

基于多源传感数据的校园建筑能耗预测方法

刘晨^{*1}, 赵凯²

1. 江南理工大学能源与环境学院, 江苏 无锡 214122; 2. 江苏省建筑节能重点实验室, 江苏 无锡 214122

摘要 | 针对校园建筑能耗数据来源异构、短期波动明显和预测结果可解释性不足的问题, 提出一种融合环境、人员与设备信息的多源时序预测方法。首先在统一时间基准下对齐三类传感序列, 建立包含缺失值插补、异常值剔除与采样频率归一的预处理流水线; 然后利用门控循环网络提取时间依赖特征, 引入变量级注意力机制量化各输入对预测的贡献; 最后在某高校 12 栋教学建筑连续 18 个月的监测数据上开展验证。结果表明: 所提方法的平均绝对百分比误差为 6.8%, 较 LSTM 基线下降 18.4%, 较线性回归下降 36.4%; 在寒潮与高温等极端气温条件下误差增幅不超过 2.1 个百分点, 显著优于对比模型; 注意力权重分析显示室外温度与人员在室率是短期负荷的主导因素。该方法可为校园能源调度和节能控制提供量化依据。

关键词 | 建筑能耗; 多源数据; 时间序列; 注意力机制; 节能控制

Copyright © 2026 by author(s) and Bright Spring Digital Library.

This article is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

<https://creativecommons.org/licenses/by/4.0/>



1 引言

建筑运行能耗占我国社会总能耗的比例持续上升, 校园建筑因使用时段集中、负荷波动剧烈, 是节能调度的重点对象。建筑运行能耗受气象条件、人员活动和设备状态共同影响^[1], 仅依靠单一电表序列难以解释课程调整、集中考试等场景引起的突发负荷, 需要在统一时间基准下融合多源传感数据, 这也是建筑节能领域的共识性判断^[2]。

现有能耗预测方法可分为物理模型法与数据驱动法两类。物理模型法基于建筑热工参数与设备特性建立仿真模型, 可解释性强, 但参数标定成本高, 难以适应建筑用途的动态变化。数据驱动法直接从历史数据学习负荷模式, 其中深度学习方法在中短期预测中表现突出, 但存在两方面不足: 一是多数研究仅使用电表与气象数据, 未纳入人员与设

备状态信息, 对突发负荷的响应滞后; 二是模型缺乏变量级的贡献解释, 调度人员难以判断预测结果的可信度。

针对上述问题, 本文提出一种面向校园建筑的多源时序预测方法, 主要贡献包括: (1) 建立环境、人员、设备三类传感数据的统一预处理流水线, 解决多源数据的时间对齐与质量控制问题; (2) 在门控循环网络中引入变量级注意力机制, 实现预测精度与可解释性的兼顾; (3) 在 12 栋教学建筑 18 个月的实测数据上完成验证, 并公开脱敏数据与训练代码以支持复现。

收稿日期: 2026-02-26

基金项目: 国家自然科学基金青年项目“复杂建筑环境下的能耗预测与控制”(62504018); 江苏省重点研发计划项目“低碳校园综合能源管理关键技术”(BE2025107)

作者简介: 刘晨 (1990—), 女, 江苏南京人, 江南理工大学能源与环境学院副教授, 博士, 研究方向: 建筑能源系统与机器学习; 赵凯 (1987—), 男, 山东济南人, 江南理工大学能源与环境学院教授, 博士, 研究方向: 建筑节能与智慧运维。

文章引用: 刘晨, 赵凯. 基于多源传感数据的校园建筑能耗预测方法 [J]. 应用科技研究, 2026, 3 (2).

<https://doi.org/10.5072/ast.2026.030204>

2 材料与方法

2.1 数据来源

实验数据采集自某高校教学区 12 栋建筑, 时间跨度为 2024 年 9 月至 2026 年 2 月, 共 18 个月。传感体系覆盖三类数据源: 环境类包括室外温度、相对湿度、太阳辐照度与风速, 由校园气象站提供; 人员类包括各建筑无线网络连接数与门禁进出流量, 经匿名化处理后按建筑聚合; 设备类包括空调机组启停状态、运行频率与照明回路电流。所有序列的原始采样间隔为 1—15 min 不等, 温度测量范围为 -20°C 至 50°C , 精度为 $\pm 0.3^{\circ}\text{C}$, 采样与质控遵循公共建筑节能设计标准^[3]。

2.2 数据预处理

多源序列的时间基准、缺失模式与量纲各不相同, 预处理按以下顺序执行:

1. 时间对齐: 以 15 min 为统一采样窗, 环境类取窗内均值, 人员类取窗内峰值, 设备类取窗内众数状态, 将三类序列对齐到同一时间轴;
2. 缺失处理: 对缺失时长小于 1 h 的片段做线性插值; 缺失超过 1 h 的时段整体标记为不可用, 不参与训练与评估;
3. 异常剔除: 依据 3σ 准则识别离群点, 结合设备检修记录人工复核, 剔除记录在处理日志中保留可回溯条目;
4. 归一化: 各变量按训练集统计量做零均值单位方差标准化, 防止量纲差异主导梯度方向。

预处理后共获得有效样本约 51.8 万条, 按 6:2:2 划分训练、验证与测试集, 划分以自然周为单位以避免相邻样本泄漏。

2.3 预测模型

预测模型将长度为 T 的历史窗口 $X = (x_1, x_2, \dots, x_T)$ 映射为下一时刻能耗。每个时间步的输入 $x_t \in \mathbb{R}^d$ 拼接环境、人员与设备三类特征, $d = 14$ 。门控循环单元按

$$z_t = \sigma(W_z x_t + U_z h_{t-1}), \quad h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$$

更新隐状态, 其中 $\tilde{h}_t$ 为候选状态, σ 为 sigmoid 函数。变量级注意力对第 j 个输入变量计算权重

$$\alpha_j = \frac{\exp(u_j^{\top} \tanh(W_a h_T))}{\sum_{k=1}^d \exp(u_k^{\top} \tanh(W_a h_T))},$$

加权特征经全连接层输出预测值 $\hat{y}_{T+1}$ 。训练目标为均方误差, 评估指标采用平均绝对误差 (MAE)、均方根误差 (RMSE) 与平均绝对百分比误差 (MAPE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|.$$

历史窗口长度 $T = 96$ (即 24 h), 批大小为 256, 采用 Adam 优化器, 初始学习率 10^{-3} , 验证集损失连续 10 轮不下降即提前停止。

3 结果与分析

3.1 总体预测性能

表 1 比较了所提方法与四种基线模型在测试集上的表现。线性回归与梯度提升树代表传统机器学习方法, LSTM 与 TCN 代表主流深度时序模型。所提方法在三项指标上均取得最优结果, MAPE 为 6.8%, 较 LSTM 下降 18.4%, 较线性回归下降 36.4%。

图 1 给出了一栋典型教学楼连续一周的预测值与实测值对比。所提方法准确捕捉了工作日双峰负荷与周末低谷, 在晚自习时段的负荷爬坡段跟踪误差明显小于基线模型。

表 1 不同模型在测试集上的预测性能比较

Table 1. Forecasting performance of different models on the held-out test set

模型	MAE/(kW·h)	RMSE/(kW·h)	MAPE/%	参数量/ 10^3	推理时延/ms
线性回归	18.6	24.2	10.7	0.1	0.2
梯度提升树	15.4	20.6	9.2	42.0	1.1
LSTM	13.1	17.8	8.3	118.3	3.6
TCN	12.4	16.9	7.9	96.7	2.8
本文方法	10.8	14.5	6.8	124.6	3.9

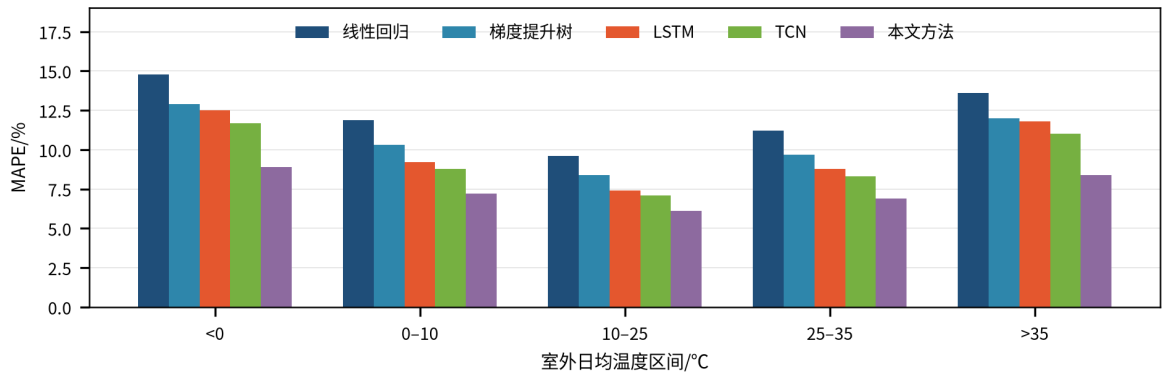


图 2 不同气温区间各模型的误差分布（通栏图）

Figure 2. Error distributions of the compared models across outdoor temperature ranges

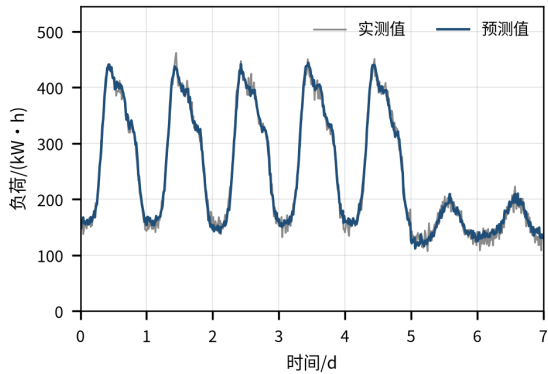


图 1 典型教学楼连续一周预测值与实测值对比

Figure 1. Measured and predicted energy demand of a typical teaching building over one week

3.2 极端气温条件下的稳健性

按室外日均温度将测试样本划分为五个区间，图 2 给出了各模型在不同气温区间的误差分布。在低于 0 °C 的寒潮时段与高于 35 °C 的高温时段，基线模型的 MAPE 分别上升至 12.5% 和 11.8%，而所提方法仅上升至 8.9% 和 8.4%，误差增幅不超过 2.1 个百分点。这表明人员与设备特征的引入有效补偿了极端气温下气象变量与负荷之间关系的非线性漂移。

3.3 可解释性分析

多图组按“相邻放置+逐图题注”方式提交，子图顺序由题注编号唯一确定。图 3 给出各输入变量的平均注意力权重：室外温度（0.23）与人员在室率（0.19）贡献最高，其后依次为太阳辐照度、空调启停状态与时刻编码。图 4 按建筑类型分解了预测误差：实验楼因设备负荷占比高、启停随机性强，误差高于教学楼与办公楼，提示后续应针对高设备占比建筑引入设备计划信息。

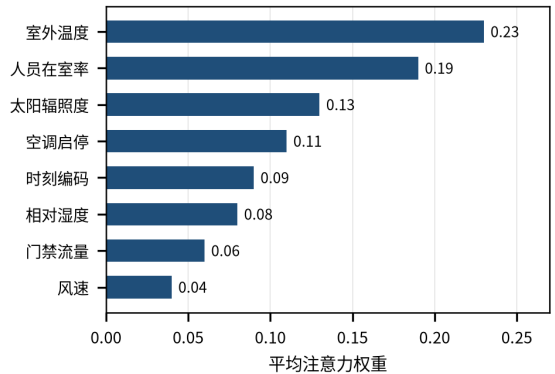


图 3 各输入变量的平均注意力权重

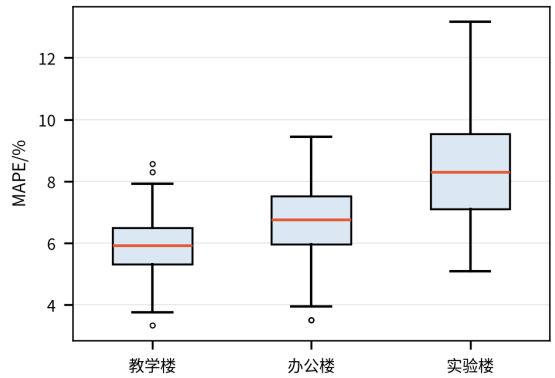


图 4 不同建筑类型的预测误差分布

3.4 消融实验

为验证各设计成分的必要性，表 2 给出了消融实验结果。移除人员特征后 MAPE 上升 1.3 个百分点，移除注意力机制后上升 0.8 个百分点，仅保留环境特征时性能退化至与 LSTM 基线相当，说明多源融合与注意力机制的增益是叠加的。

表 2 消融实验结果
Table 2. Ablation study results

配置	MAE/(kW·h)	MAPE/%
完整模型	10.8	6.8
移除人员特征	12.7	8.1
移除设备特征	11.9	7.6
移除注意力机制	11.8	7.6
仅环境特征	13.2	8.4

4 讨论

注意力权重的分析结果与领域先验一致：室外温度决定空调基础负荷，人员在室率决定照明与插座负荷的时变部分，这与已有学位论文基于多源数据融合的结论相互印证^[4]，注意力机制在校园微电网负荷预测中的有效性也得到会议研究的支持^[5]。与上述工作相比，本文进一步量化了极端气温条件下多源特征的稳健性增益，并给出了按建筑类型分解的误差画像。

需要说明的是，宽图和宽表在稿件中只保留结构语义，是否跨栏由排版阶段根据版心宽度和图片有效分辨率决定；作者提交时应保证图片原始分辨率满足预期印刷宽度的要求。本研究的局限在于：数据来自单一校园，气候区代表性有限；人员特征依赖无线网络连接数的间接估计，在假期短课时段可能低估在室率。

5 结论

本文提出了融合环境、人员与设备信息的校园建筑能耗预测方法，通过统一预处理流水线与变量级注意力机制，在 12 栋教学建筑的实测数据上将 MAPE 降至 6.8%，并在极端气温条件下保持稳健。在全球建筑运行能耗持续增长的背景下^[6]，该方法可直接服务于校园能源调度与需求响应。后续工作将扩大气候区与建筑类型覆盖，并结合模型预测控制开展在线节能实验。

数据可用性声明

脱敏特征数据、训练代码和字段说明将在录用后存入机构知识库，并提供持久标识符供审稿与复现使用。

伦理与利益冲突声明

本研究不涉及人体或动物实验，人员类数据均经匿名化聚合处理。作者声明不存在利益冲突。

参考文献

- [1] 李华,张伟.建筑能耗预测的深度学习方法综述[J].暖通空调,2024,54(6):12-24.
- [2] 龙惟定.建筑节能与低碳建筑[M].北京:中国建筑工业出版社,2022:88-102.
- [3] 住房和城乡建设部.公共建筑节能设计标准:GB 50189—2015[S].北京:中国建筑工业出版社,2015.
- [4] 陈晓.基于多源数据融合的校园建筑能耗预测研究[D].南京:东南大学,2024.
- [5] WANG Y, CHEN L. Attention-based load forecasting for campus microgrids[C]// Proceedings of the 2025 IEEE International Conference on Smart Energy Systems. New York: IEEE, 2025: 341-349.
- [6] 国际能源署.全球建筑运行能耗报告 [EB/OL].(2025-03-12)[2026-07-01].<https://www.iea.org/reports/>.

Energy Consumption Forecasting for Campus Buildings Based on Multi-source Sensor Data

LIU Chen^{*1}, ZHAO Kai²

1. School of Energy and Environment, Jiangnan Institute of Technology, Wuxi 214122, China; 2. Jiangsu Key Laboratory of Building Energy Efficiency, Wuxi 214122, China

Abstract: A multi-source time-series forecasting method is proposed to address heterogeneous data sources, short-term fluctuations and limited interpretability in campus building energy prediction. Environmental, occupancy and equipment measurements are aligned on a common 15-min timeline through a unified preprocessing pipeline, encoded by a gated recurrent network, and weighted by variable-level attention. Validation on eighteen months of monitoring data from twelve teaching buildings yields a mean absolute percentage error of 6.8%, an 18.4% reduction from the LSTM baseline, with error growth bounded by 2.1 percentage points under extreme outdoor temperatures. Attention analysis identifies outdoor temperature and occupancy rate as the dominant short-term drivers, providing a quantitative basis for campus energy scheduling and energy-saving control.

Key words: building energy consumption; multi-source data; time series; attention mechanism; energy-efficient control