

DOI: 10.11835/j.issn.2096-6717.2025.072



开放科学(资源服务)标识码OSID:



住宅空调能耗分形分析及短期预测模型优化

符程皓, 刘猛, 李子桥

(重庆大学土木工程学院; 国家级低碳绿色建筑国际联合研究中心; 绿色建筑与人居环境营造国际合作联合实验室, 重庆400045)

摘要:住宅建筑的空调系统能耗受用户行为多样性影响,且在数据质量不高的情况下,根据传统统计特征难以准确识别使用模式差异,使得短期能耗预测模型难以平衡准确性和计算成本。分形理论通过量化时间序列的非线性与多尺度复杂度,为识别用户行为差异提供了独特优势。提出一种结合分形分析与数据驱动方法的短期能耗预测框架,通过分形特征划分用户为低、中、高复杂度群体,从滑动窗口、数据量和特征选择3方面精选输入数据和特征,规避噪声干扰,并保留关键模式信息,以兼顾预测精度和计算成本。对夏季住宅空调运行数据研究发现,滑动窗口大小与复杂度负相关:低复杂度群体倾向较长窗口,中、高复杂度群体需较短窗口。数据量需求上,低、中复杂度需求较少,高复杂度群体需更全面数据。特征选择中,低、中复杂度群体依赖大幅波动特征,高复杂度群体依赖小幅波动特征。基于2000户样本验证,模型MAPE为9.82%,CV-RMSE为11.40%,相较于逐户预测模型,误差降低41.52%,并节省大量计算时间。

关键词:住宅建筑;空调能耗;分形分析;聚类分析;能耗预测;暖通空调

中图分类号:TU831

文献标志码:A

文章编号:2096-6717(XXXX)XX-0001-12

Residential air conditioning energy fractal analysis and short-term prediction model optimization

FU Chenghao, LIU Meng, LI Ziqiao

(School of Civil Engineering; National Centre for International Research of Low-carbon and Green Buildings; Joint International Research Laboratory of Green Buildings and Built Environments (Ministry of Education), Chongqing University, Chongqing 400045, P. R. China)

Abstract: The energy consumption of air conditioning systems in residential buildings is significantly influenced by the diversity of user behaviors. Traditional statistical features struggle to accurately discern differences in usage patterns under conditions of inadequate data quality, rendering short-term energy consumption prediction models challenged in balancing accuracy and computational cost. Fractal theory, however, offers a distinct

收稿日期:2025-03-22

基金项目:重庆市研究生科研创新项目(CYB240070);中央高校基本科研业务费专项(2024CDJYXTD-003、2023CDJKYJH102);

作者简介:符程皓(1999-),男,主要从事室内环境与大数据应用研究,E-mail:chenghaofu1999@163.com。

刘猛(通信作者),男,教授,博士生导师,E-mail:liumeng2033@126.com。

Received: 2025-03-22

Foundation items: Graduate Research and Innovation Foundation of Chongqing, China (No. CYB240070); the Fundamental Research Funds for the Central Universities (Nos. 2024CDJYXTD-003, 2023CDJKYJH102)

Author brief: FU Chenghao (1999-), main research interests: indoor environments and big data applications, E-mail: chenghaofu1999@163.com.

LIU Meng (corresponding author), professor, doctoral supervisor, E-mail: liumeng2033@126.com.

advantage in addressing these challenges by quantifying the nonlinear and multi-scale complexity of time series data, thereby enabling the differentiation of user behavior patterns. This study proposes a short-term energy consumption prediction framework that integrates fractal analysis with a data-driven approach. By employing fractal features, users are categorized into low, medium, and high complexity groups, with input data and features optimized across sliding window size, data volume, and feature selection to mitigate noise interference, preserve key patterns, and balance prediction accuracy with computational efficiency. Based on summer residential air conditioning data, the study finds that sliding window size is negatively correlated with complexity: low-complexity groups prefer longer windows, while medium- and high-complexity groups require shorter ones. For data volume, low- and medium-complexity groups need less, whereas high-complexity groups demand more comprehensive data. In feature selection, low- and medium-complexity groups rely on large-fluctuation features, while high-complexity groups depend on small-fluctuation features. Validation with a 2000-household sample demonstrates a MAPE of 9.82% and a CV-RMSE of 11.40%, achieving a 41.52% error reduction compared to per-household prediction models while significantly saving computational time.

Keywords: residential buildings; energy consumption of air conditioning systems; fractal analysis; cluster analysis; energy prediction; HVAC

建筑行业占全球最终能源消耗的30%^[1-2]。在建筑领域中,供暖、通风和空调(HVAC)系统占据了能耗的显著比例^[3-5],尤其在住宅建筑中,HVAC能耗受用户行为多样性和生活习惯的影响显著^[6-7]。因此,优化住宅建筑的HVAC能耗,特别是通过短期预测,即针对未来一天至一个月的能耗变化(以小时或日尺度为主),支持实时运行调整和节能策略制定,已成为一项关键任务。数据驱动模型因在预测HVAC能耗、支持系统运行优化和能源管理策略制定方面的潜力而被广泛应用,特别是短期能耗预测模型的实效性。例如,薛凯^[8]通过多维度数据识别空调使用群体特征实现短期能耗预测。晏璐^[9]基于空调大数据量化了夏热冬冷地区家用空调的区域用能特性。然而,这些模型的性能高度依赖于输入数据的质量和特征选择的合理性^[10-12]。在住宅建筑中,受限于设备条件和用户行为的多样性,获取高质量数据并确保模型的高效性与实用性面临显著挑战^[13]。传统预测方法多依赖统计特征或机器学习模型(如支持向量机、深度学习)^[12, 14-19],虽在一定程度上提高了预测精度,但往往忽视了计算效率,且对数据质量的依赖性较高,限制了其在资源受限环境中的应用。因此,亟需开发能够在提升预测精度的同时兼顾计算成本的方法。

分形理论作为一种研究自相似模式和跨尺度不规则结构的数学工具^[20-22],能够度量时间序列复杂性^[23-24],为分析非线性时间序列数据提供了有效框架。分形特征因其能够捕捉复杂系统的多尺度动态特性,在多个领域展现出显著应用价值,例如,地质地形分析^[25-27]、气象模式研究^[28, 29]、金融市场波动性分析^[30-32]、电力预测和异常诊断^[33-35]以及医学信

号处理^[36-37]。在能源消耗预测领域,分形特征在表征能耗数据的非线性与多尺度特性方面具有独特优势。相较于传统统计方法或常规机器学习技术,分形分析不仅能提供更细致的动态信息,还能生成可解释的指标,为优化预测模型提供了新的可能性。

尽管如此,分形分析在建筑能源系统中的应用仍处于探索阶段。已有研究表明了其潜力,例如,Mughees等^[38]提出了一种结合单分形与多分形分析的特征提取方法,用于识别低功率电器的能耗特性;Yang等^[39]基于分形相关维数(FCD)开发了空调系统故障检测技术,提升了检测精度;祝用华等^[40]利用关联维数对空调系统进行能耗分析;Knowles等^[41]通过多分形去趋势波动分析(MFDFA),揭示了住宅建筑能耗的多尺度特性。然而,这些研究多集中于特征识别或故障检测,鲜有将分形分析用于基于用户复杂度差异优化能耗预测模型的尝试,尤其是在平衡预测准确性与计算成本方面。

笔者通过住宅空调能耗分形分析,量化用户行为的复杂度,将用户划分为不同群体,并针对各群体优化能耗预测模型。研究重点在于根据复杂度划分优化滑动窗口、数据量选择、特征提取等关键参数,以在提升预测精度的同时降低计算成本。相较于传统方法,本研究通过分形特征捕捉用户行为的非线性与多尺度特性,结合针对性的模型优化策略,不仅克服了数据质量不足的局限,还在准确性与效率之间实现了平衡。

1 数据准备

数据来源主要分为两部分:1)大数据平台;2)

气象数据平台。其中大数据平台的优势在于获取的数据多样,能够反映用户的实际使用习惯、设备运行状态等。

1.1 数据基本信息

通过中国某知名空调企业物联网系统的大数据平台获取了住宅建筑中空调使用数据,数据内容包括:Device(设备ID)、Stamp(年/月/日/秒)、故障标志、设定温度(℃)、运行状态(开机/关机)、总耗电量(kW·h)和城市。该大数据平台在上传数据后进行了脱敏,所以设备ID并不能获知用户的建筑维护结构和具体地理位置信息,仅作为用户的识别标签,不涉及泄露用户隐私的问题。

获取的空调原始数据文件格式均为“txt”格式的文件,与本研究相关的具体数据格式和参数描述如表1和表2所示。对于空调数据还需作如下补充:1)获取的空调设备均为空气源热泵和壁挂式空调;2)数据每行代表一个时间步,平均时间步长为3 min,反映的空调操作包括设备启停、设定温度调节;3)根据长江流域城市分布,筛选重庆、武汉和上海3座典型城市的数据;4)选取的数据时间范围为2021年6月1日至2021年9月1日(夏季)。通过数据筛选和预处理后,共得到了1 143台空调设备的运行数据。

采用的室外气象数据均来自气象数据平台 <https://zh.weatherspark.com/>,获取的数据格式和参数如表3所示。其中,气象数据的时间范围和空调设备数据保持一致,时间步长为1 h。

表1 空调数据源文件数据格式

Table 1 Data format of the air conditioner data source file

Device	Stamp	故障标志	设定温度	运行状态	总耗电量	城市
XXX	yyyy/MM/dd/HH/mm/ss	0/1	XX	0/1	XXX	City

表2 空调数据参数描述

Table 2 Description of air conditioner data parameters

参数	单位/数据形式	数据描述
Device		空调设备ID
Stamp	yyyy/MM/dd/HH/mm/ss	平均时间步长为3 min
故障标志	0/1	故障/正常
设定温度	℃	空调设定温度
运行状态	0/1	停机/开机
总耗电量	kW·h	逐时累积耗电量
城市		重庆/武汉/上海

表3 气象参数数据格式

Table 3 Data format of meteorological parameters

城市编号	年	月	日	小时	温度/℃
XXX	yyyy	MM	dd	HH	XX

注:城市编号为气象数据平台设定的城市标签。

1.2 数据预处理

由于获取的原始空调使用数据和原始气象参数数据存在数据格式不统一、数据质量层次不齐和时间粒度不一致等问题,需要对原始数据进行数据预处理和格式整合。预处理流程如下:

1)按照设备ID将原始数据拆分成相同长度的时间序列csv文件;

2)对每个csv文件的缺失值进行线性插值处理;

3)进行一阶差分处理生成单位时间步长的耗电量;

4)识别并标记负向阶跃的时间点,一阶差分时间序列中标记点会产生负值;

5)通过3倍标准差法则识别异常点;

6)对识别出的负向阶跃点和异常值进行线性插值处理;

7)根据式(1)对一阶差分时间序列进行1 h重采样,获取每台设备2 232个时间步长的耗电量数据。这样处理后的数据集具有极高的时间序列连续性和完整性,远远超过了50%的最低水平,时间序列的缩放指数和缩放行为不会受到严重影响^[42-43],且实现了空调设备运行数据和气象数据时间步长的统一。

$$E_i = \frac{(E_{\max} - E_{\min})}{(t_{\max} - t_{\min})} \quad (1)$$

式中: E_i 为第*i*小时内的耗电量; $t_{\max}$ 为第*i*小时内的最大时刻; $t_{\min}$ 为第*i*小时内的最小时刻; $E_{\max}$ 为 $t_{\max}$ 对应的耗电量; $E_{\min}$ 为 $t_{\min}$ 对应的耗电量。

2 研究方法

研究的技术路线如图1所示。首先,对预处理后的空调能耗时间序列进行多重去趋势波动分析,提取多重分形特征(谱宽 Δh_q 、谱高 ΔD_q 、截断类型和不对称系数*B*);其次,基于分形特征进行K-means聚类,划分不同复杂度的空调使用用户群体;最后,基于不同复杂度群体构建能耗预测模型优化方案,并分别建立个性化短期能耗预测模型。

2.1 形理论

多重分形去趋势波动分析(MF-DFA)是一种用于分析非平稳时间序列中多重分形特性的工具,通过估计时间序列的多重分形谱来识别时间序列中具有不同波动特性的局部区域,从而识别时间序列的复杂度和自相似性(复杂度与自相似性成反比关系)。采用Ihlen^[44]的MF-DFA计算方法,其中,奇异性指数 h_q 和奇异性维度 D_q 的符号和其他文献^[45]存在不同,但在计算意义上是相同的。通常,多重

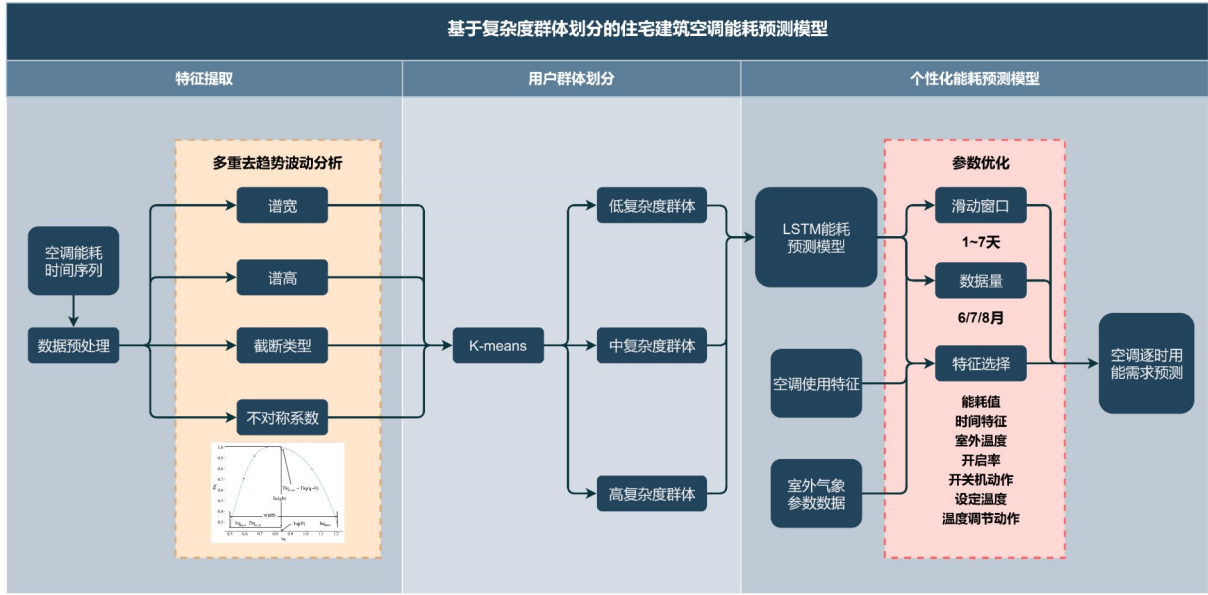


图1 技术路线图

Fig. 1 Technology roadmap

去趋势波动分析得出的多重分形谱具有倒置的抛物线形状,如图2所示,主要由4个特征参数来描述:1)谱宽 $\Delta h_q = h_{q_{\max}} - h_{q_{\min}}$; 2)谱高 $\Delta D_q = D_{q_{\max}} - D_{q_{\min}}$; 3)截断类型; 4)不对称系数 B 。截断类型可以分为对称的、左截断或右截断,采用以下划分规则^[28]:

- 1)L:左截断。
- 2)LL:左侧高度截断。(左侧截断位置超过中间点)
- 3)R:右截断
- 4)S:无截断或对称的。(左右两侧没有明显截断)

其中,左截断类型时间序列的多重分形结构对大幅度的局部波动不敏感,右截断类型时间序列对小幅度的局部波动不敏感,对称类型时间序列对大小幅度波动表现出相同的敏感度^[28]。

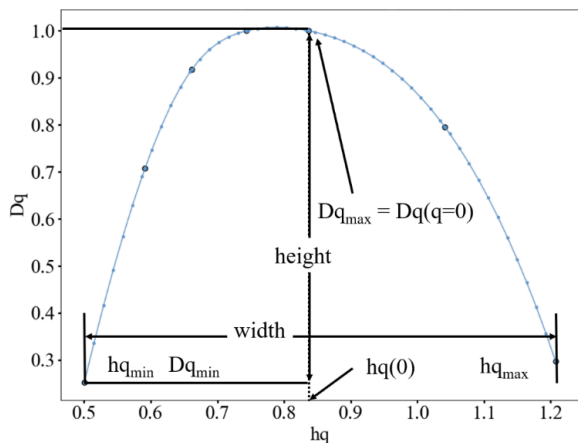


图2 多重分形谱及其主要特征示意图

Fig. 2 Schematic diagram of multifractal spectrum and its main features

$$G(h_q) = A(h_q - h_q(0))^2 + B(h_q - h_q(0)) + C(2)$$

对多重分形谱进行二阶方程拟合,得到方程(2)。其中,在3个常数 A 、 B 、 C 中,最重要的是不对称参数 B ,用来描述多重分形谱的偏移情况。当 $B = 0$ 时,光谱的形状是对称的;当 B 为正值时,光谱是右偏的;当 B 为负值时,光谱是左偏的。右偏斜谱的时间序列更加聚焦于高分形指数,左偏斜谱的时间序列更聚焦于低分形指数。具有高谱高、宽谱宽和右偏态的多重分形谱被认为比具有相反特征的时间序列更“复杂”^[44]。

2.2 K-means 聚类

由于分形特征能反映系统内的复杂性和自相关性,基于多重分形特征,采用 K-means 聚类方法来识别住宅建筑空调用户群体。采用“Z-score”标准化方法对聚类参数标准化,采用 DBSCAN 方法识别出离群点^[46],根据“Elbow Method”确定最佳 k 值,选择 SSE 和轮廓系数作为聚类效果的评价指标。

2.3 LSTM 神经网络模型

长短期记忆网络(LSTM)是一种特殊的循环神经网络(RNN)架构^[47]。LSTM 网络被设计用来克服传统 RNN 在处理长序列数据时所面临的梯度消失和梯度爆炸问题,从而有效地捕捉时间序列数据中的长期依赖关系。LSTM 网络的核心组件是 LSTM 单元,如图3所示,它通过由遗忘门、输入门和输出门组成的门控机制来控制信息的流动和更新。

LSTM 神经网络模型的具体步骤和参数设置如下:

- 1)确定输入和输出参数的形状,并对其进行归

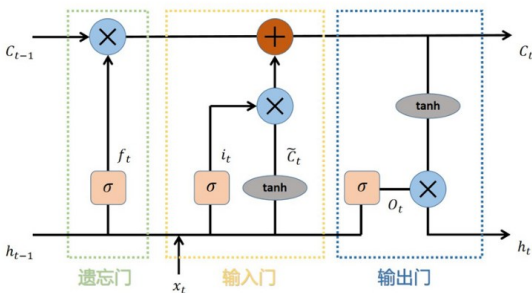


图3 LSTM循环单元结构

Fig. 3 LSTM recurrent unit structure

一化处理;

2)将输入数据集按照60%、20%和20%比例划分为训练集、验证集和测试集;

3)通过Python语言的Skearn库和TensorFlow库构建LSTM神经网络模型和参数设置,并选择网格搜索法进行超参数优化;

4)训练模型,输出预测的短期能耗时间序列

5)通过评估指标来评价模型,本研究选择的评估指标包括准确性指标(MAPE、CV-RMSE和拟合优度 R^2)和训练速度指标(平均训练时长)。

2.4 模型优化策略

在能耗预测模型中,滑动窗口、数据量和特征数量是影响预测准确性和计算成本的核心因素。空调运行特征为模型提供丰富信息,但滑动窗口、特征和数据量并非越多越好。增加窗口、数据量和特征数量会提高计算成本吗,甚至引入噪声或冗余,导致模型鲁棒性下降、预测准确性降低,甚至引发“维度灾难”和过拟合风险。因此,针对不同复杂度群体优化滑动窗口、数据量和特征选择,以平衡预测准确性和计算成本。

2.4.1 滑动窗口

滑动窗口用于从历史能耗时间序列中提取固定长度的数据段,以1 h步长滑动,生成含能耗值、时间特征等的输入向量,以捕捉能耗的动态变化。为了捕捉日周期性的能耗模式,按照1 d的整数倍选取如表4所示的滑动窗口范围分别训练能耗预测,通过对比滑动窗口大小下模型的准确性指标和训练速度指标,旨在为不同复杂度群体选择合适的滑动窗口大小选择提供参考依据。滑动窗口的大小反映了用于预测未来24 h能耗的时间范围,不同窗口大小的设置导致的预测模型评价指标差异能够反映用户使用模式变化情况。

2.4.2 数据量

将整个夏季时间序列按照月份拆分为3个数据集(6月、7月和8月),通过对比不同数据量输入下

表4 滑动窗口选取范围

Table 4 Sliding window selection range	
滑动窗口/h	天数
24	1
48	2
72	3
96	4
120	5
144	6
168	7

模型的准确性指标和训练速度指标,旨在为不同复杂度群体选择合适的数据输入量提供参考依据。

2.4.3 特征选择

选取如表5所示的特征参数分别训练能耗预测模型,通过对比不同输入特征下模型的准确性指标和训练速度指标,并结合shap方法分析各个特征对模型的贡献度,旨在为不同复杂度群体选择合适的特征参数提供参考依据。

表5 基于空调运行数据提取的特征参数

Table 5 Feature parameters extracted from air conditioning operation data	
特征参数	描述
能耗值	逐时空调耗电量
时间特征	年/月/日/周
室外温度	逐时室外温度
开启率	同时刻运行概率
开关机动作	同时间开关机概率
设定温度	同时刻设定温度均值
温度调节动作	同时刻调节温度概率

注:1.开关机动作反映时间序列大幅波动;2.温度调节动作反映时间序列小幅波动。

3 用户群体识别

如图4所示,通过分形特征Spearman相关性分析发现,谱宽和谱高这两个多重分形谱的关键参

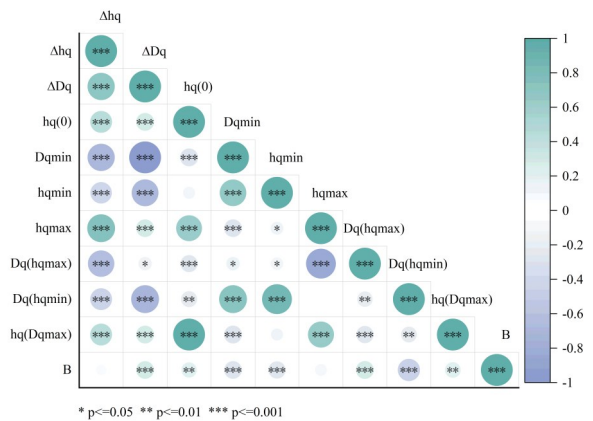


图4 分形特征相关性分析

Fig. 4 Correlation analysis of fractal features

数,不仅彼此之间呈现极强的正相关关系($\rho \rightarrow 1$, $p < 0.001$),而且与其他多重分形特征参数均存在显著的相关性($p < 0.05$ 或更低)。因此,将谱宽和谱高作为分形特征的综合性量化指标。

基于谱宽和谱高进行K-means聚类,通过肘部图(图5)确定最佳簇数 $K=3$,并在表6中以0、1、2标签标注。3类群体的数值范围对比显示其多重分

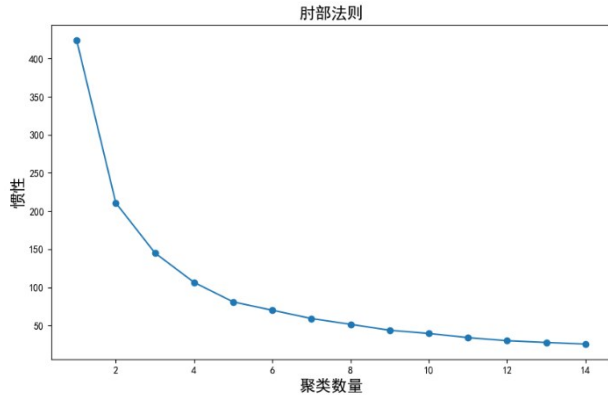


图5 肘部图

Fig. 5 Elbow plot

形特征差异:第0类谱宽和谱高最小,多重分形特征较弱,能耗时间序列波动性和复杂度较低,自相似性高;第1类分形特征增强,复杂度升高,自相似性降低;第2类谱宽和谱高最大,多重分形特征最强,时间序列复杂度最高,自相似性最低。由于重庆市、武汉市和上海市3座城市都属于长江流域夏热冬冷地区,存在气候和空调使用相似性,在后续的能耗预测模型参数优化中进行合并处理。

表6 分形特征聚类分析概述

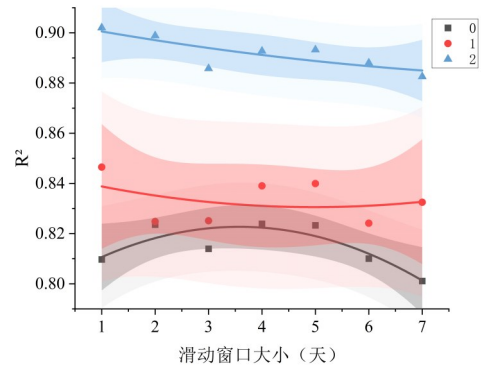
Table 6 Overview of fractal feature clustering analysis

城市	分形特征	用户数量	谱宽 Δh_q	谱高 ΔD_q
重庆市	0	49	0.07~0.55	0.10~1.073
	1	96	0.29~1.01	0.63~1.73
	2	67	0.41~1.43	1.02~2.34
武汉市	0	138	0.07~0.91	0.28~1.43
	1	157	0.50~1.35	0.84~1.97
	2	32	1.00~1.80	1.41~2.61
上海市	0	224	0.23~1.40	0.06~0.66
	1	292	0.33~1.08	0.65~2.05
	2	79	0.63~1.44	1.08~2.41

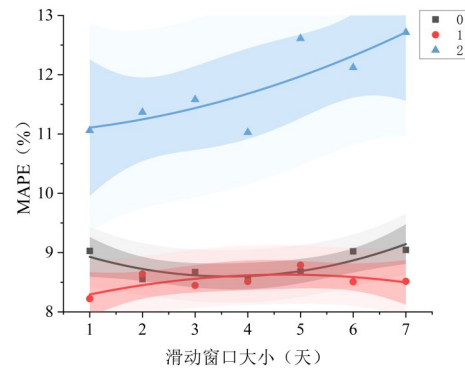
4 空调能耗预测模型优化

4.1 滑动窗口分析

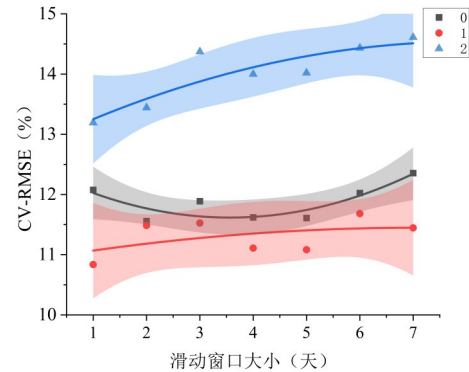
用 R^2 评估模型鲁棒性,用MAPE和CV-RMSE比较模型误差,用平均训练时长衡量计算成本。多滑动窗口大小的能耗预测评价指标如图6所示,随着滑动窗口增大,低复杂度0类用户误差先减后增,



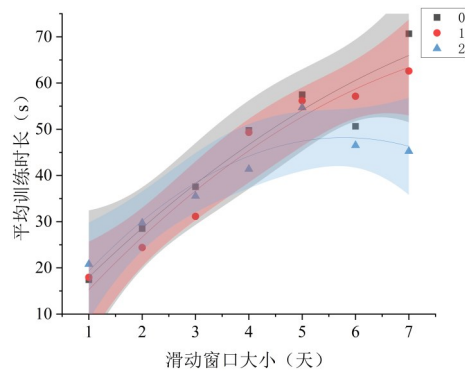
(a) 拟合优度评价



(b) 平均绝对百分比误差



(c) 变异系数均方根误差



(d) 平均训练时长

图6 多滑动窗口大小能耗预测评价指标

Fig. 6 Energy Consumption Prediction Evaluation Metrics Across Multiple Sliding Window Sizes

鲁棒性先升后降,训练时长最大且持续增加,原因是其长程相关性强,使用模式稳定,适度增大窗口(2~4 d)能提升准确性,但过大窗口导致过拟合,降低测试集表现。高复杂度2类用户误差持续增大,鲁棒性持续下降,训练时长增长最慢,因使用模式不稳定,噪声数据干扰预测,适配最小窗口(1 d)。中复杂度1类用户适配较小窗口(1~2 d)。表7总结了不同复杂度群体的滑动窗口选择标准。

表 7 滑动窗口大小选择标准
Table 7 Selection criteria for sliding window sizes

用户群体复杂度	滑动窗口大小/h
低	48、72、96
中	24、48
高	24

注:用户群体复杂度是根据样本聚类分析得出的相对标签。

4.2 数据量分析

数据量选择需考虑能耗时间序列的季节性趋势和长程相关性。不同数据量的能耗预测评估指标如图7所示, R^2 方面,除0类用户7月为正值外,其余群体和月份均为负值,表明数据量减少后模型鲁棒性下降。0类用户差值较小,因其高自相似性,模式一致性强;2类用户差值最大,因自相似性低且使用模式多变,6月、7月、8月分别对应夏季初、盛夏和末期,季节周期差异显著。中复杂度1类用户分形特征和指标差异居中。误差评估显示,高复杂度2类用户对季节周期更敏感,误差增幅大;低复杂度0类用户误差增幅最小。值得注意的是,2类用户7月误差低于夏季整体,因盛夏高温下使用模式更稳定。训练时长方面,数据量减少后各月训练时长显著降低,7月降幅最大,因高温导致使用模式更稳定。表8总结了数据量选择标准。

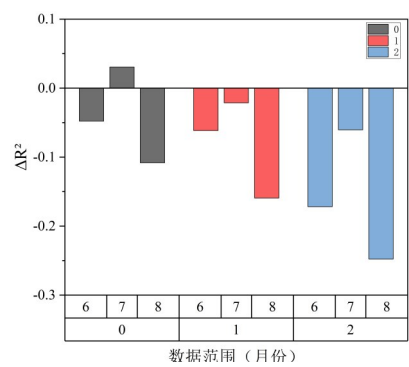
表 8 基于复杂度划分的数据量选择标准
Table 8 Selection criteria for data volume based on complexity classification

复杂度	数据量	数据特点描述
低	少(1个月)	具有代表性的环境因素稳定时段内数据集
中	中(2-3个月)	环境因素稳定和部分非稳定时段内数据集
高	多(3个月)	完整数据集

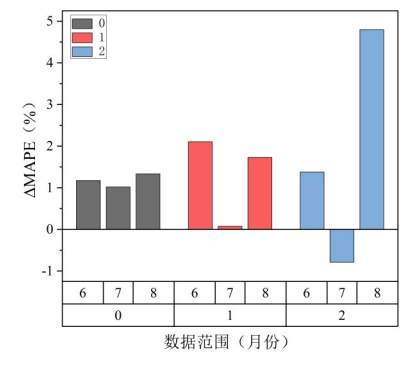
注:1.用户群体复杂度是根据样本数据聚类分析得出的相对标签;2.数据量给出的具体大小仅针对本研究样本。

4.3 特征选择分析

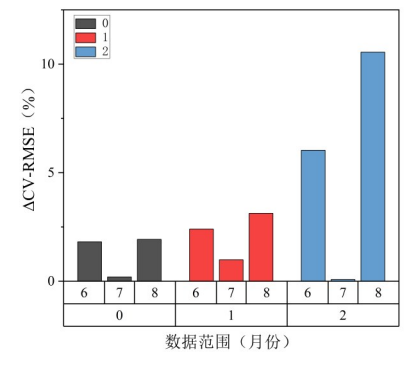
不同特征组合的能耗预测评价指标如图8所示,低复杂度群体中,输入特征增多时, R^2 略升,MAPE和CV-RMSE小幅下降,预测准确性提高,但训练时间显著增加。中复杂度群体和低复杂度



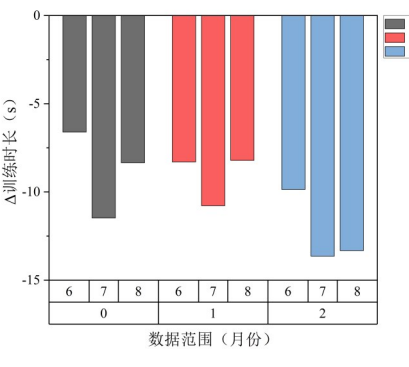
(a) 拟合优度评价



(b) 平均绝对百分比误差



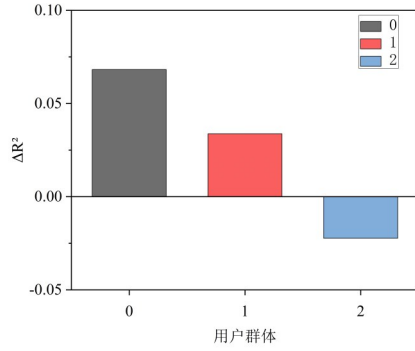
(c) 变异系数均方根误差



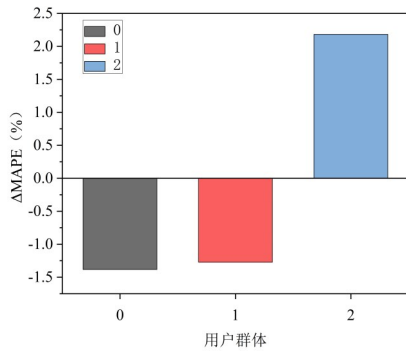
(d) 平均训练时长

图 7 不同数据量的能耗预测评价指标(评价指标为每个月与整个夏季结果的差值)

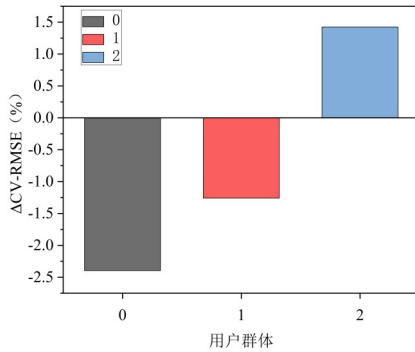
Fig. 7 Energy consumption prediction evaluation metrics across different data volumes (evaluation metrics represent the difference between monthly results and the entire summer results)



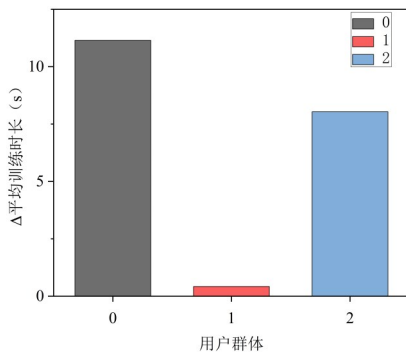
(a) 拟合优度评价



(b) 平均绝对百分比误差



(c) 变异系数均方根误差



(d) 平均训练时长

图8 不同特征组合的能耗预测评价指标(评价指标为每个月与整个夏季结果的差值)

Fig. 8 Energy consumption prediction evaluation metrics across different feature combinations (evaluation metrics represent the difference between monthly results and the entire summer results)

群体变化一致,仅在训练时长上的差异变小。高复杂度群体中,输入特征增多时, R^2 提升,MAPE和CV-RMSE增大,预测准确性降低,训练时间显著增加。综上,特征数量影响因群体复杂度而异:低复杂度群体少量特征即可,过多的特征输入会显著增加训练时长;中复杂度群体选择较全面特征,在训练时长增幅不大的前提下提高准确性。高复杂度群体则需要精选特征,避免噪声特征的输入导致预测准确性降低。

采用SHAP分析方法生成图9特征贡献度排序结果,分析各特征对能耗预测模型的贡献。3类用户群体贡献度较高的均为开启率、设定温度、时间特征,作为基准特征。但开关机动作和室外温度在不同复杂度群体中差异明显。低、中复杂度(0、1类)用户多重分形特征以左截断为主,对大幅波动不敏感,加入开关机动作特征可提升预测准确性;高复杂度(2类)用户多重分形特征以右截断为主,对小幅波动不敏感,加入室外温度特征可增强预测精度。总结得出表9所示特征选择参考标准。

表9 基于复杂度划分的特征选择标准

Table 9 Selection criteria for features based on complexity classification

复杂度	特征数量	特征描述
低	少	少量特征,聚焦反映大幅波动的特征
中	多	全面特征,聚焦反映大幅波动的特征
高	少	规避噪声特征,聚焦反映小幅波动的特征

注:用户群体复杂度是根据样本数据聚类分析得出的相对标签。

4.4 区域尺度短期空调能耗预测:以区域住宅建筑为例

通过由不同复杂度比例构成的2000户样本进行区域案例分析,以验证前述基于复杂度的个性化能耗预测模型的性能,样本构成和对于参数设置如表10所示。

提出的模型与逐户预测模型的对比结果如表11所示,预测结果MAPE为9.82%相对较小,CV-RMSE为11.40%符合ASHRAE手册对小时尺度能耗预测模型要求的30%以内误差,测试集与训练集和验证集的 R^2 差异较小且值较高,表明模型拟合较好且未出现过拟合现象。相较于逐户预测模型,每户的平均训练时长为20.45s,2000户总训练时长约为40900s,提出的模型平均训练时长为39.90s,并将CV-RMSE降低41.19%,展现出更高的预测精度和计算效率。图10给出了夏季末连续一周的逐时能耗预测值和真实值对比,结果表明,很好地预测了真实数据中的趋势和波动特征。逐时能耗分布于130 kW·h至650 kW·h之间。

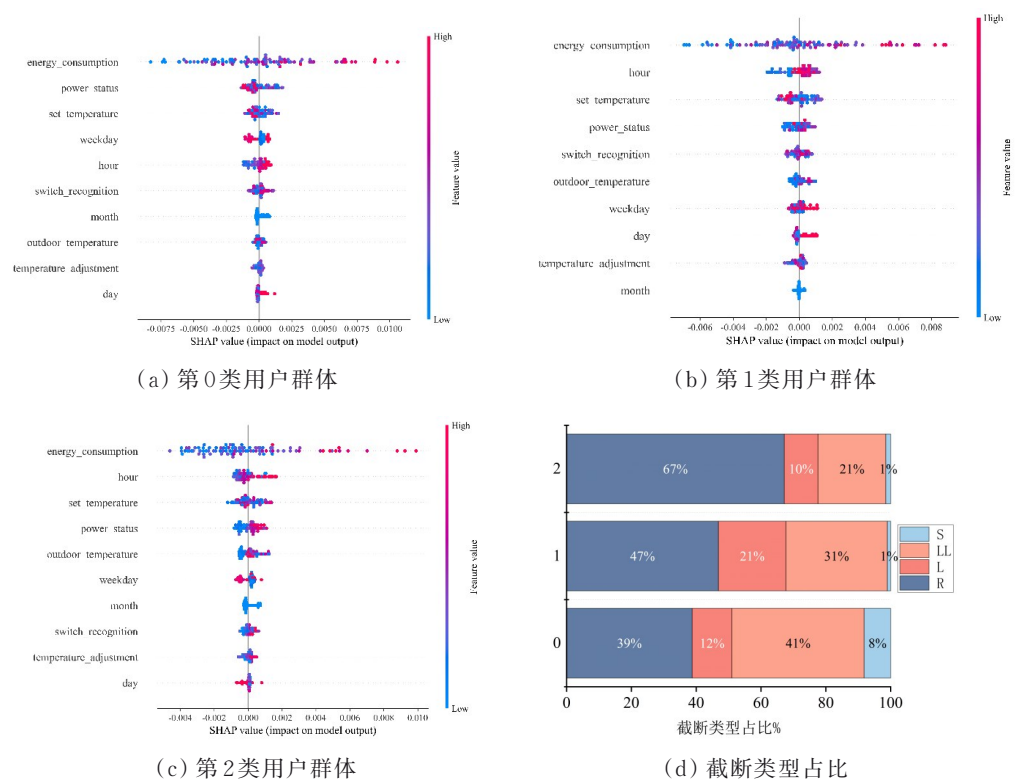


图9 Shap分析结果可视化图和截断类型占比

Fig. 9 Visualization of SHAP analysis results and proportion of truncation types

表 10 区域案例样本构成和参数设置

Table 10 Composition of regional case samples and parameter settings

复杂度	用户占比/%	滑动窗口/h	数据量	输入特征
低	40	48	7月1日—7月31日	开启率、设定温度、时间特征、开关机动作
中	50	24	6月1日—9月1日	开启率、设定温度、时间特征、开关机动作
高	10	24	6月1日—9月1日	全特征

表 11 案例分析结果

Table 11 Case analysis results

模型	数据集	R^2	MAPE/%	CV-RMSE/%	平均训练时长/s
复杂度分类模型	训练集	0.94	10.22	14.21	39.90
	验证集	0.93	9.95	11.79	
	测试集	0.90	9.82	11.40	
逐户预测模型	训练集	0.43	42.20	50.60	20.45
	验证集	0.38	43.02	46.66	
	测试集	0.21	47.06	52.59	

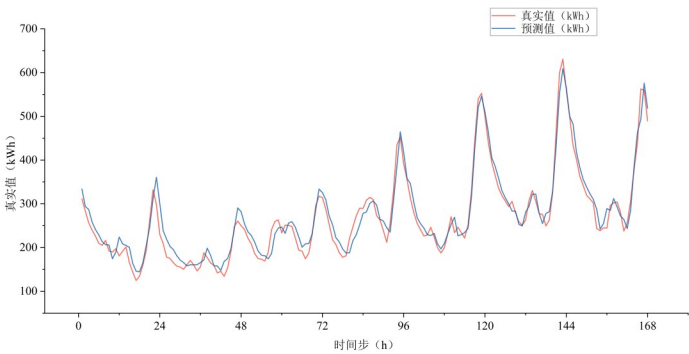


图 10 一周内逐时能耗预测值和真实值对比

Fig. 10 Comparison of hourly energy consumption predictions and actual values over a week

5 结论

基于数据驱动方法,利用住宅建筑空调运行数据,结合分形分析和聚类分析,构建了区域尺度短期能耗预测模型,并针对不同复杂度用户群体设计个性化设置方案,得出以下结论:

1)提出了一种基于分形特征的K-means聚类与LSTM网络相结合的能耗预测框架。通过分形特征有效识别出低、中、高复杂度住宅空调用户群体,并针对不同群体,优化了滑动窗口、数据量及特征选择。相较于采用多维度特征实现用户群体分类,分形理论通过降维处理,仅从能耗数据中提取谱宽和谱高特征,就能量化能耗复杂度从而有效区分用户群体。

2)基于群体复杂度,从滑动窗口、数据量和特征选择3方面精选输入数据和特征,规避噪声干扰,同时保留关键模式信息,有效地兼顾了预测模型的准确性和计算成本。滑动窗口方面,低复杂度群体适宜2~4 d窗口以捕捉长程相关性,中、高复杂度群体适配1~2 d窗口以减少噪声干扰;数据量选择上,低复杂度群体因高自相似性对数据量变化不敏感,高复杂度群体需根据季节周期(如7月盛夏)调整数据量;特征选择中,SHAP分析显示低、中复杂度群体受益于大幅波动特征(如开关机动作),高复杂度群体依赖小幅波动特征(如室外温度),时间特征、设定温度和开启率为通用基准特征。

3)模型性能验证表明,其在区域尺度能耗预测中具有优异表现。基于2021年夏季2000户样本数据,模型预测MAPE为9.82%,CV-RMSE为11.40%,满足ASHRAE小时尺度误差要求,测试集与训练集 R^2 差异小,展现出良好的拟合与泛化能力,能有效捕捉能耗趋势与波动。与逐户预测模型相比,CV-RMSE降低了41.19%,训练时长得到大幅降低。

本研究基于空调大数据,利用分形理论优化住宅空调短期能耗预测,取得一定进展。未来可进一步整合多样化数据特征(如热工参数)并扩展至非住宅建筑,深化其应用研究。

参考文献

[1] CAO X D, DAI X L, LIU J J. Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade [J]. *Energy and Buildings*, 2016, 128: 198-213.

[2] DEB C, LEE S E. Determining key variables influencing energy consumption in office buildings through cluster analysis of pre- and post-retrofit building data [J].

Energy and Buildings, 2018, 159: 228-245.

- [3] KHAN A N, KHAN Q W, LEE J, et al. Dynamic temporal analysis and modeling of residential lighting consumption for energy efficiency and sustainability [J]. *IEEE Access*, 2024, 12: 154365-154380.
- [4] PAPADOPOULOS S, KONTOKOSTA C E, VLA-CHOKOSTAS A, et al. Rethinking HVAC temperature setpoints in commercial buildings: The potential for zero-cost energy savings and comfort improvement in different climates [J]. *Building and Environment*, 2019, 155: 350-359.
- [5] PÉREZ-LOMBARD L, ORTIZ J, POUT C. A review on buildings energy consumption information [J]. *Energy and Buildings*, 2008, 40(3): 394-398.
- [6] DAI L K, LI Z Q, CHEN X Y, et al. Usage behavior characteristics of household air-conditioners during the extremely hot summer: A case study of Chongqing [J]. *Building and Environment*, 2023, 234: 110160.
- [7] DE WILDE P, ALY D, CHO S, et al. Occupant behavioural freedom in building energy use [J]. *Applied Energy*, 2025, 377: 124682.
- [8] 薛凯.住宅房间空调器使用群体特征及其用能预测方法研究[D].重庆:重庆大学,2022.
- XUE K. Characteristic of residential room air conditioners usage groups and its energy demand prediction method [D]. Chongqing: Chongqing University, 2022. (in Chinese)
- [9] 晏璐.夏热冬冷地区分散式家用空调供冷的区域用能特性量化研究[D].重庆:重庆大学,2022.
- YAN L. The quantitative study of district energy use characteristics for distributed household air conditioners' cooling in hot summer and cold winter zone [D]. Chongqing: Chongqing University, 2022. (in Chinese)
- [10] ALA' RAJ M, RADI M, ABBOD M F, et al. Data-driven based HVAC optimisation approaches: A systematic literature review [J]. *Journal of Building Engineering*, 2022, 46: 103678.
- [11] XIAO Z W, GANG W J, YUAN J Q, et al. Impacts of data preprocessing and selection on energy consumption prediction model of HVAC systems based on deep learning [J]. *Energy and Buildings*, 2022, 258: 111832.
- [12] SHA H J, XU P, HU C H, et al. A simplified HVAC energy prediction method based on degree-day [J]. *Sustainable Cities and Society*, 2019, 51: 101698.
- [13] NEALE A, KUMMERT M, BERNIER M. Discriminant analysis classification of residential electricity smart meter data [J]. *Energy and Buildings*, 2022, 258: 111823.
- [14] AHMAD T, CHEN H X, GUO Y B, et al. A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy

- demand: A review [J]. *Energy and Buildings*, 2018, 165: 301-320.
- [15] LIANG R, LE-HUNG T, NGUYEN-THOI T. Energy consumption prediction of air-conditioning systems in eco-buildings using hunger games search optimization-based artificial neural network model [J]. *Journal of Building Engineering*, 2022, 59: 105087.
- [16] LIU H H, LIU Y C, GUO X, et al. An energy consumption prediction method for HVAC systems using energy storage based on time series shifting and deep learning [J]. *Energy and Buildings*, 2023, 298: 113508.
- [17] ZHUANG D, GAN V J L, DUYGU TEKLER Z, et al. Data-driven predictive control for smart HVAC system in IoT-integrated buildings with time-series forecasting and reinforcement learning [J]. *Applied Energy*, 2023, 338: 120936.
- [18] MOHAN R, PACHAURI N. An ensemble model for the energy consumption prediction of residential buildings [J]. *Energy*, 2025, 314: 134255.
- [19] 廖文强, 王江宇, 陈焕新, 等. 基于长短期记忆神经网络的暖通空调系统能耗预测[J]. *制冷技术*, 2019, 39(1): 45-50, 54.
- LIAO W Q, WANG J Y, CHEN H X, et al. Energy consumption prediction of HVAC system based on long short-term memory neural network [J]. *Chinese Journal of Refrigeration Technology*, 2019, 39(1): 45-50, 54. (in Chinese)
- [20] WEST G B, BROWN J H, ENQUIST B J. The fourth dimension of life: Fractal geometry and allometric scaling of organisms [J]. *Science*, 1999, 284(5420): 1677-1679.
- [21] JACQUIN A E. Image coding based on a fractal theory of iterated contractive image transformations [J]. *IEEE Transactions on Image Processing*, 1992, 1(1): 18-30.
- [22] LI Y X, ZHANG S, LIANG L L, et al. Multivariate multiscale higuchi fractal dimension and its application to mechanical signals [J]. *Fractal and Fractional*, 2024, 8(1): 56.
- [23] 胥永刚, 何正嘉. 分形维数和近似熵用于度量信号复杂性的比较研究[J]. *振动与冲击*, 2003, 22(3): 25-27, 5.
- XU Y G, HE Z J. Research on comparison between approximate entropy and fractal dimension for complexity measure of signals [J]. *Journal of Vibration and Shock*, 2003, 22(3): 25-27, 5. (in Chinese)
- [24] 吕安民, 李成名, 林宗坚, 等. 基于分形的时间序列模式挖掘方法及其应用[J]. *郑州大学学报(自然科学版)*, 2001, 33(4): 59-62.
- LÜ A M, LI C M, LIN Z J, et al. Mining time series model based on fractal theory and its application [J]. *Journal of Zhengzhou University (Natural Science Edition)*, 2001, 33(4): 59-62. (in Chinese)
- [25] YE D Y, LIU G N, MA T R, et al. The mechanics of frost heave with stratigraphic microstructure evolution [J]. *Engineering Geology*, 2023, 319: 107119.
- [26] YANG R, HE S, YI J Z, et al. Nano-scale pore structure and fractal dimension of organic-rich Wufeng-Longmaxi shale from Jiaoshiba area, Sichuan Basin: Investigations using FE-SEM, gas adsorption and helium pycnometry [J]. *Marine and Petroleum Geology*, 2016, 70: 27-45.
- [27] POTIRAKIS S M, MINADAKIS G, EFTAXIAS K. Sudden drop of fractal dimension of electromagnetic emissions recorded prior to significant earthquake [J]. *Natural Hazards*, 2012, 64(1): 641-650.
- [28] KALAMARAS N, TZANIS C G, DELIGIORGI D, et al. Distribution of air temperature multifractal characteristics over Greece [J]. *Atmosphere*, 2019, 10(2): 45.
- [29] ASTAF'EVA N M. Wavelet analysis: Basic theory and some applications [J]. *Uspekhi Fizicheskikh Nauk*, 1996, 166(11): 1145.
- [30] LI B, ZHANG T X, ZHANG C. Investigation of financial bubble mathematical model under fractal-fractional caputo derivative [J]. *Fractals*, 2023, 31(5): 2350050.
- [31] WANG H Y, WANG T T. Multifractal analysis of the Chinese stock, bond and fund markets [J]. *Physica A: Statistical Mechanics and Its Applications*, 2018, 512: 280-292.
- [32] 孙凌芸, 张金林. 投资者情绪与股票市场收益的相关性研究: 基于多重分形分析方法[J]. *金融评论*, 2017, 9(5): 92-102, 126.
- SUN L Y, ZHANG J L. Investor sentiment and stock market returns: A study based on multiple fractal analysis [J]. *Chinese Review of Financial Studies*, 2017, 9(5): 92-102, 126. (in Chinese)
- [33] 殷佳敏, 谢宁, 王承民, 等. 基于分形理论的新型电力系统规划场景生成方法[J]. *电力自动化设备*, 2024, 44(7): 94-100, 188.
- YIN J M, XIE N, WANG C M, et al. Generation method of planning scenario for new power system based on fractal theory [J]. *Electric Power Automation Equipment*, 2024, 44(7): 94-100, 188. (in Chinese)
- [34] 马忠娇, 张吉礼. 基于POD-LSE-FCD的iBES异常能耗数据检测方法[J]. *建筑热能通风空调*, 2022, 41(9): 1-5, 47.
- MA Z J, ZHANG J L. Detection method of iBES abnormal energy consumption data based on POD-LSE-FCD [J]. *Building Energy & Environment*, 2022, 41(9): 1-5, 47. (in Chinese)
- [35] 徐建军, 王硕昌, 袁硕, 等. 基于改进的分形理论的短期电力负荷预测[J]. *吉林大学学报(信息科学版)*, 2022, 40(3): 347-353.
- XU J J, WANG S C, YUAN S, et al. Short-term

- power load forecasting based on improved fractal theory [J]. *Journal of Jilin University (Information Science Edition)*, 2022, 40(3): 347-353. (in Chinese)
- [36] FRANÇA L G S, MIRANDA J G V, LEITE M, et al. Fractal and multifractal properties of electrographic recordings of human brain activity: toward its use as a signal feature for machine learning in clinical applications [J]. *Frontiers in Physiology*, 2018, 9: 1767.
- [37] THAMRIN C, STERN G, FREY U. Fractals for physicians [J]. *Paediatr Respir Rev*, 2010, 11(2): 123-131.
- [38] MUGHEES A, KAMRAN M. Novel fractal-based features for low-power appliances in non-intrusive load monitoring [J]. *Computers, Materials & Continua*, 2024, 80(1): 507-526.
- [39] YANG X B, JIN X Q, DU Z M, et al. A novel model-based fault detection method for temperature sensor using fractal correlation dimension [J]. *Building and Environment*, 2011, 46(4): 970-979.
- [40] 祝用华, 杜志敏, 晋欣桥, 等. 关联维数在空调系统能耗分析中的应用[C]//上海市制冷学会. 上海市制冷学会 2009 年学术年会论文集, 2009: 523-526.
- ZHU Y H, DU Z M, JIN X Q, et al. Application of correlation dimension in energy consumption analysis of air conditioning systems [C]//Shanghai Refrigeration Society. *Proceedings of the 2009 Annual Academic Conference of Shanghai Refrigeration Society*, 2009: 523-526. (in Chinese)
- [41] KNOWLES H S, HOSTETLER M E, LIEBOVITCH L S. Describing the dynamics, distributions, and multiscale relationships in the time evolution of residential building energy consumption [J]. *Energy and Buildings*, 2018, 158: 310-325.
- [42] CHEN Z, IVANOV P C, HU K, et al. Effect of nonstationarities on detrended fluctuation analysis [J]. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, 2002, 65(4 Pt 1): 041107.
- [43] MIRZAYOF D, ASHKENAZY Y. Preservation of long range temporal correlations under extreme random dilution [J]. *Physica A: Statistical Mechanics and Its Applications*, 2010, 389(24): 5573-5580.
- [44] IHLEN E A F. Introduction to multifractal detrended fluctuation analysis in Matlab [J]. *Frontiers in Physiology*, 2012, 3: 141.
- [45] KANTELHARDT J W, ZSCHIEGNER S A, KOSCIELNY-BUNDE E, et al. Multifractal detrended fluctuation analysis of nonstationary time series [J]. *Physica A: Statistical Mechanics and Its Applications*, 2002, 316(1-4): 87-114.
- [46] LIMWATTANAPIBOOL O, ARCH-INT S. Determination of the appropriate parameters for K-means clustering using selection of region clusters based on density DBSCAN (SRCD-DBSCAN) [J]. *Expert Systems*, 2017, 34(3): e12204.
- [47] ISLAM S, ELMEKKI H, ELSEBAI A, et al. A comprehensive survey on applications of transformers for deep learning tasks [J]. *Expert Systems with Applications*, 2024, 241: 122666.

(编辑 XXX)