

一种面向算力基础设施的 AI 能耗管控方案

燕 艳¹ 徐一伟¹ 宋德华¹ 王艺皓¹
YAN Yan XU Yiwei SONG Dehua WANG Yihao

摘 要

数据中心面临能耗过高的严峻挑战,如制冷系统冗余度过高、冷冻水系统及末端设备设置不合理,以及调控机制响应滞后等方面。为更好地解决这些问题,文章提出了一种基于数据驱动、平台支撑的算力基础设施 AI 能耗管控方案。该方案结合专家经验、数据挖掘、AI 自学习三大引擎,构建了一个 AI 能耗管控平台,同时创新性设计了末端空调按需自动调控技术和冷冻水系统及末端机房多点自动联控技术,实现了对数据中心制冷系统的实时动态监控与管理。通过该方案,数据中心平均制冷节能率提升了 20%,PUE 平均降幅了 0.13,有效缓解了数据中心能耗过高的问题。

关键词

AI 能耗管控; 自动调控; 多点联控; 专家经验; 数据挖掘

doi: 10.3969/j.issn.1672-9528.2025.06.031

0 引言

近年来,我国数据中心规模高速增长,但 PUE 居高不下、高能耗问题突出,为推动信息通信行业绿色低碳高质量发展,国家陆续出台具备战略性、引领性和创新性的系列政策和举措。2021 年 7 月 4 日,工业和信息化部印发了《新型数据中心发展三年行动计划(2021—2023 年)》^[1],计划中提到:“新建大型及以上数据中心 PUE 降低到 1.3 以下。新建边缘数据中心 PUE 一般不高于 1.5。加速改造升级‘老旧小散’数据中心”。2022 年 8 月 22 日,工业和信息化部等七部门印发了信息通信行业绿色低碳发展行动计划(2022—2025 年)^[2],计划中提到:“最终实现到 2025 年,全国新建大型、超大型数据中心电能利用效率(PUE)降到 1.3 以下,改建核心机房 PUE 降到 1.5 以下”。与此同时,国务院将碳达峰、碳中和纳入经济社会发展全局,以经济社会发展全面绿色转型为引领,以能源绿色低碳发展是关键,全方位构建节约资源和保护环境的产业结构、生产方式、生活方式与空间格局,坚定不移走生态优先、绿色低碳的高质量发展道路。在此战略背景下,“东数西算”工程加速推进,数据中心作为数字经济的关键基础设施,其绿色化、低碳化发展成为落实国家战略的重要抓手。

在数据中心的发展历程中,能源消耗问题始终是一个不容忽视的挑战。随着数字化和人工智能计算需求的不断增长^[3],新型数据设施的投入使用,使得数据中心的电力需求呈现显著上升趋势。其中,空调系统的能耗大约占据了总能

耗的 45%^[4],而 IT 设备自身的能耗紧随其后。因此,提高冷源的利用效率成为降低数据中心能耗的关键所在。

本文通过深入分析数据中心能耗问题,针对以下方面进行重点研究:

(1) 末端制冷冗余。具体表现为机柜的承载能力未达预期,规划与实际存在显著差异,长期实际装机容量远低于设计容量。此外,机房空调的温度设置普遍依赖于先验经验,而未根据各机柜的实际运行状态进行动态调整。这一做法不仅导致了末端制冷的冗余,还造成了能耗浪费。

(2) 冷源配置不合理。当前冷水机组的参数配置完全基于人工经验,这种方式虽能降低单个设备能耗,但其关联设备能耗激增。数据中心未能根据气温及末端热负载动态变化进行及时调整,使得整个冷冻水系统未能实现最优运行状态。这一状况最终导致制冷量过剩,影响了整体的能源使用效率。

(3) 机房内高能耗问题定位机制尚未建立。机房内的气流组织存在多种问题,其中“跑、冒、滴、漏”等现象尤为显著,这些因素导致部分冷通道功能失效。机房管理者缺乏有效的工具和手段来快速定位能耗问题,因此往往采取降低空调温度的临时措施来应对,这不仅使机房的温湿度控制变得困难,而且无法从根本上解决高能耗的问题。由此,机房的能源效率优化需求迫切,需采取更为科学的策略来确保能耗问题的有效管理。

针对以上问题,本文提出了一种面向算力基础设施的 AI 能耗管控解决方案,构建了一个 AI 能耗管控平台,并创新性设计了机房末端空调按需自动调控技术、冷冻水系统及末端机房多点自动联控技术。旨在解决机房温湿度控制难、末

1. 浪潮通信信息系统有限公司 山东济南 250101

端空调制冷冗余、冷冻水系统设置不合理等问题，推动数据中心向高效、低碳、安全的运行模式转变。

1 AI 能耗管控平台

1.1 平台架构

本文通过分析研究，提出 AI 能耗管控平台。如图 1 所示，该平台基于设计域、控制域、评估域的闭环业务流程，为机房管理者提供全局寻优、测优、调优、至优的智能支撑手段。

AI 能耗管控平台采用云原生架构，以容器、微服务等技术为基础，有效提高系统的灵活性、可扩展性、易用性和可靠性^[5]。此外，该平台采用“云+端”两级部署，云侧基于数据底座进行数据治理、AI 训练/推理/策略、节能验证；端侧部署保护箱、采集网关，负责安全监测、指令自动执行。同时平台采用 Devops 流水线方式，完成自动化独立部署，简化部署难度，提升运维效率。

1.2 三大引擎

三大引擎联动，最大限度寻求全局最优。为提升 AI 能耗管控平台的性能表现，本文结合专家经验引擎、数据挖掘引擎和 AI 引擎，最大限度寻求数据中心能效的最优全局解。接下来对三大引擎进行说明。

(1) 构建专家经验引擎。本文通过提炼历史运维数据，构建专家经验知识库，并在此基础上，建立了一套问题分析引擎，能够快速识别当前能耗问题，并针对性地提出节能方案。此外，构建了一个基于专家经验的多层级问题诊断体系，该体系包括定方向、分原因、剖指标、给建议等，每一层级的分析规则均可由专家根据实际情况灵活调整，从而实现对算力基础设施能耗问题的深度诊断及快速分析。

(2) 构建数据挖掘引擎。鉴于机房环境的复杂性和多样性，针对不同情况，系统需调整不同参数以实现精确的节能调控。数据挖掘引擎通过深度分析机房环境和可调控策略之间的关系，以及构建有监督的策略数据集和策略回归模型，为 AI 能耗管控平台提供了精准的节能策略支持。

(3) 构建 AI 自学习引擎。该引擎能够与实时数据进行快速交互，并根据交互结果输出节能策略。同时，该引擎具备自主学习能力，能够不断优化调整，以深度挖掘节能潜力，进而实现精细化节能管理。通过构建 AI 算法模型，辅助推动数据中心保持高效率、高节能模式运转。

2 关键技术分析

2.1 机房末端空调按需自动调控技术

本文基于环境变量、能耗、设备运行等各种机房数据，创新设计了一种混合 AI 模型自动寻优方法，同时遵循“一房一策”的分析理念，提出了一种机房末端空调按需自动调控技术。该技术实现了对跨厂家、多型号空调设备的按需远程开启与精确调节，能够智能化生成空调最优运行策略，确保机房制冷输出与 IT 设备所需热量的最佳匹配，进而优化机房的热管理效率。此外，该技术也促进实现机房自动化智能调节，使得空调系统的节能效率提升至 20% 以上。接下来将从 6 个方面对本技术原理进行阐述。

(1) 机房环境异常检测与分析

由于 CBLOF (cluster-based local outlier factor) 算法在大规模数据集和捕捉局部异常方面的良好表现^[6]，本文采用该算法对机房环境进行分析，挖掘机房异常问题点。CBLOF 算法是一种基于聚类结果的离群点检测方法。其核心思想是利用聚类算法将数据集划分为多个簇，然后根据每个数据点在

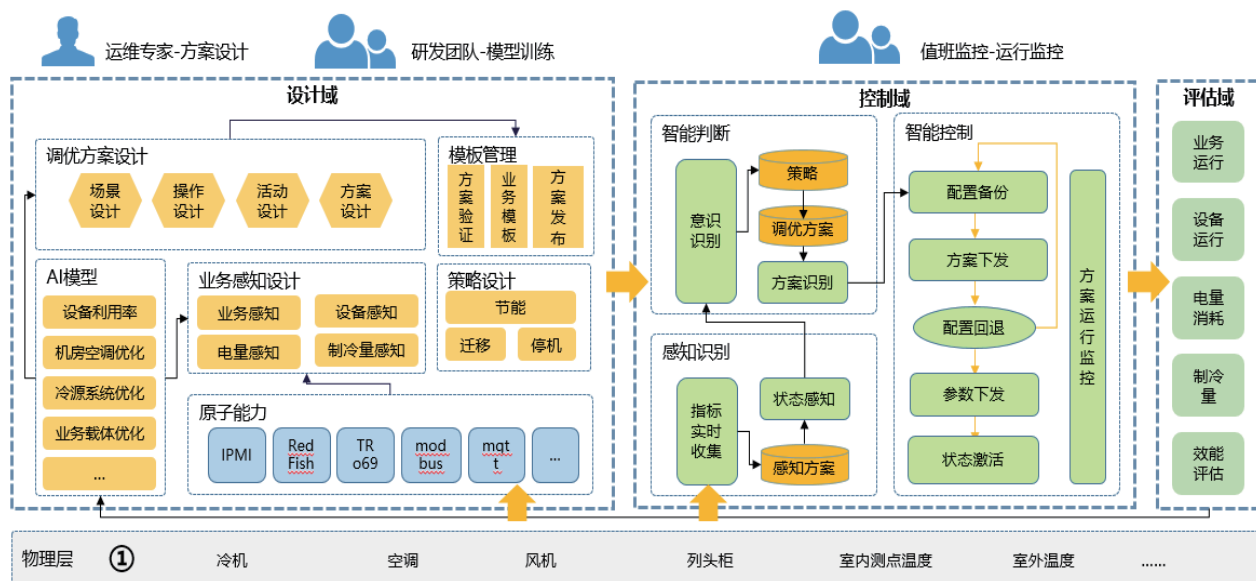


图 1 平台架构

其所在簇中的局部密度以及该簇相对于其他簇的大小来评估其离群程度。本文选用 K-Means 算法进行聚类分析, 结合肘部法则计算其最佳簇数。设定每个簇 C_i , 其大小为 $|C_i|$ 。对于该簇 C_i 中每个数据点 o , 其重要性 $I(o)$ 为:

$$I(o) = \frac{|C_i|}{|D|} \quad (1)$$

式中: $|D|$ 为数据集大小。

该数据点的局部密度 $L(o)$, 其计算为:

$$L(o) = \frac{1}{\text{avg}(o, C_i)} \quad (2)$$

式中: $\text{avg}(o, C_i)$ 是数据点 o 到其所在簇 C_i 中其他数据点的平均距离。

其离群因子 $\text{LOF}(o)$ 为:

$$\text{LOF}(o) = \frac{1}{\min(I(o), L(o))} \quad (3)$$

CBLOF 算法生成的局部离群因子对于每个数据点提供了直观异常程度评分, 根据异常因子的大小, 判断该点是否异常, 从而实现异常点检测。本文通过 CBLOF 算法实现了对机房环境和空调测点温度的深度分析, 当某个时刻的温度值超出聚类检测动态范围时, 则认为此刻的温度是异常的。

(2) 能耗 - 温度相关性分析

本文根据每个机房独特布局、数据特性和气流走向, 选用随机森林 (Random Forest) 算法进行能耗与温度相关性分析, 实现持续动态调优。随机森林是一种以决策树为基分类器的集成学习算法, 通过构建多个决策树并进行投票来提高预测的准确性和稳定性。基于基尼不纯度或信息增益, 随机森林在构建每一棵决策树时, 采用每次划分中考虑 k 个特征的随机采样方式, 并选取最优特征进行划分, 以达到减小 Gini 指数或增加熵值的目的, 最大程度上提升模型分类的准确性。基尼不纯度公式为:

$$\text{Gini}(D) = 1 - \sum_{k=1}^K p_k^2 \quad (4)$$

式中: p_k 为数据集中第 k 类样本的比例。

信息增益公式为:

$$\text{IG}(D, a) = \text{Entropy}(D) - \sum_{v \in \text{Values}(a)} \frac{|D_v|}{D} \text{Entropy}(D_v) \quad (5)$$

(3) 温度聚类构建机房温场

本文采用 K-means 算法对机房的温湿度数据进行聚类分析, 以辅助构建机房的温场模型。首先, 随机选取 K 个温湿度初始点作为质心; 然后为每个数据点找到距离最近的质心, 并将其分配至相应的簇, 随后每个簇的质心更新为该簇内所有数据点的平均值, 直至整个温湿度数据集中的所有点都距离其对应的质心最近。其中, 簇质心更新公式为:

$$\mu_i = \frac{1}{|S_i|} \sum_{x \in S_i} x \quad (6)$$

式中: μ_i 是第 i 个簇的质心; S_i 是第 i 个簇; $|S_i|$ 是簇 S_i 中数据点的数量。

距离计算公式为:

$$d(x_i, \mu_i) = \|x - \mu_i\|^2 = \sum_{j=1}^n (x_j - \mu_{ij})^2 \quad (7)$$

(4) 空调 - 机房温场相关性挖掘

在机房环境中, 多台空调共同作用于测点温度, 这种相互作用呈现非线性关系。本文采用 XGBoost 算法建立空调出风温度与测点温度之间的回归关系。XGBoost 算法是一种高效的梯度提升决策树算法, 通过迭代式训练每个决策树以最小化目标函数, 进而构建回归模型。在 XGBoost 算法中, 目标函数由损失函数和正则化项组成, 即:

$$L(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (8)$$

式中: $l(y_i, \hat{y}_i)$ 是损失函数; $\hat{y}_i$ 是第 i 个样本的预测值; K 是树的数量; f_k 是第 k 棵树的模型; $\Omega(f_k)$ 是正则化项。

针对回归问题, 本文采用平方损失函数, 即:

$$l(y_i, \hat{y}_i) = \frac{1}{2} (y_i - \hat{y}_i)^2 \quad (9)$$

设 T 为叶子节点数量, 则正则化项为:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (10)$$

式中: w_j 是第 j 个叶子节点向量值。

本文基于机房空调历史数据, 利用 XGBoost 算法进行相关性分析, 得到各个空调对测点的影响系数, 进而根据机柜温度情况快速定位需调控空调, 这一方法大幅缩短了空调压缩机的启停波动周期, 有效实现机柜进风温度的均衡控制。

(5) 能耗走势组合预测

本文采用了 LSTM+Prophet 组合方法, 对机房能耗数据进行预测分析。首先, 结合注意力机制对 LSTM 进行了优化。注意力机制能够关注输入序列中的关键信息, 从而提升针对特定任务的模型泛化能力^[7]。引入注意力机制旨在让模型学习更多重要信息, 而忽略无关信息。

首先, 定义注意力权重为:

$$W = (W^1, W^2, \dots, W^l) \quad (11)$$

式中: W^l 为第 l 个特征的权重。

然后对特征间的相对重要性进行设定:

$$\tilde{X}_i = (x_i^1 W^1, x_i^2 W^2, \dots, x_i^l W^l) \quad (12)$$

将设定的 $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_T)$ 作为输入变量放入 LSTM。则加入注意力机制的 LSTM 单元计算为:

$$g_{\text{forget}} = \text{Sigmoid}(W_{\text{fg}} \tilde{X}_i + W_{\text{hfg}} h_{t-1} + b_{\text{fg}}) \quad (13)$$

$$g_{\text{input}} = \text{Sigmoid}(W_{\text{ig}} \tilde{X}_i + W_{\text{hig}} h_{t-1} + b_{\text{ig}}) \quad (14)$$

$$g_{\text{output}} = \text{Sigmoid}(W_{\text{og}} \tilde{X}_i + W_{\text{hog}} h_{t-1} + b_{\text{og}}) \quad (15)$$

$$(C)_t = (C)_{t-1} \otimes (g_{\text{forget}})_t + (g_{\text{input}})_t \otimes (\tanh(W_C \tilde{X}_t + W_{hC} h_{t-1} + b_C)) \quad (16)$$

$$h_t = g_{\text{output}} \otimes \tanh((C)_{t-1}) \quad (17)$$

式中： W 为权重系数； b 为偏置系数； h_{t-1} 为隐藏层单元，通过公式（13）~（15）逐步实现遗忘门、输入门、输出门三个门的加权， $(C)_t$ 为加权后的存储单元； h_t 为隐藏单元的输出。

假定 t 为时间步长，则：

$$\tilde{y}^t = h^{t-1} \quad (18)$$

则最终输出结果可表示为：

$$\tilde{y}_T = (\tilde{y}^1, \tilde{y}^2, \dots, \tilde{y}^T) \quad (19)$$

本文通过 LSTM 对固定输入输出区间的能耗数据实现预测。同时，结合 Prophet 算法，在有限的数据集中动态调整输入输出区间，进一步优化机房能耗预测结果。

（6）空调调控最优解输出

本文采用遗传算法与 PID 控制器的组合方法来求解最优调控策略。遗传算法能够从全局范围内搜索出最优解，而 PID 控制器通过结合比例控制（P）、积分控制（I）和微分控制（D）三种控制方式，以得到最优的控制输出。这种组合方法通过不断监测系统的实际输出与期望输出之间的误差，并根据误差的大小来调整控制器的输出，从而实现柔性调控。具体而言，测点温度与合理值之间的差值被用作需要调控的空调温度的范围，关联性最大的空调被调整，以得到调控值和调控时长，进而实现对机房温度的精细化控制。

2.2 冷冻水系统及末端机房多点自动联控技术

本文深入考虑末端机房与中央制冷源之间的联动效应，并充分利用 PID 控制算法在温度控制等闭环系统中的优势^[8]，提出了一种冷冻水系统及末端机房多点自动联控技术。该技术能够充分考虑 IT 设备的散热需求、机房环境的变化以及机房空调的制冷输出。基于末端空调的冷量需求，系统动态调整冷源侧的运行参数，并自动下发执行节能策略。通过该技术，机房实现了末端机房制冷设备的启停控制、冷量调节，以及冷源侧制冷单元的组合优化和制冷量控制（包括频率和流速等多点联控）。最终，达到了冷机与末端联合调优、最大限度节能的目标。

3 总结

针对数据中心末端制冷冗余、冷源配置不合理、调控不及时等问题，本文提出了一套面向算力基础设施的 AI 能耗管控解决方案。首先，本文结合专家经验、数据挖掘、AI 自主学习三大引擎，构建了一个 AI 能耗管控平台，实现了对数

据中心制冷系统的实时动态监控与管理；其次，本文基于混合 AI 模型自动寻优策略，设计了末端空调按需自动调控技术和冷冻水系统及末端机房多点自动联控技术等关键技术；最后，通过 AI 能耗管控平台和关键技术的有效配合，有效解决了机房制冷过剩和温湿度控制难等问题，实现制冷量“精准够用”，平均制冷节能率提升了 20%，PUE 平均降幅了 0.13。下一步的工作是在此基础上持续优化算法模型，并不断迭代升级平台，从而提升模型性能和平台能力。

参考文献：

- [1] 工业和信息化部印发《新型数据中心发展三年行动计划（2021—2023 年）》[J]. 中国信息化, 2021(9):16-20.
- [2] 何璐.《信息通信行业绿色低碳发展行动计划（2022—2025 年）》发布[N]. 机电商报, 2022-09-19(A01).DOI:10.28408/n.cnki.njdsb.2022.000365.
- [3] 国家数据局. 全国数据资源调查报告（2023 年）[R/OL]. [2024-07-08]. https://zfsd.gd.gov.cn/xxfb/sjyw/content/post_4434806.html.
- [4] 杨玉军, 胡恺. 二级机房空调能耗智能管理系统[J]. 广播电视网络, 2021, 28(4):73-74.
- [5] 韩全磊, 洪金磊, 高川. 新一代算力中心运维支撑平台的研究与设计[J]. 信息技术与信息化, 2024(11):14-17.
- [6] BOUKERCHE A, ZHENG L N, ALFANDI O. Outlier detection: methods, models, and classification[J]. ACM computing surveys (CSUR), 2020, 53(3):1-37.
- [7] 闫永权. 基于注意力机制的云数据中心资源消耗预测[J/OL]. 武汉大学学报(理学版), 1-10[2024-09-13]. <https://doi.org/10.14188/j.1671-8836.2024.0031>.
- [8] DENG W, ZHANG X X, ZHOU Y Q, et al. An enhanced fast non-dominated solution sorting genetic algorithm for multi-objective problems[J]. Information sciences, 2022, 585: 441-453.

【作者简介】

燕艳（1983—），女，甘肃庆阳人，本科，中级工程师，研究方向：基础设施智能运维、绿色能源、数字孪生、大数据。

徐一伟（1977—），女，山东济南人，本科，中级工程师，研究方向：通信网络、智能运维、大数据、云计算。

宋德华（1978—），女，山东济南人，本科，中级工程师，研究方向：人工智能、大数据、智能运维、5G。

王艺皓（1995—），男，山东济南人，硕士，中级工程师，研究方向：人工智能、大数据、智能运维。

（收稿日期：2024-12-18 修回日期：2025-04-29）