

文章编号:0253-4339(2025)05-0115-09

doi:10.12465/j.issn.0253-4339.2025.05.115

基于特征工况挖掘和先验概率引导的暖通空调系统参数辨识方法

贺佳宁 鲁洁 赵阳

(浙江大学制冷与低温研究所 杭州 310027)

摘要 暖通空调系统实测数据的高度冗余显著降低了其模型参数辨识的计算效率。为了解决上述问题,提出一种基于特征工况挖掘和先验概率引导的参数辨识方法。特征工况挖掘方法对运行数据进行相关性分析,选择暖通空调系统运行工况的关键变量,并根据关键变量维度进行网格采样,得到更具代表性的特征工况,从而提升模型单次计算效率。模型参数辨识过程中,建立待辨识参数的先验概率模型,基于先验概率进行待辨识参数的先验区间估计和改进目标函数,引导模型更快收敛。该方法在武汉某工厂暖通空调系统供冷季一个月的运行数据集上进行了验证,结果表明:该方法中的 MAPE(平均绝对百分比误差)和 CV-RMSE(均方根误差变异系数)相比基于 K-means 聚类的方法分别降低了 16.0% 和 12.0%,比基于原始数据的方法分别降低了 20.9% 和 15.2%, NMBE(归一化平均偏差误差)更接近 0, R^2 (决定系数)相比 2 种方法分别提升了 4.7% 和 8.5%, 计算效率提高约 39.3%。该方法为实现高效准确的暖通空调系统建模提供了技术指导和数据支撑。

关键词 HVAC 系统;冷水机组建模;参数辨识;特征工况;粒子群算法

中图分类号:TU831.4; TP183

文献标识码:A

Feature Condition Mining and Prior Probability Guidance Based Model Calibration Methodology for HVAC System

He Jianing¹ Lu Jie¹ Zhao Yang¹

(Institute of Refrigeration and Cryogenics, Zhejiang University, Hangzhou, 310027, China)

Abstract The high redundancy of the measured data from heating, ventilation, and air conditioning (HVAC) systems significantly reduces the computational efficiency of model calibration. To address this challenge, a model calibration method based on mining feature operating conditions and a priori probability guidance was introduced in this study. Correlation analysis was conducted on the operational data for mining feature operating conditions. Feature variables related to HVAC system operation were selected, and a grid sampling technique based on these characteristic variables was employed to obtain representative operating conditions, enhancing the efficiency of the model calculations. Additionally, a prior probability model was established for the parameters to be calibrated during the model calibration process. A priori interval estimation was then performed, and the objective function was improved based on the prior probability to guide the model towards faster convergence. The proposed method was validated using a one-month operational dataset from a cooling plant in an industrial building located in Wuhan, China. The results indicated that the proposed method achieved significant improvements in performance metrics. Specifically, mean absolute percentage error (MAPE) and cross-validated root mean square error (CV-RMSE) were reduced by 16.0% and 12.0%, respectively, compared to the K-means clustering-based method, and by 20.9% and 15.2%, respectively, compared to the baseline data-based method. Furthermore, the normalized mean bias error (NMBE) was closer to zero, and the coefficient of determination (R^2) increased by 4.7% and 8.5%, respectively, compared to the two aforementioned methods. Additionally, our method enhanced the computational efficiency by approximately 39.3%. This method provides technical guidance and data support for achieving an efficient and accurate modeling of HVAC systems.

Keywords HVAC system; chiller modeling; model calibration; feature condition; particle swarm algorithm

建筑运行能耗约占全国能源消费总量的 21%^[1],其中约 50% 的建筑运行能耗来自暖通空调

(ventilation and air conditioning, HVAC) 系统^[2]。HVAC 系统节能运行对达成“双碳”目标至关重要。

基金项目:国家自然科学基金(52161135202)资助项目。(The project was supported by the National Natural Science Foundation of China (No. 52161135202).)

收稿日期:2024-04-23;修回日期:2024-07-13;录用日期:2024-08-30

近年来,随着大数据、物联网和人工智能等技术的快速发展,基于数字孪生的 HVAC 系统优化运行技术^[3]得到了工业界和学术界的普遍关注,具有很大的发展潜力。参数辨识是构建可靠的数字孪生模型最重要的环节,它通过实际数据来辨识出模型参数,让模型能够正确反映实际系统特性^[4-5]。

模型辨识方法可分为手动方法和自动方法^[6]。手动方法需要专业知识和反复试验迭代^[7],耗费大量的人力。自动化辨识方法通常依赖于数学和建模工具,利用目标函数来减少测量数据和仿真数据之间的差异。优化算法如遗传算法(GA, genetic algorithm)和粒子群优化算法(PSO, particle swarm optimization)等已被用于自动辨识研究中。R. G. Ramos 等^[8]采用多目标遗传算法 NSGA-II 来辨识建筑围护结构。M. G. Larochelle 等^[9]使用元启发式算法自动辨识 EnergyPlus 暖通空调子系统模型。近年来,随着支持模型辨识的计算资源和工具变得更加经济,自动辨识得到了蓬勃发展^[10]。

在 HVAC 系统实际的运行过程中,其运行工况一般重复出现,导致运行数据冗余度较高。直接使用此类数据进行模型参数辨识会显著增加计算量,且对模型精度的提升有限^[11]。部分研究人员尝试通过少量运行工况下的数据进行模型参数辨识从而提高模型计算效率。M. G. Larochelle 等^[9]使用单一运行工况下的测量数据对 HVAC 系统模型进行参数辨识。结果表明,在该运行工况下,辨识模型的准确性很高,然而随着运行工况的改变,模型的平均绝对百分比误差从 10% 增至近 50%。Lu Jie 等^[11]提出了一种基于聚类压缩的方法,将 K-means 聚类算法得到的每个簇中心点作为该簇中数据的代表点,实现数据压缩,提高了 HVAC 系统模型校准的计算效率。然而, K-means 聚类方法在计算簇中心点时,同一簇内数据的运行工况存在差异,计算出的代表数据并不能够真实反映实际运行情况,在训练数据有限的情况下,模型容易过度拟合这些不真实的合成数据,从而影响参数辨识的准确性和可靠性。因此,如何在保证模型精度的同时,去除冗余数据从而提高参数辨识计算效率是当下亟待解决的关键问题。

基于采样的数据挖掘方法能保证所提取数据的真实性,得到代表整个数据集的特征数据。欠采样方法能减少多数类样本的样本数量^[12],在一定程度上能够降低数据冗余度,如随机欠采样^[13]、EasyEnsemble^[14]等。然而,这些方法是针对分类任务提出的^[12]。若能够利用采样的思想挖掘运行过程的特征工况,进行参数辨识,有很大的潜力能够解决 HVAC

系统运行数据中存在的冗余问题。

如何科学定义参数寻优区间,并减少前期的寻优负担是 HVAC 系统辨识问题所面临的另一个挑战。过宽的参数区间会增加参数辨识的复杂性和时间成本。目前的研究中通常会设置较宽的搜索区间,增加了探索最优解的计算负担,在优化算法目标函数的设计上,已有研究大多数仅衡量仿真值与实际值差距^[6-7]。若能够在辨识过程中融入先验知识,为参数提供合理的先验区间,并对目标函数进行改进,在目标函数中引入先验知识项,就能够引导模型更快地收敛,提高辨识效率。

针对上述 2 个问题,本研究提出了一种基于特征工况挖掘和先验概率引导的 HVAC 系统参数辨识方法。提出了一种基于网格采样的特征工况挖掘方法来提取 HVAC 系统的代表性工况,能够以一种直观可解释的方式提取运行数据,得到系统真实的特征工况,保证了模型能够准确捕捉真实的系统特征,同时减少了模型单次计算所需时间,提升了参数辨识的时间效率;提出了一种先验概率引导的方法来建立待辨识参数的先验概率模型,基于先验知识改进寻优区间和目标函数,以引导模型更快收敛。该方法在某工厂 HVAC 系统供冷季一个月的运行数据集上进行验证。

1 暖通空调系统参数辨识方法

1.1 方法流程图

暖通空调系统参数辨识方法流程如图 1 所示。本研究提出的基于特征工况挖掘的 HVAC 系统参数辨识方法主要包括 2 个步骤:特征工况挖掘和基于先验概率引导的模型参数辨识。

1.2 特征工况挖掘

特征工况挖掘流程如图 2 所示。特征工况挖掘分为 2 个步骤:关键变量筛选和基于网格采样的特征工况挖掘。关键变量筛选旨在通过相关性分析选择 HVAC 系统运行工况的代表性变量。基于网格采样的特征工况挖掘依据关键变量维度进行数据重构,在新的数据空间中进行网格均匀采样提取特征工况。

1.2.1 关键变量筛选

HVAC 系统的运行数据包括多个变量维度,如功率、流量和温度等。关键变量选择对于减少多重共线性至关重要,即从高度相关的变量中选择一个或一组具有代表性的变量,实现数据降维。相关性分析是实现这一目标的有效方法^[15]。在本研究中,使用皮尔逊相关系数^[16]来确定变量之间的相关性:

$$\rho = \frac{\text{cov}(X, Y)}{s(X) \times s(Y)} \quad (1)$$

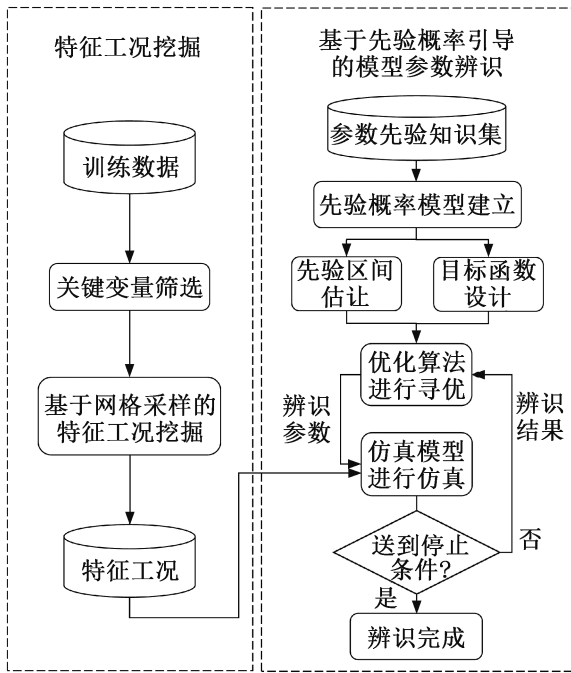
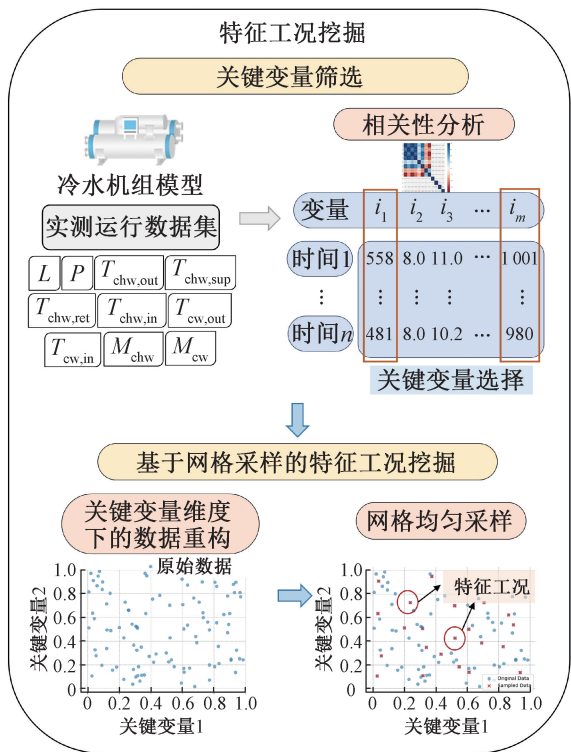


图 1 暖通空调系统参数辨识方法流程图

Fig.1 Flowchart of HVAC system model calibration method



L 冷负荷, kW; P 功率, kW; $T_{chw,out}$ 冷冻水出口温度, $^{\circ}\text{C}$;
 $T_{chw,sup}$ 冷冻水供应温度, $^{\circ}\text{C}$; $T_{chw,ret}$ 冷冻水返回温度, $^{\circ}\text{C}$;
 $T_{chw,in}$ 冷冻水入口温度, $^{\circ}\text{C}$; $T_{cw,out}$ 冷却水出口温度, $^{\circ}\text{C}$;
 $T_{cw,in}$ 冷却水入口温度, $^{\circ}\text{C}$; M_{chw} 冷冻水流量, m^3/h ;
 M_{cw} 冷却水流量, m^3/h 。

图 2 特征工况挖掘流程图

Fig.2 Flowchart of feature condition mining method

式中: ρ 为皮尔逊相关系数; X 和 Y 为要分析的 2 个

变量; $\text{cov}(X, Y)$ 为 X 和 Y 之间的协方差; $s(X)$ 和 $s(Y)$ 分别为 X 和 Y 的标准差。

通过相关性分析, 确定哪些变量之间存在高度相关性。对于相关系数较高的变量组 ($\rho > 0.75$), 选择其中一个作为代表, 替代其他变量, 得到反映 HVAC 系统运行工况的关键变量, 实现运行数据集在变量维度上的降维。

1.2.2 基于网格采样的特征工况挖掘

采用基于网格采样的特征工况挖掘方法, 从测量数据中提取具有代表性的运行工况, 具体步骤如下:

1) 进行关键变量维度下的数据重构。首先, 基于所选的关键变量维度对运行数据进行重构, 并确定在这些新维度下的数据空间范围。然后, 对数据进行归一化处理^[17], 调整至 $[0, 1]$ 的范围内, 使变量维度单位一致, 如式(2)所示。

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (2)$$

式中: X 为原始数据值; X_{\min} 为数据集中最小值; X_{\max} 为数据集中最大值; X' 为归一化后的数据值。

2) 计算需采样的特征工况点数量。设置训练集的采样率, 根据采样率和训练集大小, 计算需要经过采样得到的特征工况点数量。

3) 对重构的数据空间进行网格划分。确定数据空间中每个维度的划分粒度, 将每一维度分别划分, 形成网格空间。

4) 进行特征工况点的采样。首先确定每个网格的采样数量, 然后在数据点数量超过采样数量的网格中进行随机采样, 选取采样数量的数据点。对于数据点数量少于或等于采样数量的网格, 不进行采样。

5) 提取特征工况数据。从每个网格中提取所选的特征工况数据, 得到的特征工况可视为在整个运行数据空间中均匀分布的工况点。

1.3 基于先验概率引导的模型参数辨识

先验概率引导的第 1 步是通过参数的先验概率分布建立待辨识参数 g_{ij} 的先验概率模型。仿真软件 Dymola 提供了麦克维尔、约克、开利和特灵等生产商不同型号冷水机组 DOE-2 模型性能曲线系数集^[18], 将此作为领域先验知识, 得到各参数的先验分布, 建立待辨识参数的先验概率模型。

其次, 基于先验知识为待辨识的参数进行先验区间估计。选择与所需要辨识的冷水机组型号一致的 DOE-2 模型性能曲线系数集, 其中包含了各个待辨识参数的取值。在定义参数搜索空间时, 利用先验概率模型来估计模型参数的范围, 从而科学定义寻优区间。

然后,构建融合数据和知识的模型参数匹配程度的目标函数,如式(3)所示,该目标函数由 2 项组成,一项是数据项,本文选用平均绝对误差(MAE, mean absolute error),该项衡量辨识后模型的仿真输出与实测数据的差距;另一项是先验知识项 $f(g_{ij}, r_{ij})$,该项评估了当前的模型参数 g_{ij} 与其先验概率分布的一致性情况。 k 为权重系数,定义了先验分布知识指导在该评分体系中发挥的作用权重。

$$F_{\text{obj}} = \frac{1}{n} \sum_{i=1}^n |s_i - m_i| + kf(g_{ij}, r_{ij}) \quad (3)$$

式中: s 为冷水机组模型仿真的逐时功率,kW; m 为实测功率,kW; g_{ij} 为参数寻优值; n 为用于辨识的样本数量。

对于先验知识项 $f(g_{ij}, r_{ij})$,采用能够度量 2 个分布之间相似性的高斯核函数,以引导优化算法更偏向先验知识,从而加速收敛。

$$f(g_{ij}, r_{ij}) = e^{-\gamma \|g_{ij} - r_{ij}\|^2} \quad (4)$$

式中: $\|g_{ij} - r_{ij}\|$ 为 2 个特征向量之间的平方欧几里得距离; γ 为高斯核函数的超参数。

最后,进行粒子群算法寻优。模型仿真的训练集为前述所获得的特征工况。采用粒子群算法(PSO)辨识 HVAC 系统模型, PSO 算法的基本思想是通过模拟个体(粒子)在解空间中的移动来搜索最优解。首先,随机初始化粒子群的位置和速度。然后,对于每个粒子,利用目标函数评估每个个体的性能,即生成参数的辨识精度,从而计算其适应度,并更新个体的位置。每个粒子根据个体最优位置和全局最优位置,更新其速度和位置。重复这一过程,直至达到停止标准,如达到最大迭代次数,从而得到模型的最优解。

1.4 模型参数辨识精度的评价指标

使用 4 个指标来评估辨识精度,包括平均绝对百分比误差(MAPE, mean absolute percentage error)、归一化平均偏差误差(NMBE, normalized mean bias error)、均方根误差变异系数(CV-RMSE, cross-validated root mean square error)和决定系数(R^2 , coefficient of determination)。MAPE、NMBE 和 CV-RMSE 越小,仿真值和实测值越接近,模型辨识的精度越高; R^2 越接近 1,模型辨识的精度越高。

$$M_{\text{MAPE}} = \frac{1}{n} \sum_{i=1}^n \frac{|m_i - s_i|}{m_i} \quad (5)$$

$$N_{\text{NMBE}} = \frac{1}{m} \frac{\sum_{i=1}^n (m_i - s_i)}{n} \quad (6)$$

$$C_{\text{CV-RMSE}} = \frac{1}{m} \sqrt{\frac{\sum_{i=1}^n (m_i - s_i)^2}{n}} \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (m_i - s_i)^2}{\sum_{i=1}^n (m_i - \bar{m}_i)^2} \quad (8)$$

式中: m 为实测值; s 为仿真值; \bar{m} 为平均实测值; n 为样本个数。

2 验证结果与分析

2.1 仿真模型建立

采用最常用的 DOE-2 模型^[19]来建立冷水机组的物理模型。DOE-2 模型是基于三条多项式曲线的经验模型。 $f(C_{\text{CAP}}, T)$ 表示可用制冷量与蒸发器和冷凝器温度函数关系的曲线:

$$f(C_{\text{CAP}}, T) = g_{11} + g_{12}T_{\text{chws}} + g_{13}T_{\text{chws}}^2 + g_{14}T_{\text{cws}} + g_{15}T_{\text{cws}}^2 + g_{16}T_{\text{chws}}T_{\text{cws}} \quad (9)$$

$f(E_{\text{EIR}}, T)$ 表示满负荷效率与蒸发器和冷凝器温度关系的曲线:

$$f(E_{\text{EIR}}, T) = g_{21} + g_{22}T_{\text{chws}} + g_{23}T_{\text{chws}}^2 + g_{24}T_{\text{cws}} + g_{25}T_{\text{cws}}^2 + g_{26}T_{\text{chws}}T_{\text{cws}} \quad (10)$$

$f(E_{\text{EIR}}, P_{\text{PLR}})$ 表示效率与部分负荷率关系的曲线:

$$f(E_{\text{EIR}}, P_{\text{PLR}}) = g_{31} + g_{32}P_{\text{PLR}} + g_{33}P_{\text{PLR}}^2 \quad (11)$$

式中: T_{chws} 为蒸发器供水温度,℃; T_{cws} 为冷凝器供水温度,℃; E_{EIR} 为 COP 的倒数,即轴功率与制冷量之比; P_{PLR} 为部分负荷率; $g_{11} \sim g_{33}$ 为待辨识的参数。

当前运行工况下的冷水机组最大可用制冷量:

$$C_{\text{CAP}} = C_{\text{CAP rated}} f(C_{\text{CAP}}, T) \quad (12)$$

式中: $C_{\text{CAP rated}}$ 为冷水机组的额定制冷量,kW; C_{CAP} 为当前运行工况下的最大制冷量,kW。

基于上述 3 条性能曲线,可以计算冷水机组的实际功率:

$$P = \frac{f(E_{\text{EIR}}, P_{\text{PLR}}) C_{\text{CAP}}}{\text{COP}_{\text{rated}} f(E_{\text{EIR}}, T)} \quad (13)$$

式中: P 为冷水机组的功率,kW。

采用武汉某工厂 HVAC 系统中 1#冷水机组的运行数据集(采样间隔为 5 min)对所提方法进行验证。该 HVAC 系统在 TRNSYS^[20]上建立的系统模型如图 3 所示。数据集的测量时间为 2019 年 7 月 22 日—9 月 30 日。运行数据变量包括功率(P)、冷负荷(L)、冷冻水出口温度($T_{\text{chw, out}}$)、冷冻水供应温度($T_{\text{chw, sup}}$)、冷冻水返回温度($T_{\text{chw, ret}}$)、冷冻水入口温度

($T_{chw,in}$)、冷却水出口温度($T_{cw,out}$)、冷却水入口温度($T_{cw,in}$)、冷冻水流量(M_{chw})及冷却水流量(M_{cw})。

为了模拟实际建模过程中,仅根据较短时间段内的运行数据对模型参数辨识,并确保模型在未来较长时间内保持较好的辨识精度,本研究选取 3 d 的数据作为原始训练集,包括 7 月 22 日—7 月 24 日的 1#冷水机组开启的运行数据,共 862 个数据点,测试集为 7 月 25 日—9 月 30 日期间 1#冷水机组开启的时间,共 5 986 个数据点。从工厂收集的运行数据被用作辨识模型的模型输入。输入数据包括冷冻水入口流量(M_{chw})、冷却水入口流量(M_{cw})、冷负荷(L)、室外干球温度(T_d)及湿球温度(T_w)。

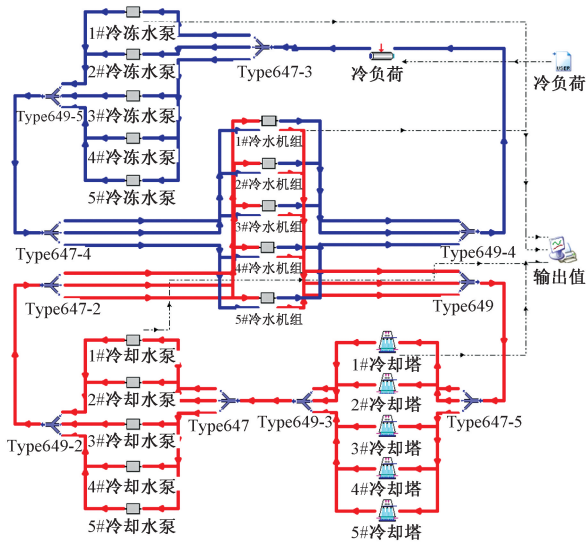


图 3 TRNSYS 中系统模型

Fig.3 Modeling of the cooling plant system in TRNSYS

2.2 特征工况挖掘结果

利用相关性分析衡量运行数据集变量之间的关系,图 4 所示为使用皮尔逊相关系数获得的运行数据变量之间的相关性热图。 $P-L$ ($r=0.99$)、 $P-T_{chw,ret}$ ($r=0.91$)、 $P-T_{chw,in}$ ($r=0.91$)和 $P-T_{cw,out}$ ($r=0.95$) 之间的相关系数很高,这意味着可以用功率变量有效替代其他变量。 $M_{chw}-T_{chw,out}$ ($r=0.78$)和 $M_{chw}-T_{chw,sup}$ ($r=0.78$) 之间的相关系数表明,可以用 M_{chw} 替代 $T_{chw,out}$ 和 $T_{chw,sup}$ 。 M_{cw} 保持稳定,因此与其他变量的相关系数较低,在选择关键变量时将其排除在外。根据上述分析,选择 P 和 M_{chw} 作为关键变量。

原始训练数据集在所选关键变量空间中归一化后的表示如图 5 所示,共计 862 个数据点。归一化后功率和流量的范围均为 $[0, 1]$,按照 0.2 的间隔进行网格划分。可以看出,数据高度集中在某些值上,表明许多数据点反映了相同的运行工况,存在数据冗余。

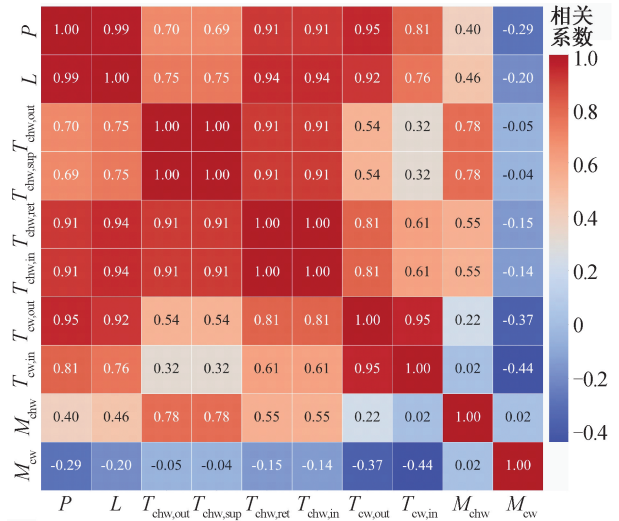


图 4 运行数据集变量间的相关性热图

Fig.4 Correlation coefficient between variables of operation data

对每个网格中的数据进行均匀采样以选择特征工况点。对于每个网格,若有 5 个以上的数据点,则随机选取其中 5 个进行均匀采样。采样数量可根据原始数据集的大小进行调整。通过均匀网格抽样获得的特征工况,共计 80 个点,为原始训练集的 9.3%,如图 5 所示。采样后的特征工况数据集将作为模型参数辨识的训练集。

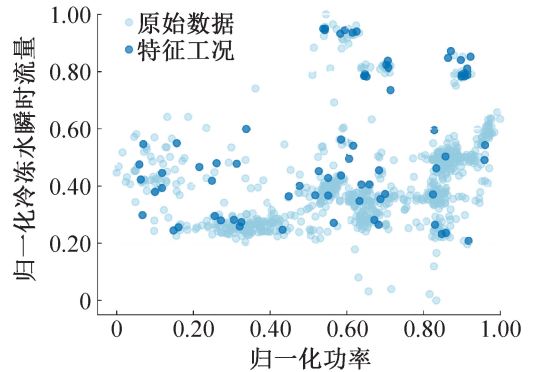


图 5 特征工况挖掘前后的数据分布图

Fig.5 Distribution of representing data before and after feature condition mining

2.3 基于先验概率的粒子群算法性能评估

从 Dymola 模型性能曲线系数集得到型号为开利的冷水机组参数样本,进行统计分析,确定每个参数的先验区间;使用正态分布拟合参数分布,得到各参数的概率密度函数。以待辨识参数 $g_{11} \sim g_{13}$ 为例,参数的先验分布及概率密度函数曲线如图 6 所示。

将冷机性能曲线系数的最大值和最小值分别作为对应参数寻优范围的上下边界,如表 1 所示。

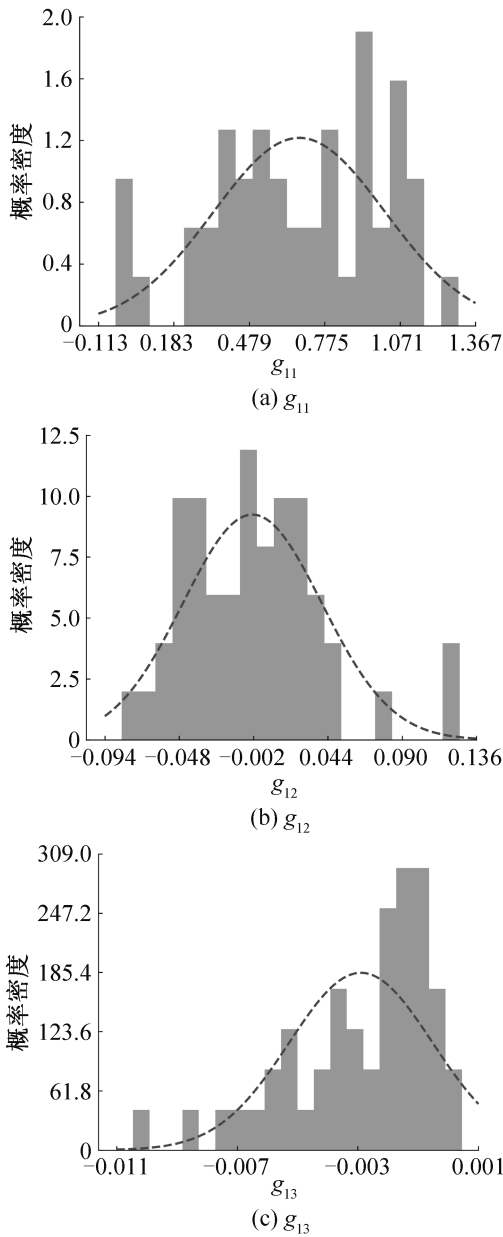


图 6 待辨识参数 $g_{11} \sim g_{13}$ 的先验分布及概率密度曲线
Fig.6 A priori distribution of the parameters $g_{11} \sim g_{13}$

搭建了 Python 与 TRNSYS 的联合仿真平台,进行 TRNSYS 和 Python 之间的数据传输。基于原始训练数据集采用 Python 的粒子群算法对 1#冷水机组的 DOE-2 模型进行目标函数中超参数 k 与 γ 的寻优预实验,采用网格搜索法,在预设的超参数空间内进行穷举搜索。 k 的取值为 $\{10, 20, 50, 100\}$, γ 的取值为 $\{0.0001, 0.001, 0.01, 0.1\}$, 最终确定 $k = 20, \gamma = 0.01$ 。本文提出的目标函数包括数据项与先验知识项 $f(g_{ij}, r_{ij})$, 如式(14)所示。

$$F_{obj} = \frac{1}{n} \sum_{i=1}^n |s_i - m_i| - 20 \text{Exp}(-0.01 \|g_{ij} - r_{ij}\|^2) \quad (14)$$

表 1 1#冷水机组的待辨识参数及其搜索区间

Tab.1 Parameters of 1# chiller and associated search spaces

参数	下边界	上边界
g_{11}	-0.050	1.487
g_{12}	-0.090	0.149
g_{13}	-0.010	0.001
g_{14}	-0.010	0.140
g_{15}	-0.004	0.000
g_{16}	-0.002	0.008
g_{21}	0.150	1.243
g_{22}	-0.090	0.159
g_{23}	-0.010	0.004
g_{24}	-0.020	0.144
g_{25}	-0.004	0.002
g_{26}	-0.003	0.008
g_{31}	0.050	0.627
g_{32}	-0.600	0.986
g_{33}	-0.006	1.454

由于粒子群算法具有随机性,在目标函数有先验知识项和无先验知识项的情况下进行了 5 组辨识实验,粒子群算法的收敛性能对比如表 2 所示。结果表明,基于先验概率的粒子群算法通过目标函数中先验知识项的引导,能够使模型更快收敛,缩短了 26.4% 的辨识时间,大幅提升了模型参数辨识的效率。

表 2 目标函数有无先验知识项的收敛性能对比

Tab.2 Comparison of convergence performance of objective functions with and without prior knowledge terms

	平均迭代次数/次	平均辨识时间/s
有先验知识项	92	212
无先验知识项	125	288

2.4 基于特征工况挖掘的参数辨识方法性能评估

为了研究所提特征工况挖掘方法的性能,选择了 2 种方法进行对比。其中一种方法使用原始训练数据集,另一种方法对其进行 K-means 聚类,对于每个簇,计算其数据点的平均值,将这些平均值作为簇中心。将每个簇的中心作为训练集,簇的数量与本文方法训练集大小一致。粒子群算法中种群大小和最大

迭代次数均为10。

3种方法得到的训练数据集情况如图7所示。基于采样的方法相比基于K-means聚类的方法,特征工况的分布更加均匀。基于K-means聚类的方法通过对聚类簇中的工况点取平均值来得到特征工况,因此存在一些实际中未出现的假工况,对参数辨识造成干扰;同时数据集中缺失了冷水机组功率低于570 kW、冷冻水流量低于960 m³/h的低负荷工况,可能会影响参数辨识中对于低负荷工况的拟合精度。

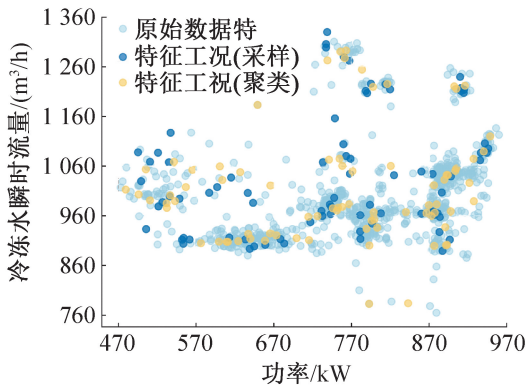


图7 3种方法的训练数据集情况

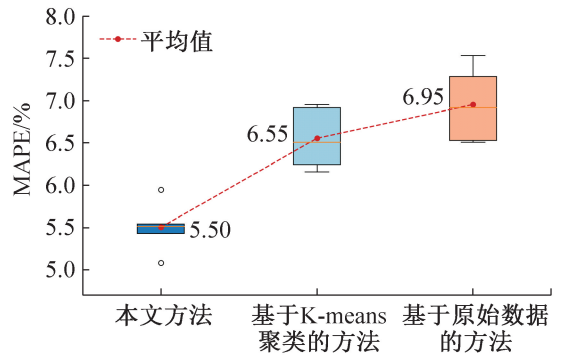
Fig.7 Training dataset for three methods

在1#冷水机组运行数据的测试集上对上述3种方法的辨识精度进行了评估。由于粒子群算法具有随机性,进行了5次重复实验,结果如图8所示。评价指标是针对1#冷水机组的功率计算的。

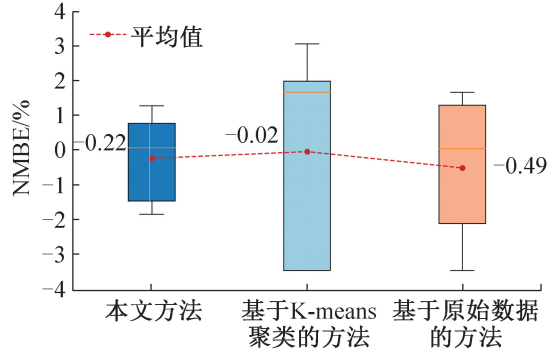
与其他方法相比,本文的方法具有更高的精度。本方法中MAPE和CV-RMSE的平均值相比基于K-means聚类的方法分别降低了16.0%和12.0%,比基于原始数据的方法分别降低了20.9%和15.2%,NMBE更接近0, R²的平均值比基于K-means聚类的方法提升了4.7%,比基于原始数据的方法提升了8.5%。

图9所示为3种方法在测试集上的模型测量和仿真拟合结果的对比。基于原始数据的方法的训练数据集为原始3d的数据,由图7中训练集情况可知,在冷水机组低功率情况下的工况数据点占比较少,中高功率数据在辨识过程中占据较大比例。因此,在测试集上,低功率工况的拟合情况较差。基于聚类的方法在低功率工况的拟合精度有所提升,但由于训练数据集中均值工况的干扰,拟合精度仍不及本文提出的方法。

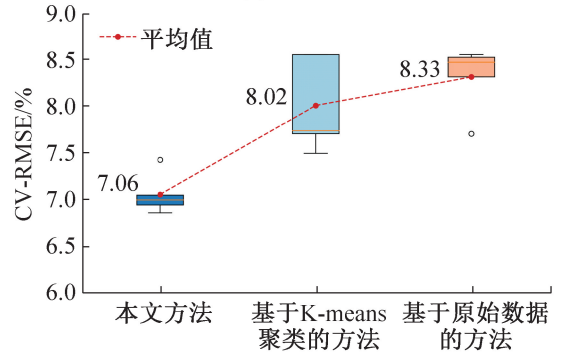
对3种方法的时间成本进行评估,如图10所示。本文方法通过去除冗余工况的数据,减少了



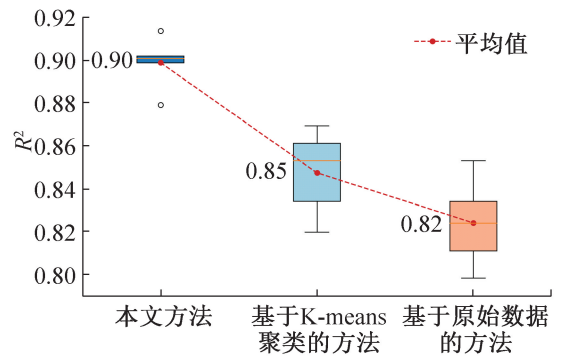
(a) MAPE对比



(b) NMBE对比



(c) CV-RMSE对比



(d) R²对比

图8 3种方法在测试集上模型辨识精度对比

Fig.8 Comparison of model calibration accuracy of three methods on test set

模型单次仿真的计算时间,比基于原始数据的方法缩短了39.3%的计算时间,大幅提升了辨识的时间效率。

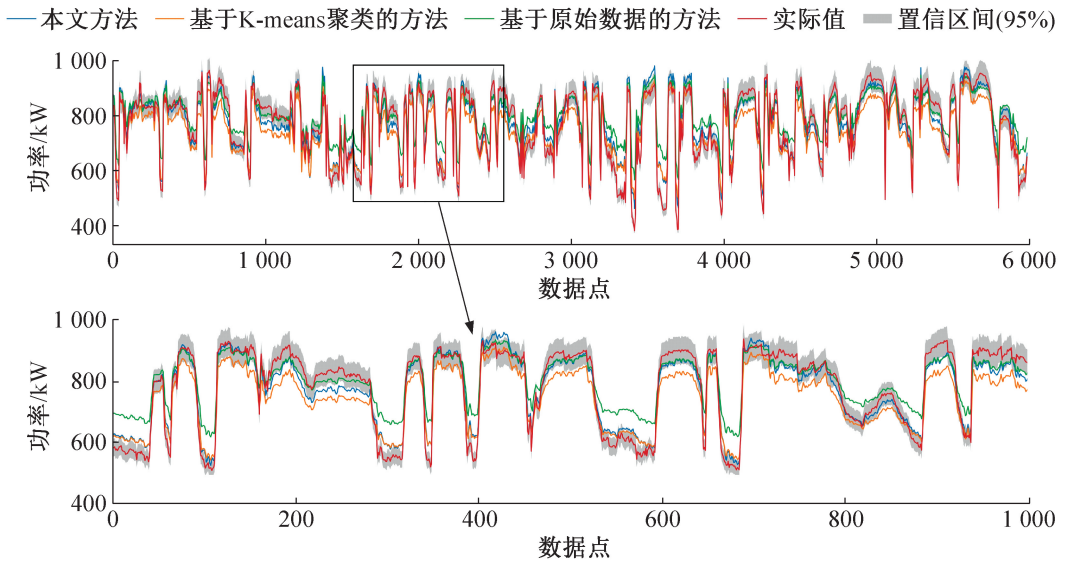


图 9 3 种方法在测试集上模型功率拟合曲线对比

Fig.9 Comparison of the power fitting curves of the three methods on the test set

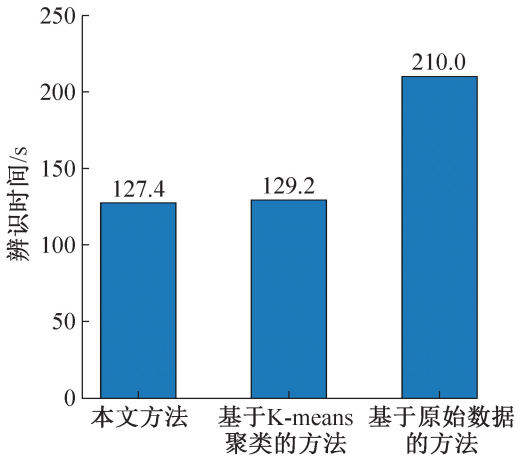


图 10 不同辨识方法的时间成本

Fig.10 Time costs of calibration process for the three methods

3 结论

本文提出一种基于特征工况挖掘和先验概率引导的 HVAC 系统参数辨识方法。首先,提出了一种基于网格采样的特征工况挖掘方法,用于提取 HVAC 系统的代表性工况,显著提高了参数辨识的时间效率。其次,本文提出了一种先验概率引导的方法用于构建待辨识参数的先验概率模型。得到如下结论:

1)通过对运行数据中特征工况的挖掘,保证了模型能够准确捕捉真实的系统特征,降低了运行数据的冗余;对辨识算法进行先验概率引导,基于先验知识改进寻优区间和目标函数,提升了参数辨识的效率。

2)提出的基于特征工况挖掘的方法提升了模型

精度。1#冷水机组功率的 MAPE、NMBE、CV-RMSE、 R^2 分别为 5.5%、-0.22%、7.06%、0.90。与基于 K-means 聚类的方法相比,NMBE 更接近 0, MAPE、NMBE 和 R^2 的精度分别提高了 16.0%、12.0% 和 4.7%;与基于原始数据的方法相比,精度分别提高了 20.9%、15.2% 和 8.5%。

3)本方法显著提升了参数辨识的时间效率。基于先验概率的粒子群算法在辨识中具有更好的收敛性能,目标函数中加入先验知识项后,收敛速度提升了 26.4%;通过特征工况挖掘,去除了仿真中的冗余工况,比基于原始数据的方法计算效率提升了 39.3%。

4)本文只选择了冷水机组模型作为研究对象进行评估,但该方法也可应用于其他暖通空调系统设备的物理模型,如冷却塔、水泵等。未来工作将围绕提升暖通空调系统参数辨识效率展开。

参考文献

[1] 中国城市科学研究会.《中国建筑节能年度发展研究报告 2023(城市能源系统专题)》[M]. 北京:中国建筑工业出版社, 2023. (Chinese Society for Urban Studies. China building energy efficiency annual development research report 2023 urban energy systems theme [M]. Beijing: China Architecture Building Press, 2023.)

[2] JUNG W, JAZIZADEH F. Human-in-the-loop HVAC operations: a quantitative review on occupancy, comfort, and energy-efficiency dimensions [J]. Applied Energy, 2019, 239: 1471-1508.

[3] CHEN K, ZHU X, ANDUV B, et al. Digital twins model

- and its updating method for heating, ventilation and air conditioning system using broad learning system algorithm [J]. *Energy*, 2022, 251: 124040.
- [4] DENG Zhang, CHEN Yixing, YANG Jingjing, et al. Archetype identification and urban building energy modeling for city-scale buildings based on GIS datasets [J]. *Building Simulation*, 2022, 15(9): 1547–1559.
- [5] HONG T, LANGEVIN J, SUN K. Building simulation: ten challenges [J]. *Building Simulation*, 2018, 11(5): 871–898.
- [6] SUN K, HONG T, KIM J, et al. Application and evaluation of a pattern-based building energy model calibration method using public building datasets [J]. *Building Simulation*, 2022, 15(8): 1385–1400.
- [7] FABRIZIO E, MONETTI V. Methodologies and advancements in the calibration of building energy models [J]. *Energies*, 2015, 8(4): 2548–2574.
- [8] RAMOS R G, FERNÁNDEZ B C, GÓMEZ-ACEBO T T, et al. Genetic algorithm for building envelope calibration [J]. *Applied Energy*, 2016, 168: 691–705.
- [9] LAROCHELLE M G, MONFET D, NOUANEGUE H F, et al. Energy calibration of HVAC sub-system model using sensitivity analysis and meta-heuristic optimization [J]. *Energy and Buildings*, 2019, 202: 109382.
- [10] ZHAO Yang, LI Tingting, ZHANG Xuejun, et al. Artificial intelligence-based fault detection and diagnosis methods for building energy systems: advantages, challenges and the future [J]. *Renewable and Sustainable Energy Reviews*, 2019, 109: 85–101.
- [11] LU Jie, TIAN Xiangning, FENG Chenxin, et al. Clustering compression-based computation-efficient calibration method for digital twin modeling of HVAC system [J]. *Building Simulation*, 2023, 16(6): 997–1012.
- [12] ZHANG Chaobo, LI Junyang, ZHAO Yang, et al. Problem of data imbalance in building energy load prediction: concept, influence, and solution [J]. *Applied Energy*, 2021, 297: 117139.
- [13] REN Xinyuyang, ZHANG Chaobo, ZHAO Yang, et al. A data mining-based method for revealing occupant behavior patterns in using mechanical ventilation systems of Dutch dwellings [J]. *Energy and Buildings*, 2019, 193: 99–110.
- [14] LIU Xuying, WU Jianxin, ZHOU Zhihua. Exploratory undersampling for class-imbalance learning [J]. *IEEE Transactions on Systems, Man, and Cybernetics Part B, Cybernetics*, 2009, 39(2): 539–550.
- [15] LEE K, LIM H. Correlation analysis of building parameters according to ASHRAE Standard 90.1 [J]. *Journal of Building Engineering*, 2024, 82: 108130.
- [16] LEE R J, ALAN N W. Thirteen ways to look at the correlation coefficient [J]. *The American Statistician*, 1988, 42(1): 59–66.
- [17] YU Liping, PAN Yuntao, WU Yishan. Research on data normalization methods in multi-attribute evaluation [C]// 2009 International Conference on Computational Intelligence and Software Engineering. Wuhan: IEEE, 2009.
- [18] 张丽珠, 章超波, 陈琦, 等. 基于遗传算法和人工神经网络的冷水机组模型参数辨识及误差补偿方法 [J]. *制冷学报*, 2021, 42(3): 93–99. (ZHANG Lizhu, ZHANG Chaobo, CHEN Qi, et al. Genetic-algorithm-based parameter identification and artificial-neural-network-based error compensation for chiller model [J]. *Journal of Refrigeration*, 2021, 42(3): 93–99.)
- [19] HYDEMAN M, GILLESPIE JR K L. Tools and techniques to calibrate electric chiller component models [J]. *ASHRAE Transactions*, 2002, 108(1): 733–741.
- [20] KLEIN S A, BECKMAN W A. TRNSYS 16: a transient system simulation program: mathematical reference [J]. *TRNSYS*, 2007, 5: 389–396.

通信作者简介

赵阳,男,长聘副教授,浙江大学能源工程学院,18814803300, E-mail: youngzhao@zju.edu.cn。研究方向:建筑能源系统大数据分析 with 认知计算,制冷设备故障检测与诊断。

About the corresponding author

Zhao Yang, male, tenured associate professor, College of Energy Engineering, Zhejiang University, 86–18814803300, E-mail: youngzhao@zju.edu.cn. Research fields: artificial intelligence-based building energy systems, fault detection and diagnosis.