营销服务中心, 江苏省 南京市 210000)

Flexible Aggregated Control of Building Air-Conditioning Loads Based on LLM-PINN

CHEN Sheng^{1*}, ZHAO Yu¹, HUI Hongxun², SU Huiling³, WEI Zhinong¹

- (1. College of Electrical and Power Engineering, Hohai University, Nanjing 211100, Jiangsu Province, China;
2. State Key Laboratory of Internet of Things for Smart City (Department of Electrical and Computer Engineering, University of Macau), Macao 999078, China;
3. Marketing Service Center, State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210000, Jiangsu Province, China)

ABSTRACT: To improve the response accuracy of aggregated control for massive distributed building air-conditioning loads under multiple operating scenarios, this paper proposes a flexible aggregate control framework integrating multimodal large language models (LLM) and physics-informed neural networks (PINN). First, the inertial electrical response of air-conditioning units and the multi-node thermal dynamics of rooms are decoupled, and a hierarchical PINN simulation model is developed to derive room-level dynamic power safety boundaries and support continuous thermal safety verification during the response process. Second, a multimodal LLM is used to parse operational semantics, functional topology, and visual scene information, converting unstructured rules and occupancy states into room-level structured constraint parameters. These parameters drive feature-wise linear modulation Transformer (FiLM-Transformer)-based candidate allocation and quadratic programming (QP)-based feasible-domain projection. A semantic-constraint marginal value assessment method is further developed based on QP dual information, enabling semantics-aware fast feasible allocation and value quantification of constraint relaxation. Case studies show

that the proposed method characterizes room-level regulation boundary variations caused by operational semantics and visual states, achieves second-level feasible allocation at the building-cluster scale while satisfying power balance, room boundary, and dynamic thermal safety constraints, and provides high-accuracy dynamic response evaluation. The dual-marginal-value-based constraint relaxation strategy covers capacity deficits with reduced comfort compensation costs for high-sensitivity constraints.

摘要:为提高多运行场景下海量分散建筑空调负荷聚合调控的响应精度,提出一种融合多模态大语言模型(LLM)与物理信息神经网络(PINN)的楼宇空调负荷聚合灵活调控框架。首先,将空调电功率惯性响应与房间多节点热动态过程解耦,构建电热分层 PINN 仿真模型,形成房间级动态功率安全边界,并用于响应过程的连续热安全校核。其次,利用多模态 LLM 解析运行语义、功能拓扑和视觉场景,将非结构化规则与占用状态转化为房间级结构化约束参数,驱动特征调制变换器(FiLM-Transformer)候选分配与二次规划(QP)可行域投影,并基于 QP 对偶信息提出语义约束边际价值评估方法,实现语义感知下的快速可行分配与约束放宽价值量化。算例表明,所提方法能够刻画运行语义与视觉状态引起的房间级调节边界变化,在满足功率平衡、房间边界和动态热安全约束的同时,实现楼宇群秒级可行分配与动态响应的高精度评估;考虑对偶边际价值的约束放宽策略可在覆盖容量缺额的同时降低高敏感约束的舒适补偿成本。

基金项目: 国家自然科学基金资助项目(52377091); 江苏省基础研究计划项目(BK20250186)

Project supported by the National Natural Science Foundation of China (Grant No. 52377091) and the Basic Research Plan of Jiangsu Province (Grant No. BK20250186)

关键词: 楼宇空调负荷; 多模态大语言模型; 物理信息神经网络; 两阶段分配; 聚合灵活调控

0 引言

随着高比例可再生能源持续接入新型电力系统, 净负荷波动与局部功率失衡在短时间尺度上日益突出[1-2], 需求侧柔性资源成为支撑电网短时平衡的重要补充[3-4]。虚拟电厂通过对规模化灵活资源进行动态聚合与协同调控, 为分散资源参与系统运行提供了重要组织形态[5]。商业楼宇空调负荷容量规模大、热惯性可利用, 是具备虚拟储能特性的代表性需求侧资源[6-9]。其调节过程依托室内热舒适裕度与围护结构蓄放热, 无需刚性停产或额外储能配置, 具备较高的工程可部署性[10-11]。

然而, 楼宇空调负荷的实际可调能力受房间用途、人员占用、局部热扰动及运行规则等多源因素共同制约[12-14]。不同房间在同一时段内对应不同的舒适要求、业务优先级与控制权限, 房间级可调能力呈现显著的时变性与空间异质性[15-17]。在海量分散房间聚合场景下, 上述差异会放大聚合容量边界的不确定性, 使总功率请求难以精确分解, 并增加逐房间建模、通信与约束维护成本; 非结构化用户需求和场景状态难以直接进入优化过程, 也会影响聚合调控结果的可行性与响应精度。

在热动力学建模方面, 文献[18]基于热阻热容网络建立了多节点热回路组装方法, 文献[19]采用等效热参数模型构建楼宇综合能源系统的预测控制框架。在调节能力评估方面, 文献[20]提出建筑空调系统的实时灵活性量化方法, 文献[21]开展了计及舒适度的空调负荷调控潜力差异化评估, 文献[22-23]分别从需求灵活性表征与城市建筑柔性资源禀赋角度分析了不同气候条件下的调节容量边界。文献[24]进一步面向大规模暖通空调(heating, ventilation and air conditioning, HVAC)负荷构建连续时间聚合模型, 刻画异质设备在不确定条件下的聚合响应能力。上述工作为刻画建筑柔性来源提供了重要基础, 但多采用解析积分或逐步差分求解, 大规模异构房间场景下计算开销显著。为提高仿真预测效率, 文献[25]将物理信息神经网络(physics-informed neural network, PINN)引入建筑热建模, 通过在损失函数中嵌入热力学方程残差实现了物理一致的快速预测; 文

献[26]提出面向控制的因果机器学习建模方法。然而, 现有研究多局限于单区域或简化工况, 对空调功率响应惯性与多节点热网络级联过程的仿真精度仍有待提升。

在功率分配与约束表征方面, 文献[27]提出面向商业建筑的分布式实时 HVAC 控制方案, 文献[28-29]分别采用经济模型预测控制与鲁棒模型预测控制处理多时段协调问题, 文献[30]建立温控负荷聚合模型用于微网能量管理, 文献[31-32]从分布式与分层调度角度协调聚合负荷灵活性。上述方法物理约束保证充分, 但高维异构场景下计算负担较重。人工智能方法已在智能电网运行控制、调度决策和在线优化中得到广泛研究[33-34]。文献[35]利用 Transformer 处理多能负荷预测问题, 数据驱动方法推理速度快, 但难以天然保证功率守恒与安全边界等硬约束。在约束表征层面, 现有框架处理运行规则多依赖预定义规则表或人工标注, 面对语义多样、场景动态变化的实际需求泛化不足。大语言模型(large language model, LLM)在电力系统通用智能与生成式应用方面已受到关注[36-37], 并在结构化信息抽取和多模态语义理解方面展现出应用潜力, 但在楼宇级约束生成及与下游物理优化尚属空白。

基于此, 本文提出一种融合多模态 LLM 与 PINN 的楼宇空调负荷聚合灵活调控框架。考虑空调电功率响应与房间热动态过程的差异, 构建电热分层 PINN 仿真模型, 形成房间级动态功率安全边界; 引入多模态 LLM 解析运行语义、功能拓扑和视觉场景, 建立非结构化信息到房间级约束参数的转化方法; 进一步构建基于特征线性调制(feature-wise linear modulation, FiLM)的 Transformer 候选分配与二次规划(quadratic programming, QP)可行域投影相结合的两阶段功率分配模型, 并基于 QP 对偶信息实现语义感知下的快速可行分配与约束放宽价值量化。

1 基于电热分层 PINN 的楼宇空调负荷动态仿真建模

空调聚合负荷的动力学包含两类时间尺度差异显著的过程: 压缩机电冷能量转换的快动态与围护结构热传递的慢动态。本文将上述过程分解为分层 PINN 仿真模型的级联映射, 其整体结构如图 1 所示。PINN-I 内嵌双指数衰减物理先验, 拟合 HVAC 电功率对阶跃指令的工况自适应响

应；PINN-II 基于三阻三容 (three-resistance three-capacitance, 3R3C) 热网络的物理残差约束，预测空气、墙体与地板三节点温度演化轨迹。每级子模型仅拟合单一物理过程，降低学习难度的同时保留端到端的误差可追踪性。

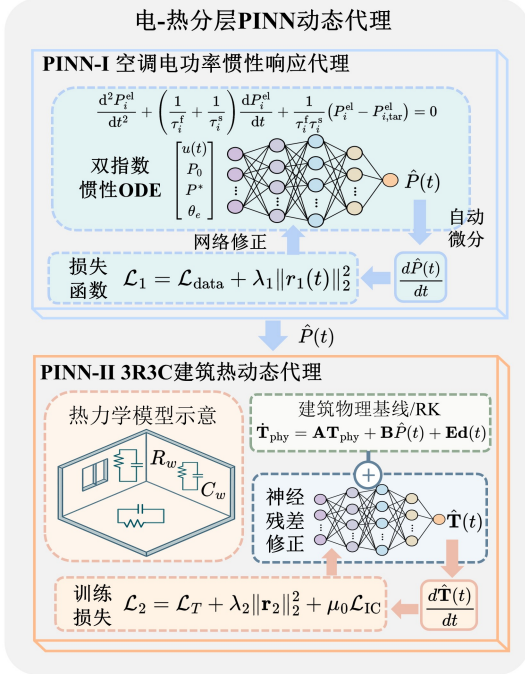


图 1 电-热分层 PINN 动态仿真建模框架

Fig. 1 Hierarchical Electric-Thermal Dynamic Modeling Framework Based on PINN

1.1 空调电功率惯性响应建模

HVAC 系统电功率对阶跃指令的响应受压缩机变频驱动与制冷剂循环惯性的共同作用，可近似为双时间常数的惯性过程，其响应表达式为：

$$P_i^{\text{el}}(t) = P_{i,\text{tar}}^{\text{el}} - \Delta P_i^{\text{el}} \left[\alpha_i \exp\left(-\frac{t}{\tau_i^f}\right) + (1 - \alpha_i) \exp\left(-\frac{t}{\tau_i^s}\right) \right] \quad (1)$$

式中：\$P_{i,0}^{\text{el}}\$ 和 \$P_{i,\text{tar}}^{\text{el}}\$ 分别为房间基准与目标电功率，kW；\$\Delta P_i^{\text{el}}\$ 为功率偏差；\$t\$ 为阶跃指令后的连续时间，s；\$\alpha_i\$ 为快模态分量系数；\$\tau_i^f\$ 和 \$\tau_i^s\$ 分别为快、慢时间常数，s，且 \$\tau_i^f < \tau_i^s\$；

上述双指数结构等价于如下二阶 ODE：

$$\frac{d^2 P_i^{\text{el}}}{dt^2} + \left(\frac{1}{\tau_i^f} + \frac{1}{\tau_i^s} \right) \frac{dP_i^{\text{el}}}{dt} + \frac{1}{\tau_i^f \tau_i^s} (P_i^{\text{el}} - P_{i,\text{tar}}^{\text{el}}) = 0 \quad (2)$$

响应参数依赖于工况条件，采用参数化神经网络学习其非线性映射：

$$[\alpha_i, s_i^f, s_i^s] = f_\theta(\mathbf{z}_i) \quad (3)$$

$$\mathbf{z}_i = [P_{i,0}^{\text{el}}, P_{i,\text{tar}}^{\text{el}}, \Delta P_i^{\text{el}}, \tau_i^{\text{eff}}, T^{\text{out}}, \eta_i, \rho_{i,0}, \rho_{i,\text{tar}}]^T \quad (4)$$

式中 \$f_\theta\$ 为参数化神经网络，\$\theta\$ 为其网络参数；\$\mathbf{z}_i\$

为工况特征向量；\$\tau_i^{\text{eff}}\$ 为有效惯性时间常数，s；\$T^{\text{out}}\$ 为室外温度，°C；\$\eta_i\$ 为性能系数；\$\rho_{i,0}\$ 和 \$\rho_{i,\text{tar}}\$ 分别为基准和目标部分负荷率；\$s_i^f\$ 和 \$s_i^s\$ 为快、慢时间常数缩放因子。

快、慢时间常数由网络输出的缩放因子与有效惯性时间常数共同确定。训练阶段，将式(1)以网络预测参数实例化得到功率轨迹，并将式(2)左端定义为物理残差：

$$r_i^1(t) = \frac{d^2 P_i^{\text{el}}}{dt^2} + \left(\frac{1}{\tau_i^f} + \frac{1}{\tau_i^s} \right) \frac{dP_i^{\text{el}}}{dt} + \frac{1}{\tau_i^f \tau_i^s} (P_i^{\text{el}}(t) - P_{i,\text{tar}}^{\text{el}}) \quad (5)$$

式中 \$r_i^1(t)\$ 为一阶段物理残差。时间导数由自动微分获得。PINN-I 的训练损失由物理残差损失和数据监督损失组成：

$$\mathcal{L}^1 = \lambda_{\text{phy}}^1 \frac{1}{N_c} \sum_{j=1}^{N_c} \|r_i^1(t_j)\|^2 + \frac{1}{N_s} \sum_{j=1}^{N_s} \|P_i^{\text{el}}(t_j) - P_i^{\text{el}}(t_j)\|^2 \quad (6)$$

式中 \$P_i^{\text{el}}(t_j)\$ 为采样时刻 \$t_j\$ 下 ODE 数值解提供的监督标签，kW；\$N_s\$ 和 \$N_c\$ 分别为数据监督点和物理配点数量；\$\lambda_{\text{phy}}^1\$ 为物理损失权重系数。

1.2 基于 3R3C 热网络的房间热动态建模

采用 3R3C 等效热参数模型描述单房间热动力学，将热系统简化为空气、墙体与地板 3 个集总节点。定义房间的温度状态向量为：

$$\mathbf{x}_i(t) = [T_i^{\text{air}}(t), T_i^{\text{wall}}(t), T_i^{\text{floor}}(t)]^T \quad (7)$$

式中 \$T_i^{\text{air}}(t)\$、\$T_i^{\text{wall}}(t)\$ 和 \$T_i^{\text{floor}}(t)\$ 分别为空气节点、墙体节点和地板节点温度，°C。

各节点经内外热阻耦合，制冷功率由 PINN-I 输出经性能系数转换得到。三节点热平衡写为如下状态方程：

$$\mathbf{C}_i \frac{d\mathbf{x}_i}{dt} = \mathbf{A}_i \mathbf{x}_i + \mathbf{B}_i^u u_i(t) + \mathbf{B}_i^d \mathbf{d}_i(t) \quad (8)$$

$$\mathbf{C}_i = \text{diag}(C_i^{\text{air}}, C_i^{\text{wall}}, C_i^{\text{floor}}) \quad (9)$$

式中 \$\mathbf{C}_i\$ 为热容对角矩阵，J/K；\$\mathbf{A}_i\$ 为由空气-墙体热阻、墙体-室外热阻、墙体-走廊热阻及地板-地基热阻构成的系统矩阵；\$u_i(t)\$ 为控制输入，\$\eta_i\$ 为性能系数；\$\mathbf{d}_i(t)\$ 为包含内热源、太阳辐射、窗户传热、通风换热及温度边界条件的外扰向量；\$\mathbf{B}_i^u\$、\$\mathbf{B}_i^d\$ 分别为控制和扰动输入矩阵。

PINN-II 采用物理基线结合神经修正架构预

测三节点温度轨迹:

$$\hat{\mathbf{x}}_i(t) = \mathbf{x}_i^{\text{base}}(t) + \psi(t) \odot \mathbf{s} \odot \tanh\left(g_\phi(\mathbf{x}_i(0), \mathbf{p}_i, P_i^{\text{el}}(t), t)\right) \quad (10)$$

式中, $\mathbf{x}_i^{\text{base}}(t)$ 为以式(8)所述 3R3C 状态方程经四阶 Runge-Kutta 数值积分得到的物理基线轨迹; g_ϕ 为参数化神经网络, 输出各节点的修正量, ϕ 为对应网络参数; $\mathbf{s} = [s_{\text{air}}, s_{\text{wall}}, s_{\text{floor}}]^T$ 为节点修正幅度缩放向量; $\psi(t)$ 为时间包络函数; $\mathbf{p}_i$ 为静态热参数向量; $\odot$ 为逐元素乘法。包络函数在初始时刻取零, 从而硬性保证预测温度的初始条件约束。

PINN-II 的物理残差定义为热平衡方程偏差, 将式(10)的预测轨迹代入式(8)的状态方程, 定义三节点物理残差向量为:

$$\mathbf{r}_i^{\text{II}}(t) = \frac{d\hat{\mathbf{x}}_i}{dt} - \mathbf{C}_i^{-1}(\mathbf{A}_i \hat{\mathbf{x}}_i + \mathbf{B}_i^u u_i(t) + \mathbf{B}_i^d d_i(t)) \quad (11)$$

式中 $\mathbf{r}_i^{\text{II}}(t)$ 为二阶段物理残差向量, 各矩阵定义同式(8), 温度对时间的导数通过自动微分计算。不同节点热容量级差异显著, 为此引入特征时间常数对角矩阵进行无量纲化:

$$\tilde{\mathbf{r}}_i^{\text{II}}(t) = \mathbf{T}_i \mathbf{r}_i^{\text{II}}(t) \quad (12)$$

$$\mathbf{T}_i = \text{diag}(\tau_{i,\text{air}}, \tau_{i,\text{wall}}, \tau_{i,\text{floor}}) \quad (13)$$

式中 $\mathbf{T}_i$ 为特征时间常数对角矩阵, 各元素由对应节点热容与总热导之比确定。

1.3 电热分层 PINN 训练与误差传播分析

分层 PINN 的级联映射形式化为:

$$\hat{\mathbf{x}}_i(t) = \mathcal{M}^{\text{II}}\left(\mathcal{M}^{\text{I}}(P_{i,\text{tar}}^{\text{el}}, \mathbf{z}_i; \boldsymbol{\theta}), \mathbf{p}_i, \mathbf{x}_i(0); \phi\right) \quad (14)$$

式中: $\mathcal{M}^{\text{I}}(\cdot; \boldsymbol{\theta})$ 和 $\mathcal{M}^{\text{II}}(\cdot; \phi)$ 分别代表 PINN-I 和 PINN-II 的参数化映射; $\mathbf{p}_i$ 为房间 i 的静态热参数向量, 包含全部热阻、热容、窗户参数及通风参数。

两级模型分阶段训练: PINN-I 先在 ODE 标签上训练并冻结参数, PINN-II 再以 PINN-I 在线推理输出为驱动输入进行训练, 使其在训练阶段即暴露于上级预测误差。

PINN-II 的总损失由数据监督、物理残差和初值约束三项组成:

$$\mathcal{L}^{\text{II}} = \mathcal{L}_{\text{data}}^{\text{II}} + \lambda_{\text{phy}}^{\text{II}} \mathcal{L}_{\text{phy}}^{\text{II}} + \lambda_{\text{ic}}^{\text{II}} \mathcal{L}_{\text{ic}}^{\text{II}} \quad (15)$$

式中 $\mathcal{L}_{\text{data}}^{\text{II}}$ 为预测温度与标签间的均方误差; $\mathcal{L}_{\text{phy}}^{\text{II}}$ 为残差向量在物理配点上的均方范数; $\mathcal{L}_{\text{ic}}^{\text{II}}$ 为初值约束惩罚; $\lambda_{\text{phy}}^{\text{II}}$ 、 $\lambda_{\text{ic}}^{\text{II}}$ 为对应权重系数。

级联结构中, PINN-I 的功率预测误差经 COP

转换传播至温度预测。对式(8)的空气节点分量进行线性化, 传播至空气温度的误差可近似为:

$$\varepsilon_{i,\text{air}}^{\text{prop}}(t) = \frac{\eta_i}{C_i^{\text{air}}} \int_0^t \varepsilon_i^{\text{I}}(s) ds \quad (16)$$

式中 $\varepsilon_i^{\text{I}}(t)$ 为 PINN-I 功率预测误差; $\varepsilon_{i,\text{air}}^{\text{prop}}(t)$ 为由功率误差传播得到的空气温度误差。式(16)表明温度误差与功率误差时间积分成正比, 高 COP 与低热容房间更敏感。定义误差传播比为:

$$r_i^{\text{prop}} = \frac{\max_t |\varepsilon_{i,\text{air}}^{\text{prop}}(t)|}{\max_t |\varepsilon_i^{\text{I}}(t)| \cdot t_{\text{hor}}} \cdot \frac{C_i^{\text{air}}}{\eta_i} \quad (17)$$

式中 r_i^{prop} 为误差传播比; t_{hor} 为预测时域长度, s。 $r_i^{\text{prop}} \approx 1$ 表明功率误差近似线性传播, $r_i^{\text{prop}} \ll 1$ 表明热容对误差有显著衰减。

2 基于多模态 LLM 的楼宇空调负荷聚合灵活调控方法

楼宇接收电网总功率调节请求后, 需要将聚合功率分解至具备不同功能属性、舒适边界和参与权限的房间。本文利用多模态 LLM 解析运行语义、功能拓扑和视觉场景, 生成房间级结构化约束参数; 在此基础上, 采用 FiLM-Transformer 生成候选分配结果, 并通过 QP 投影满足功率平衡与房间边界约束; 当可行响应容量不足时, 进一步利用 QP 对偶信息评估语义约束放宽的边际价值。整体流程如图 2 所示。

2.1 基于 LLM 的多模态解析与约束生成

商业楼宇中, 房间用途、业务优先级、温控权限及锁定指令等运行规则以自然语言形式存在, 无法直接参与优化计算。本文采用 LLM 将多模态运行语义映射为可计算的结构化约束参数。

LLM 的输入信息空间由三个模态组成: 自然语言调控指令; 被观测房间的现场监控图像, 经视觉编码器提取为视觉令牌; 楼宇功能分区拓扑, 描述功能区与房间的映射关系。上述多模态信息经系统提示模板编排为统一的令牌序列:

$$\mathbf{x} = [\mathcal{P}_{\text{sys}}; \text{Ser}(\mathcal{G}); \mathbf{H}^{\text{vis}}; \mathcal{T}] \quad (18)$$

式中 $\mathcal{P}_{\text{sys}}$ 为系统提示模板; $\mathcal{T}$ 为自然语言调控指令; $\mathcal{G} = (\mathcal{Z}, \mathcal{M})$ 为功能分区拓扑, $\mathcal{Z}$ 为功能区集合, $\mathcal{M}: \mathcal{Z} \rightarrow 2^{\{1, \dots, N\}}$ 为区域到房间的映射; $\mathbf{H}^{\text{vis}}$ 为各被观测房间视觉令牌的拼接序列, 由监控图像 $I_i \in H \times W \times 3$ 经视觉编码器提取得到, $\mathcal{I}_{\text{obs}}$ 为被观测房间索引集; $\text{Ser}(\cdot)$ 为结构化数据的文本序

列化算子； $[\cdot]$ 表示令牌拼接。LLM 基于上述输入以自回归方式生成结构化输出：

$$p(\mathbf{y}|\mathbf{x};\theta^{\text{LLM}})=\prod_{t=1}^L p_{\theta}(y_t|y_{<t},\mathbf{x}) \quad (19)$$

式中 $\mathbf{y}=(y_1,\dots,y_L)$ 为长度为 L 的输出令牌序列； θ^{LLM} 为预训练语言模型的固定参数。生成过程受预定义约束模式空间 Ω_{schema} 约束，等价于在每步解码中屏蔽不合规令牌：

$$\tilde{p}_{\theta}(y_t|y_{<t},\mathbf{x})=\frac{p_{\theta}(y_t|y_{<t},\mathbf{x})\mathbb{I}(y_t\in V_t^{\text{valid}})}{\sum_{v\in V_t^{\text{valid}}}p_{\theta}(v|y_{<t},\mathbf{x})} \quad (20)$$

式中 V 为模型词表； V_t^{valid} 为第 t 步由 Ω_{schema} 与已生成前缀 $y_{<t}$ 共同确定的有效令牌子集。 $\mathbb{I}(\cdot)$ 为示性函数，条件成立时取 1，否则取 0。

输出序列经解析与功能区展开后，得到各房间的约束向量：

$$\mathbf{c}_i=(\varphi_i^{\text{part}},\varphi_i^{\text{pri}},\varphi_i^{\text{lock}},\delta_i^+,\delta_i^-)^{\text{T}}, \quad \mathbf{c}_i\in\Omega_{\text{schema}} \quad (21)$$

式中 $\varphi_i^{\text{part}}\in[0,1]$ 为参与状态系数； $\varphi_i^{\text{pri}}\in[0,1]$ 为优先级系数，值越小表示保护优先级越高； $\varphi_i^{\text{lock}}\in[0,1]$ 为锁定系数，近零值表示温度锁定状态； δ_i^+ 和 δ_i^- 分别为升温 and 降温容忍度。

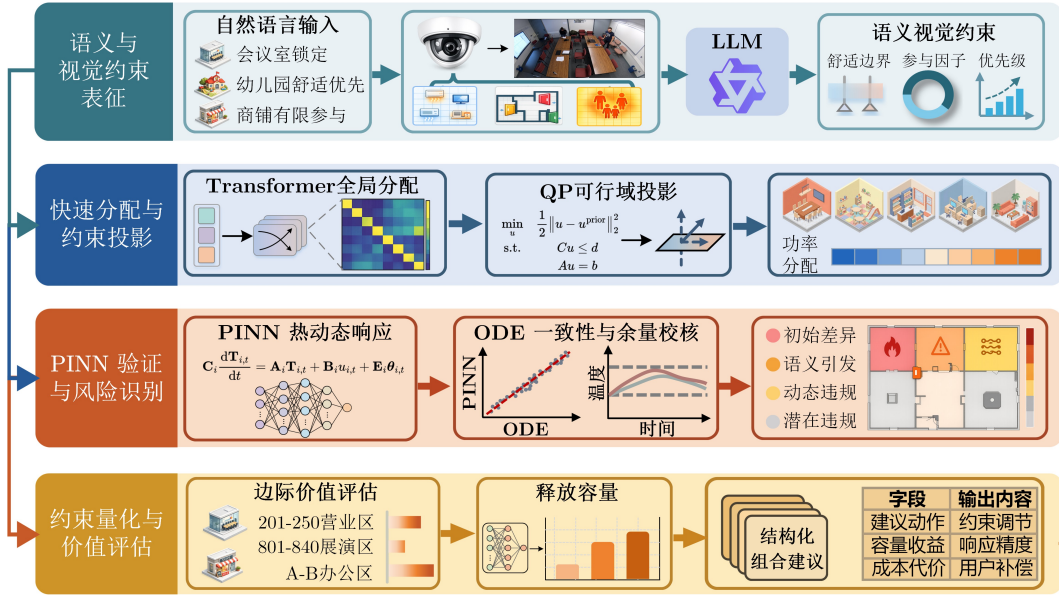


图 2 基于多模态 LLM 的聚合灵活调控流程

Fig. 2 Multimodal LLM-based aggregate flexible control workflow

视觉令牌 H^{vis} 使 LLM 能感知实际场景状态，在低占用率场景下倾向于放宽温度容忍度，高占用场景则收紧。

房间 i 的综合参与因子取各约束系数中最严格者：

$$\varphi_i = \min\{\varphi_i^{\text{part}}, \varphi_i^{\text{pri}}, \varphi_i^{\text{lock}}\} \quad (22)$$

式中 $\varphi_i\in[0,1]$ 值越小表明约束越强。 $\varphi_i=0$ 时房间不参与分配，功率锁定于基准值 $P_{i,0}^{\text{cl}}$ 。取最小值运算保证最严格的单项约束决定参与程度，避免乘积运算造成的过度衰减。

LLM 解析采用前向-后向双通道设计：前向通道生成结构化约束参数，后向通道将分配结果与对偶信息转化为可读的执行摘要。系统同时运行基于规则匹配的确定性解析器作为交叉校核通道，LLM 输出经 Ω_{schema} 验证并与确定性解析结果比对方被接受，以保证约束参数的一致性与可靠性。

对被观测房间，现场监控图像经视觉编码器映射为语义特征。编码器将图像均匀划分为 M 个非重叠图像块，经线性投影与位置编码后，由视觉 Transformer 逐层提取多尺度特征：

$$\mathbf{Z}_i^{\text{vis}} = f_{\text{ViT}}\left(\left\{\mathbf{W}_p \text{vec}(\mathbf{I}_{i,p}) + \mathbf{e}_p^{\text{pos}}\right\}_{p=1}^M; \boldsymbol{\theta}_{\text{vis}}\right) \quad (23)$$

式中 $\mathbf{I}_{i,p}$ 为第 p 个图像块； $\mathbf{W}_p$ 为图像块线性投影矩阵； $\mathbf{e}_p^{\text{pos}}$ 为可学习位置编码； $\boldsymbol{\theta}_{\text{vis}}$ 为视觉编码器的预训练参数。视觉特征经跨模态投影层映射至语言模型的嵌入空间：

$$\mathbf{H}_i^{\text{vis}} = \mathbf{W}_{\text{proj}} \mathbf{Z}_i^{\text{vis}} \in \mathbb{R}^{M' \times d_{\text{LLM}}} \quad (24)$$

式中 $\mathbf{W}_{\text{proj}}$ 为跨模态投影矩阵， M' 为压缩后的视觉令牌数， d_{LLM} 为语言模型隐层维度。投影后的视觉令牌 $\mathbf{H}_i^{\text{vis}}$ 与文本令牌拼接，使 LLM 能同时关注场景视觉特征与文本语义，实现对人员占用、遮蔽状态、门窗开度等场景要素的跨模态推理。

LLM 基于视觉令牌与提示上下文，将图像中

的定性观测映射为定量热扰动修正因子:

$$\kappa_i = h_{\theta_i}(\mathbf{H}_i^{\text{vis}} | \mathcal{P}_{\text{sys}}, \mathcal{G}), \quad i \in \mathcal{I}_{\text{obs}} \quad (25)$$

式中 h_{θ_i} 为 LLM 实现的视觉语义推理映射, 包含三步过程: 识别场景实体; 将实体状态与设计基准对比; 量化修正方向与幅度。修正后的热扰动条件为:

$$\tilde{Q}_i^{\text{int}} = \kappa_i^{\text{occ}} \cdot Q_i^{\text{int}} \quad (26)$$

$$\tilde{Q}_i^{\text{sol}} = \kappa_i^{\text{sol}} \cdot Q_i^{\text{sol}} \quad (27)$$

$$\tilde{V}_i^{\text{mech}} = \kappa_i^{\text{vent}} \cdot \dot{V}_i^{\text{mech}} \quad (28)$$

式中 Q_i^{int} 、 Q_i^{sol} 和 $\dot{V}_i^{\text{mech}}$ 分别为内热源功率、太阳辐射得热和机械通风量; κ_i^{occ} 为人员占用修正因子; κ_i^{sol} 为太阳辐射遮蔽因子; κ_i^{vent} 为通风增强因子。修正后的热扰动条件替代状态方程(8)中外扰向量 $\mathbf{d}_i(t)$ 的对应分量, 使后续安全边界计算反映实时环境状态。

基于修正后的热扰动条件和语义温度容忍度, 以级联 PINN 仿真模型为温度预测器, 定义房间 i 在预测时域 t_{hor} 内的动态功率安全边界:

$$\underline{P}_i: \max_{0 \leq t \leq t_{\text{hor}}} [T_i^{\text{air}}(t; \underline{P}_i) - T_i^{\text{air}}(0)] = \delta_i^+ \quad (29)$$

$$\bar{P}_i: \min_{0 \leq t \leq t_{\text{hor}}} [T_i^{\text{air}}(t; \bar{P}_i) - T_i^{\text{air}}(0)] = -\delta_i^- \quad (30)$$

式中 $T_i^{\text{air}}(t; P)$ 为级联 PINN 在恒定功率 P 和修正热扰动条件下对空气温度的预测值; $\underline{P}_i$ 为使温升恰好触及容忍上限的最小功率, 对应下调极限; $\bar{P}_i$ 为使温降恰好触及容忍下限的最大功率, 对应上调极限。当 $\varphi_i = 0$ 时令 $\underline{P}_i = \bar{P}_i = P_{i,0}^{\text{cl}}$ 。

上述边界值通过对功率变量的二分搜索确定: 每步以 PINN 前向推演评估温度轨迹是否越限, 直至收敛至预设精度。不同语义温度容忍度导致非对称安全域, 即同一房间在上调与下调方向的可调范围不同。

2.2 基于 FiLM-Transformer 与 QP 投影的楼宇群功率分配

直接求解含大量房间级约束的集中优化问题计算量大且难以实时响应。本文采用两阶段策略: 先由神经网络快速生成候选分配, 再经二次规划投影至可行域。

将各房间的静态热参数、动态运行状态及语义约束参数经独立编码器映射后拼接并投影为统一维度的房间编码。同时, 由总功率请求、执行时长、调节方向及基准总功率构成全局任务向量。在 Transformer 编码器的每一层, 引入 FiLM 机制, 使网络输出自适应于当前调节任务。第 l 层的调制过程为:

$$[\gamma_l, \beta_l] = f_l^{\text{gate}}(\mathbf{z}^{\text{task}}) \quad (31)$$

$$\tilde{\mathbf{h}}_l = \mathbf{h}_l \odot (1 + \delta \tanh(\gamma_l)) + \chi \beta_l \quad (32)$$

式中 $\mathbf{z}^{\text{task}}$ 为由总功率请求、执行时长、调节方向及基准总功率构成的全局任务向量; f_l^{gate} 为第 l 层的门控网络, 将任务向量映射为缩放参数 γ_l 和偏移参数 β_l ; $\mathbf{h}_l$ 为调制前的隐层特征; χ 为调制强度系数; $\tanh$ 限制缩放幅度有界。调制后的特征送入标准自注意力编码层继续处理。

网络额外引入一个全局上下文令牌, 参与全部自注意力计算以累积全局信息。最终, 各房间编码经输出头映射为分配比例:

$$P_i = P_{i,0}^{\text{cl}} + \frac{\mathbf{e}^{a_i}}{\sum_{j=1}^N \mathbf{e}^{a_j}} (P^{\text{req}} - P_0^{\text{cl}}) \quad (33)$$

$$a_i = \mathbf{w}_{\text{out}}^T [\tilde{\mathbf{h}}_i^{(L)}; \mathbf{h}_{\text{ctx}}^{(L)}] \quad (34)$$

式中 a_i 为房间 i 的分配逻辑值, 由末层房间编码 $\tilde{\mathbf{h}}_i^{(L)}$ 与全局上下文编码 $\mathbf{h}_{\text{ctx}}^{(L)}$ 拼接后经输出权重向量 $\mathbf{w}_{\text{out}}$ 线性映射得到; P_0^{cl} 为基准总功率, P^{req} 为目标总功率; 候选分配 P_i 通过 Softmax 归一化在结构上满足总功率守恒, 但不保证满足各房间的动态安全边界约束。以最小偏离候选分配为目标, 通过 QP 投影至可行域:

$$\min_P \frac{1}{2} \sum_{i=1}^N w_i (P_i - \bar{P}_i)^2 \quad (35)$$

$$\text{s.t.} \quad \sum_{i=1}^N P_i = P^{\text{req}} \quad (36)$$

$$\underline{P}_i \leq P_i \leq \bar{P}_i, \quad i = 1, \dots, N \quad (37)$$

式中 w_i 为投影误差权重系数, 可根据房间优先级调整; $\underline{P}_i$ 和 $\bar{P}_i$ 分别为动态安全下界和上界。锁定房间的上下界相等, 自动退化为等式约束 $P_i = P_{i,0}^{\text{cl}}$ 。

对参与因子较低的房间, 在投影前将其候选值按参与因子向基准值收缩, 使 QP 的参考点反映语义偏好。上述 QP 为凸二次问题, 可由算子分裂二次规划求解器(operator splitting quadratic program, OSQP)等算法在毫秒级完成求解。记最优解为 P_i^* 。

可行分配 P_i^* 满足静态功率边界约束, 但安全边界系基于恒定功率假设导出。为评估实际调控窗口内的连续热安全性, 以级联 PINN 对各房间进行温度轨迹前向校核, 定义热应力比为:

$$\sigma_i = \frac{\max_{0 \leq t \leq t_{\text{hor}}} |T_i^{\text{air}}(t; P_i^*) - T_i^{\text{air}}(0)|}{\delta_i^{\text{eff}}} \quad (38)$$

式中 δ_i^{eff} 为温变方向对应的有效容忍度：温升时取 δ_i^+ ，温降时取 δ_i^- 。 $\sigma_i < 1$ 表明存在热裕量， $\sigma_i \rightarrow 1$ 表明热裕量趋于耗尽， $\sigma_i > 1$ 表明出现动态越限风险。

2.3 基于对偶边际价值的约束灵活性评估

KKT 条件给出以下对偶变量：功率平衡约束对偶 λ 、下界约束对偶 μ_i^- 和上界约束对偶 μ_i^+ ，其中：

$$w_i(P_i^* - P_i) + \lambda + \mu_i^+ - \mu_i^- = 0, \quad i=1, \dots, N \quad (39)$$

$$\mu_i^- \geq 0, \mu_i^-(P_i^* - P_i) = 0 \quad (40)$$

$$\mu_i^+ \geq 0, \mu_i^+(P_i - P_i^*) = 0 \quad (41)$$

式中 λ 反映总功率守恒约束的边际价值： P^{req} 每增加单位功率，QP 最优目标值增加 λ 。 $\mu_i^+ > 0$ 表明房间上界约束处于激活态，放宽 1 kW 可使 QP 最优值减少 μ_i^+ 。据此，可量化识别制约调节容量的瓶颈房间及其边际贡献。

当电网调节请求超出当前可行域时，存在容量缺额：

$$Q^{\text{def}} = |P^{\text{req}} - P^{\text{feas}}| \quad (42)$$

式中 P^{feas} 为当前可行域中距 P^{req} 最近的可行总功率。此时可通过有选择地放宽部分语义约束释放调节容量。

对语义约束规则 j ，将其放宽至默认值后重新计算安全边界，得到释放的调节容量 ΔQ_j 。定义放宽规则的单位净收益为

$$\Delta \pi_j = \frac{\Delta R_j - \Delta C_j}{\Delta Q_j} \quad (43)$$

式中： ΔQ_j 为放宽规则 j 后释放的调节容量，kW； ΔR_j 为因新增响应容量获得的调节收益； ΔC_j 为放宽约束引起的舒适补偿成本。

筛选单位净收益为正的规则集合，按净收益降序排列，贪心依次选取直至累计释放容量覆盖缺额。最终选定的放宽子集满足：

$$\sum_{j \in \mathcal{J}^*} \Delta Q_j \geq Q^{\text{def}}, \quad \mathcal{J}^* \subseteq \mathcal{J}^+ \quad (44)$$

式中 $\mathcal{J}^+ = \{j: \Delta \pi_j > 0\}$ 为正收益规则集合， $\mathcal{J}^*$ 为最终选定的放宽子集。该机制通过逐条评估正收益约束的容量贡献，在容量缺额覆盖与舒适补偿成本之间形成可计算的组合选择准则。

3 算例分析

本章构建 10 栋商业楼宇空调聚合负荷算例，单栋楼宇包含 100 个异构房间，10 栋楼宇共同构成 1000 房间规模拓展场景，房间类型覆盖办公、

会议、商铺、餐饮、医疗、教学、机房及公共区域等典型功能空间。楼宇群聚合基准功率约为 10.10 MW，调节请求覆盖 0.4-1.7 MW，测试窗口设置为 5 min、15 min 和 30 min，并以 15 min 作为主要分析窗口。算例设置宽泛语义约束与严格语义约束两类运行场景，前者保留默认舒适边界及基础运行限制，后者叠加会议锁定、医疗药房保护、机房温控、商铺营业、教学活动及高敏感业务等语义约束，以形成不同程度的边界收缩和参与权限差异。

多模态约束生成环节采用 Qwen3.5-Flash 作为轻量级语义解析模型，输入包括房间运行说明、功能属性、舒适优先级及视觉场景摘要，输出为结构化约束字段。整体仿真环境基于 Python 部署，候选分配模型采用 FiLM-Transformer 实现，可行域投影由 OSQP 求解器完成，连续热安全校验由 PINN 仿真模型执行。实验平台处理器为 AMD Ryzen 7 4800U with Radeon Graphics。

3.1 视觉占用状态下动态功率边界量化分析

以房间 15 为对象，图 3 展示了不同占用状态及响应时长下的安全域边界对比。图 3 左侧给出了对应房间的空间位置及占用状态示意，右侧则展示了不同工况下安全域边界的定量比较。右侧柱状图显示，同一时间尺度下安全域总宽度在空置、稀疏、拥挤状态间近似恒定，但上/下调能力的分布呈现显著的方向性偏移。从图 3(a)可见，5 min 窗口下空置状态下调能力达 9.80 kW，而拥挤状态仅为 6.52 kW；拥挤状态的上调能力则扩展至 11.21 kW。这种非对称性随响应时长增加而放大，图 3(c)显示 30 min 时两者下调能力差距已达 93%。若采用图中“无视觉”基线方案，将对拥挤房间的下调能力高估 44%，易引入确定的越限风险。

调节总宽度在不同状态下保持稳定，主要源于房间围护结构热容量在短时窗口内近似不变，其在给定时段内可吸收或释放的总热量近似守恒。拥挤状态下调空间的显著收缩，是由空调减载与高内热源叠加导致的快速升温所致，密集的人员与设备发热加速了室温逼近舒适上限的进程。上调能力扩展则源于空调增载降温过程中，高强度内热源抵消部分制冷输入，延缓温度触碰舒适下限的过程。视觉占用状态通过内热源耦合显著重塑了房间调节能力的比例关系，且该效应

随响应时长增加呈非线性放大。

3.2 功率分配性能对比与规模拓展分析

在语义约束转化为房间级功率边界后，聚合功率请求需进一步分解至各房间执行。分配方法不仅需要满足功率平衡和房间边界约束，还需适应房间规模扩大和负荷请求增大带来的在线计算

压力。本节以 Transformer、Transformer+QP 两阶段分配、模型预测控制(model predictive control, MPC)和集中式优化为对比对象，分别在单栋 100 房间和楼宇群 1000 房间场景下分析不同方法的在线执行时间、约束可行性及规模拓展表现。

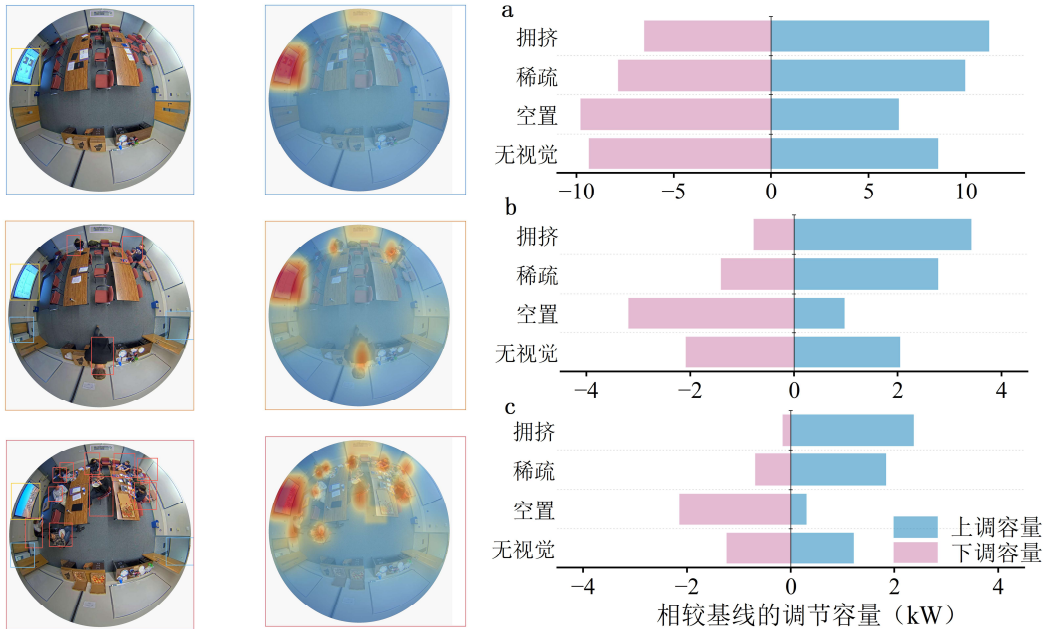


图 3 视觉占用状态下的动态功率边界

Fig. 3 Dynamic power boundaries under different visual occupancy states

表 1 单栋 100 房间不同方法的在线执行时间

Tab.1 Online execution time of different methods in the single-building 100-room case

请求/kW	Transformer/s	T+QP/s	MPC/s	集中式/s
140	0.399	4.626	39.985	65.031
180	0.226	3.251	49.982	65.91
200	0.218	3.643	55.758	78.027
220	0.226	3.687	60.821	69.138
240	0.222	3.661	66.809	91.375
260	0.228	3.424	82.392	106.702

表 1 给出了单栋 100 房间场景下不同方法的在线执行时间。Transformer 推理时间基本不随请求功率变化；T+QP 在候选分配后叠加 QP 投影和 PINN 动态校核，执行时间稳定在数秒范围且未随请求功率升高和约束压力增加出现显著波动；MPC 和集中式优化的求解时间随请求功率增加整体上升，表明迭代优化类方法对约束收紧更为敏感。其原因在于，Transformer 主要计算量来自模型推理，T+QP 额外叠加固定结构凸投影和批量 PINN 前向校核；MPC 和集中式优化的计算量

则受可行域收缩、活跃约束增多和迭代搜索过程影响。140 kW 场景下 T+QP 耗时偏高，主要与首次调用中的矩阵构建、求解器初始化、PINN 批量校核缓存建立和冷启动有关，连续负荷扫描后该部分开销不再重复出现。

单栋场景下，集中式优化可作为中等规模分配基准；当规模扩展至 10 栋 1000 房间后，全局联合优化的变量维度和耦合约束显著增加，继续用于多负荷扫描的计算代价较高。后续楼宇群算例保留滚动 MPC 作为优化类参照，重点比较 Transformer、T+QP 和 MPC 在规模拓展后的在线求解表现。

表 2 给出了楼宇群 1000 房间场景下不同负荷请求的求解结果，其中违约数按 Transformer /T+QP/MPC 的顺序统计，R²用于表征 MPC 与 T+QP 房间级分配结果的一致性。Transformer 和 T+QP 在各负荷等级下均保持秒级执行，T+QP 相较 Transformer 增加了 QP 投影和一次 PINN 动态校核时间，但仍明显低于滚动 MPC；滚动 MPC

耗时随请求功率升高持续增加，高负荷场景下达到分钟量级。随着请求功率接近聚合边界，Transformer 直接分配未显式执行房间边界与语义约束投影，容易出现局部边界违约；T+QP 通过可行域投影将语义约束纳入硬约束条件，从而消除违约，滚动 MPC 同样考虑语义条件，分配结果保持可行。MPC 与 T+QP 的分配相关性在中高负荷区间逐步提高，说明约束压力增大后，可行分配空间收缩，不同优化路径逐渐集中到相近的承载房间。

表 2 1000 房间不同负荷下算法性能对比

Tab.2 Algorithm performance under different loads in the 1000-room case

请求 /kW	Transformer /s	T+QP /s	MPC /s	违约数	R ²
400	2.051	5.610	76.827	0/0/0	—
700	2.113	4.609	91.179	0/0/0	—
1000	2.050	4.708	122.835	0/0/0	—
1300	2.038	4.797	168.129	18/0/0	0.602
1500	2.034	4.743	210.198	76/0/0	0.699
1700	2.084	4.828	240.252	163/0/0	0.905

为进一步分析高负荷下的房间级分配结构，选取表 2 中的 1700 kW 请求场景进行对比。图 4 展示了 Transformer、T+QP 和 MPC 的房间级分配结果，图 5 给出了 MPC 和 Transformer 相对于 T+QP 的分配相关性。图 4 中，T+QP 在 Transformer 候选分配的基础上对越界房间进行投影修正，使分配结果从局部高值和边界外点回到可行域内；MPC 分配结果与 T+QP 保持较高一致性，但在部分房间上呈现更明显的平滑调整特征。Transformer 与 T+QP 的相关性低于 MPC 与 T+QP，说明直接推理结果能够提供有效分配先验，但缺少房间边界和语义约束的显式校正，高负荷下容易产生局部不可行分配。

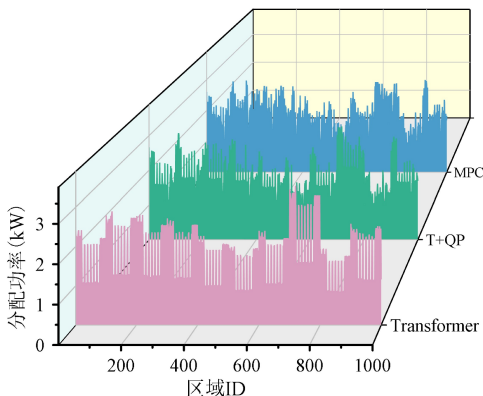


图 4 1.7 MW 请求下的房间级功率分配

Fig. 4 Room-level power allocation under the 1.7 MW request

图 5 进一步给出了同一场景下不同方法与 T+QP 结果的房间级相关性。MPC 与 T+QP 的相关性达到 0.905，高于 Transformer 与 T+QP 的 0.701。高负荷请求压缩了可行分配空间，MPC 和 T+QP 均倾向于将响应分配给边界余量较大的房间，因此二者形成相近的分配结构。Transformer 与 T+QP 之间的差异则反映了 QP 投影对原始先验的修正幅度。

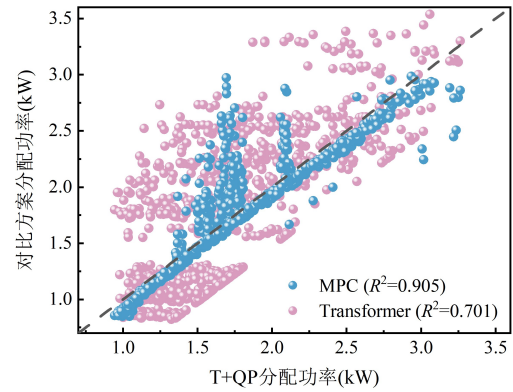


图 5 1.7 MW 请求下的分配相关性

Fig. 5 Allocation correlation under the 1.7 MW request

综上，两阶段分配方法利用 Transformer 降低候选搜索成本，并通过 QP 投影保证功率平衡和房间边界可行性，使楼宇群规模下的功率请求能够在无违约条件下实现秒级执行分解。

3.3 PINN 动态热安全校核与风险识别分析

选取 1000 房间高负荷分配场景，采用 PINN 校核 15 min 窗口内的温度演化过程。本节从典型房间三节点温度轨迹和全楼宇边界余量分布两个层面，评估分配方案在动态执行过程中的热安全性。

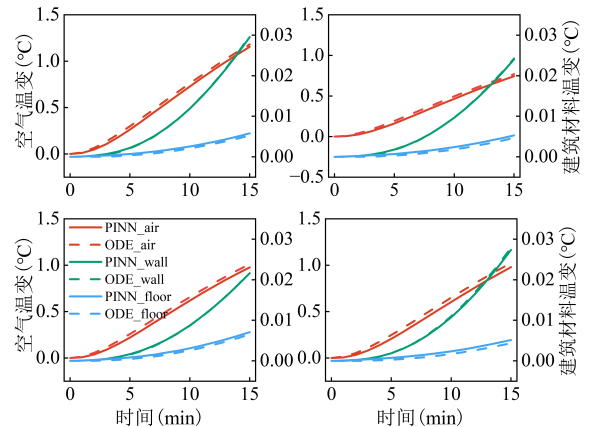


图 6 典型房间三节点温度轨迹的 PINN 校核

Fig. 6 PINN validation of three-node temperature trajectories in representative rooms

图6给出了不同热应力水平下典型房间的空气、墙体和地板三节点温度轨迹。PINN能够同时恢复空气、墙体和地板温度的非同步演化过程。空气节点对功率调节最为敏感,响应初期即出现较快变化;墙体节点受热容影响变化更平滑,温升在窗口后段逐渐累积;地板节点变化幅度最小,体现出更强热惯性。不同房间的轨迹斜率和终端温升存在差异,反映出房间热参数、初始状态和分配功率共同作用下的异构热响应。

图7进一步给出了1000个房间的动态边界余量与应力比分布。边界余量呈明显偏态分布,安全房间占主体,但仍存在一组靠近边界的尾部房间。应力比排序结果显示,高风险房间集中在排序前段,但房间编号并不连续,表明动态风险并非由空间位置单独决定,而是受分配功率、热惯性和语义约束紧度共同影响。

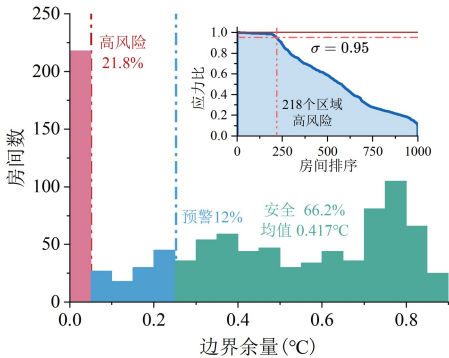


图7 1000 房间动态边界余量与应力比分布

Fig. 7 Distributions of dynamic boundary margin and stress ratio for 1000 rooms

上述结果表明,静态功率边界校验难以完全反映15 min执行过程中的热量积累与滞后传递。级联PINN通过对实际功率响应轨迹进行动态预演,可在指令执行前识别临界风险房间,从而弥补静态分配校验在连续热安全评估中的局限。

3.4 容量缺额下的语义约束量化与价值分析

严格语义约束下,PINN评估得到的1000房间可行响应容量为979.57 kW,低于调度请求1317.56 kW,形成337.98 kW容量缺额。本节基于功率平衡约束和房间边界约束的对偶变量,评估语义约束放宽的容量释放潜力及其边际经济性。图8给出了典型语义约束放宽后的释放容量与单位净收益。

结果表明,容量释放规模与边际经济性并不

一致。办公区A释放容量最大,但单位净收益低于201-250营业区和901-940营业区;实验机房释放容量接近营业区,单位净收益却降为负值。该差异表明,释放容量主要反映房间边界放宽后的物理可调潜力,单位净收益则同时受到舒适补偿成本和响应价值影响。对于医疗、机房、会议等高敏感空间,单纯增加释放容量可能带来更高补偿代价。

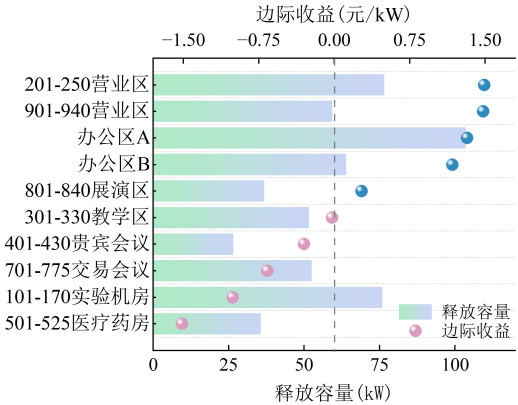


图8 语义约束放宽的容量释放与单位净收益

Fig. 8 Capacity release and unit net benefit under semantic constraint relaxation

基于正收益约束的组合结果如表3所示。各候选策略均可覆盖容量缺额,但释放容量与总净收益的排序并不一致。“营业办公+演播展厅”策略以较小的过量释放获得最高总净收益;“实验机房补足”策略释放容量最大,但因引入负收益约束,总净收益降至最低。该结果表明,容量补足不宜按释放容量单独排序,应结合缺额规模、单位净收益和高敏感约束补偿成本进行组合选择。

表3 不同语义约束放宽策略的容量释放与净收益

Tab.3 Capacity release and net benefit under different semantic constraint relaxation strategies

策略	释放容量/ kW	单位净收益/ 元/kW	总净收益 /元
营业办公+演播展厅	340.37	1.246	423.99
营业办公+医疗药房	339.21	1.061	360.02
营业办公+教育区	355.20	1.162	412.87
营业办公+交易会议	356.09	1.065	379.13
营业办公+实验机房	379.47	0.889	337.32

上述结果表明,QP对偶信息能够将约束活跃程度转化为可比较的边际经济信号。结合PINN容量评估与正收益约束组合搜索,可在容量缺额场景下识别高价值约束放宽对象,并降低高敏感规则放宽带来的补偿成本。

4 结论

本文面向复杂运行场景下楼宇空调负荷的聚合灵活调控问题,提出一种融合多模态大语言模型与物理信息神经网络的楼宇空调负荷聚合灵活调控框架。主要结论如下:

1) 电-热分层 PINN 仿真模型能够将空调电功率惯性响应与房间多节点热动态过程统一到快速评估框架。在保持物理动态一致性的同时,支持房间级动态功率安全边界生成和响应过程温度轨迹预演,使聚合灵活性评估从静态容量估计扩展到连续热安全校核。

2) 多模态 LLM 能够将运行语义、功能拓扑和视觉场景转化为房间级结构化约束参数,使非结构化场景信息进入聚合功率分配过程。结合 FiLM-Transformer 候选分配与 QP 可行域投影,两阶段方法能够在楼宇群规模下实现快速可行分配,并保证功率平衡、房间边界、语义约束和动态热安全约束满足。

3) 基于 QP 对偶信息的边际价值评估能够揭示容量释放潜力与约束协商价值之间的差异。容量缺额场景下,结合释放容量、单位净收益和舒适补偿成本进行约束组合选择,可避免单纯按容量排序导致的高敏感约束放宽,提高聚合灵活性调用的经济性与可操作性。

本文尚未显式考虑末端执行器控制延迟、通信滞后及设备非线性响应等非理想因素。后续研究将进一步引入末端执行层动态特性,并在多能源耦合楼宇场景中验证所提方法的适用性。

参考文献

- [1] 张智刚,康重庆. 碳中和目标下构建新型电力系统的挑战与展望[J]. 中国电机工程学报, 2022, 42(8): 2806-2819.
Zhang Zhigang, Kang Chongqing. Challenges and prospects for constructing the new-type power system towards a carbon neutrality future[J]. Proceedings of the CSEE, 2022, 42(8): 2806-2819(in Chinese).
- [2] Rubasinghe O, Zhang T, Zhang X, et al. Highly accurate peak and valley prediction short-term net load forecasting approach based on decomposition for power systems with high PV penetration[J]. Applied Energy, 2023, 333: 120641.
- [3] 范帅,危怡涵,何光宇,等. 面向新型电力系统的需求响应机制探讨[J]. 电力系统自动化, 2022, 46(7).
Fan Shuai, Wei Yihan, He Guangyu, et al. Discussion on demand response mechanism for new power systems[J]. Automation of Electric Power Systems, 2022, 46(7)(in Chinese).
- [4] 王彩霞,时智勇,梁志峰,等. 新能源为主体电力系统的需求侧资源利用关键技术及展望[J]. 电力系统自动化, 2021, 45(16): 37-48.
Wang Caixia, Shi Zhiyong, Liang Zhifeng, et al. Key technologies and prospects of demand-side resource utilization for power systems with renewable energy as the main body[J]. Automation of Electric Power Systems, 2021, 45(16): 37-48(in Chinese).
- [5] 康重庆,陈启鑫,苏剑,等. 新型电力系统规模化灵活资源虚拟电厂科学问题与研究框架[J]. 电力系统自动化, 2022, 46(18): 3-14.
Kang Chongqing, Chen Qixin, Su Jian, et al. Scientific problems and research framework of virtual power plant with enormous flexible distributed energy resources in new power system[J]. Automation of Electric Power Systems, 2022, 46(18): 3-14(in Chinese).
- [6] Wang Y, Tang Y, Xu Y, et al. A distributed control scheme of thermostatically controlled loads for the building-microgrid community[J]. IEEE Transactions on Sustainable Energy, 2019, 11(1): 350-360.
- [7] Utama C, Troitzsch S, Thakur J. Demand-side flexibility and demand-side bidding for flexible loads in air-conditioned buildings[J]. Applied Energy, 2021, 285: 116418.
- [8] 刘广生,李成鑫,侯治吉,等. 计及用户舒适度的空调负荷可调节能力评估及响应策略[J]. 电力系统自动化, 2023, 47(21): 58-66.
Liu Guangsheng, Li Chengxin, Hou Zhiji, et al. Adjustable capability evaluation and response strategy of air-conditioning load considering user comfort[J]. Automation of Electric Power Systems, 2023, 47(21): 58-66(in Chinese).
- [9] 靳小龙,穆云飞,贾宏杰,等. 融合需求侧虚拟储能系统的冷热电联供楼宇微网优化调度方法[J]. 中国电机工程学报, 2017, 37(2): 581-591.
Jin Xiaolong, Mu Yunfei, Jia Hongjie, et al. Optimal scheduling method for a combined cooling, heating and power building microgrid considering virtual storage system at demand side[J]. Proceedings of the CSEE, 2017, 37(2): 581-591(in Chinese).
- [10] 刘志伟,苗世洪,杨炜晨,等. 计及电气特性的空调负荷建模及集群控制策略[J]. 电力自动化设备, 2022, 42(1).
Liu Zhiwei, Miao Shihong, Yang Weichen, et al. Air-conditioning load modeling and cluster control strategy considering electrical characteristics[J]. Electric Power Automation Equipment, 2022, 42(1)(in Chinese).

- [11] Huang S, Katipamula S, Lutes R G. Experimental investigation on thermal inertia characterization of commercial buildings for demand response[J]. *Energy and Buildings*, 2021, 252: 111384.
- [12] Cordova S, Canizares C A, Lorca A, et al. Aggregate modeling of thermostatically controlled loads for microgrid energy management systems[J]. *IEEE Transactions on Smart Grid*, 2023, 14(6): 4169-4181.
- [13] Tang L, Xie H, Wang Y, Xu Z. Deeply flexible commercial building HVAC system control: a physics-aware deep learning-embedded MPC approach[J]. *Applied Energy*, 2025, 388: 125631.
- [14] Le L, Fang J, Ai X, et al. Aggregation and scheduling of multi-chiller HVAC systems in continuous-time stochastic unit commitment for flexibility enhancement[J]. *IEEE Transactions on Smart Grid*, 2023, 14(4): 2774-2785.
- [15] Jiang S, Hui H, Song Y. Credible demand response capacity evaluation for building HVAC systems based on grey-box models[J]. *Applied Energy*, 2025, 395: 126144.
- [16] Liu X, Wu Y, Wu H. Enhancing HVAC energy management through multi-zone occupant-centric approach: a multi-agent deep reinforcement learning solution[J]. *Energy and Buildings*, 2024, 303: 113770.
- [17] 周特, 周意, 王帅, 等. 变频温控负荷参与电力系统调频的综合惯量支撑技术评述[J]. *中国电机工程学报*, 2026, 46(3): 970-994.
Zhou Te, Zhou Yi, Wang Shuai, et al. Review on comprehensive inertia support technology of variable-frequency thermostatically controlled loads participating in power system frequency regulation[J]. *Proceedings of the CSEE*, 2026, 46(3): 970-994(in Chinese).
- [18] Ghiaus C, Ahmad N. Thermal circuits assembling and state-space extraction for modelling heat transfer in buildings[J]. *Energy*, 2020, 195: 117019.
- [19] Jin X, Qi F, Wu Q, et al. Integrated optimal scheduling and predictive control for energy management of an urban complex considering building thermal dynamics[J]. *International Journal of Electrical Power and Energy Systems*, 2020, 123: 106273.
- [20] Tian G, Sun Q Z, Wang W. Real-time flexibility quantification of a building HVAC system for peak demand reduction[J]. *IEEE Transactions on Power Systems*, 2022, 37(5): 3862-3874.
- [21] 金旭, 张远实, 李明, 等. 考虑热舒适度的居民空调负荷调控潜力差异化评估[J]. *电力系统自动化*, 2024, 48(1): 1-11.
Jin Xu, Zhang Yuanshi, Li Ming, et al. Differentiated evaluation of regulation potential of residential air-conditioning load considering thermal comfort[J]. *Automation of Electric Power Systems*, 2024, 48(1): 1-11(in Chinese).
- [22] Zhou X, Yap E W, Dou W, et al. Demand flexibility characterisation in non-residential buildings: a review[J]. *Applied Energy*, 2025, 393: 126100.
- [23] Qi T, Hui H, Feng W, et al. Flexible resource endowment of urban buildings considering climate diversity in China[J]. *Carbon Neutrality*, 2025, 4: 22.
- [24] Liu J, Ai X, Fang J, Cui S, Wang S, Yao W, Wen J. Continuous-time aggregation of massive flexible HVAC loads considering uncertainty for reserve provision in power system dispatch[J]. *IEEE Transactions on Smart Grid*, 2024, 15(5): 4835-4849.
- [25] Drgona J, Tuor A R, Vrabie D L. Physics-informed neural networks for control-oriented thermal modeling of buildings[J]. *Applied Energy*, 2022, 314: 118852.
- [26] van der Heijde B, Ferrando M, Salom J, et al. What-if: a causal machine learning approach to control-oriented modelling for building thermal dynamics[J]. *Applied Energy*, 2025, 377: 124550.
- [27] Yu L, Xie D, Jiang T, et al. Distributed real-time HVAC control for cost-efficient commercial buildings under smart grid environment[J]. *IEEE Internet of Things Journal*, 2018, 5(1): 44-55.
- [28] Zheng W, Wang D, Wang Z. Economic model predictive control for building HVAC system: a comparative analysis of model-based and data-driven approaches using the BOPTEST framework[J]. *Applied Energy*, 2024, 374: 123969.
- [29] Hu Z, Gao Y, Sun L, et al. Improved robust model predictive control for residential building air conditioning and photovoltaic power generation with battery energy storage system under weather forecast uncertainty[J]. *Applied Energy*, 2024, 371: 123652.
- [30] Song M, Gao C, Ma S, et al. Distributed scheduling of HVACs based on transactive energy and ADMM[J]. *Applied Energy*, 2022, 325: 119831.
- [31] Song M, Sun W, Wang Y, et al. Hierarchical scheduling of aggregated TCL flexibility for transactive energy in power systems[J]. *IEEE Transactions on Smart Grid*, 2020, 11(3): 2452-2463.
- [32] Li Y, Yu C, Yan M, Yang T, Zeng Z, Chai T. Deep reinforcement learning for smart grid operations: algorithms, applications, and prospects[J]. *Proceedings of the IEEE*, 2023, 111(9): 1055-1096.
- [33] 韩富佳, 王晓辉, 乔骥, 等. 基于人工智能技术的新型电力系统负荷预测研究综述[J]. *中国电机工程学报*, 2023, 43(22): 8569-8592.

Han Fujia, Wang Xiaohui, Qiao Ji, et al. Review of load forecasting for new power systems based on artificial intelligence technology[J]. Proceedings of the CSEE, 2023, 43(22): 8569-8592(in Chinese).

- [34] Wang C, Wang Y, Ding Z, et al. A transformer-based method of multienergy load forecasting in integrated energy system[J]. IEEE Transactions on Smart Grid, 2022, 13(4): 2703-2714.

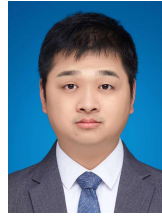
- [35] 赵俊华, 文福拴, 黄建伟, 等. 基于大语言模型的电力系统通用人工智能展望: 理论与应用[J]. 电力系统自动化, 2024, 48(6): 13-28.

Zhao Junhua, Wen Fushuan, Huang Jianwei, et al. Prospect of artificial general intelligence for power systems based on large language model: theory and applications[J]. Automation of Electric Power Systems, 2024, 48(6): 13-28(in Chinese).

- [36] 丁俐夫, 陈颖, 肖谭南, 等. 基于大语言模型的新型电力系统生成式智能应用模式初探[J]. 电力系统自动化, 2024, 48(19): 1-13.

Ding Lifu, Chen Ying, Xiao Tannan, et al. Preliminary exploration of a new generative intelligent application model for power systems based on large language

models[J]. Automation of Electric Power Systems, 2024, 48(19): 1-13(in Chinese).



陈胜

*通信作者: 陈胜 (1990), 男, 博士, 教授, 博士生导师, 主要研究方向为新型电力系统下的综合能源系统规划与运行, chenshenghu@163.com;

赵宇 (2002), 男, 硕士研究生, 主要研究方向为建筑综合能源系统优化运行与需求侧灵活资源调控, 18536203710@163.com;

惠红勋 (1992), 男, 博士, 助理教授, 博士生导师, 主要研究方向为电力系统优化与控制、需求侧资源调控和电力市场, hongxunhui@um.edu.mo;

苏慧玲 (1983), 女, 博士, 国网江苏省电力有限公司营销服务中心, 主要研究方向为电力系统电力计量、需求侧管理与负荷资源调控, suhuil@163.com;

卫志农 (1962), 男, 教授, 博士生导师, 主要研究方向为电力系统运行分析与控制、输配电系统自动化, wzn_nj@263.net。