



同濟大學

TONGJI UNIVERSITY

硕士学位论文

基于数据挖掘的公共建筑
节能潜力分析方法

姓名：何睿凯

学号：1930256

学院：机械与能源工程学院

学科门类：工学

学科名称：供热、供燃气、通风及空调工程

指导教师：许鹏 教授

二〇二二年三月



同濟大學
TONGJI UNIVERSITY

A dissertation submitted to

Tongji University in conformity with the requirements for
the degree of Master of Engineering

**Data-Mining Analysis of Energy-Saving
Potential in Public Buildings**

Candidate: Ruikai He

Student Number: 1930256

School/Department: School of Mechanical Engineering

Discipline: Engineering

Major: HVAC & Gas Engineering

Supervisor: Prof. Peng Xu

March, 2022

基于数据挖掘的公共建筑节能潜力分析方法

何睿凯

同济大学

摘要

针对既有建筑进行节能潜力分析是完成“双碳”目标的重要一环，全国公共建筑能耗监测平台的快速发展以及数据驱动算法在全行业的深度应用有助于数据挖掘技术应用于建筑节能潜力分析，这样不仅缩短传统建筑能源审计的周期，而且辅助技术人员快速定位节能潜力点。

本文提出一种公共建筑节能潜力分析框架，可以分析**建筑内热负荷**，**设备选型方案**，**冷站运维**三个方面的节能潜力。

从数据利用方面，可以利用基本建筑信息、冷站设备信息和冷站总能耗时间数据对建筑内热负荷，设备选型方案和冷站运维三方面作节能潜力分析；或者利用详细的分项计量数据和冷站设备信息在冷站运维方面做详细的节能潜力分析。

在建筑内热负荷方面，本块的节能潜力分析是建立在研究建筑能耗关键变量的基础上，根据推测的关键变量真值对建筑负荷进行节能潜力分析。首先基于 Grasshopper 和 eppy 的批量建模与建筑能耗模拟技术，构建建筑-能耗数据库；以此为基础**训练以关键变量为输入、冷站总能耗为输出的黑箱模型**并采用树模型的方法对黑箱模型进行优化；最后串联全局随机搜索算法**定性分析**建筑内热负荷构成。

在设备选型方面，根据易获取的和推测出的影响建筑负荷的关键变量，利用轻量化黑箱模型推测出建筑负荷，随后以最大化平均部分负荷率为优化目标优化冷机名义制冷量；以节能规范中对冷机性能的要求为下限选择真实冷机模型做节能潜力分析。

在冷站运维方面，考虑到不同建筑拥有的数据差异性，提出两种不同程度的分析——“浅诊断”和“深诊断”。“浅诊断”是指：仅根据冷机能耗时序数据**定性分析**冷机自身是否存在性能劣化，分析方法同样采用关键变量推测的方法；“深诊断”是指：根据分项计量数据采用统计学方法分析是否设备和系统是否存在常见的能效问题；由于分项数据存在多变量、多工况的特性，而不同的建筑的分项数据易存在五花八门的问题，如何清洗分项计量数据成为“深诊断”前亟待解决的问题，本文提出基于真实数据交叉验证的分项计量数据处理方法进行数据清洗。

最后，本文采用两个信息化程度迥异的建筑案例验证本文的节能潜力分析方法的准确性和实用性。信息化程度较低的建筑用于验证建筑内热负荷、设备选型方案及“浅诊断”三方面的节能潜力分析方法，信息化程度较高的建筑用于验证“深诊断”部分的节能潜力分析方法。

本文围绕建筑节能潜力分析方法展开，从信息化程度较低到信息化程度较高的建筑均能找到切入口并采用本文提出的方法分析节能潜力，所以本文提出的节能潜力方法具有一定的实用性。

关键词：节能潜力分析，数据质量分析，超参数调优，数据挖掘

ABSTRACT

Analysis of energy-saving potential for existing public buildings is an necessity for carbon-related mission. Thanks to the rapid development of energy monitoring platforms and widespread application of machine learning, data-mining becomes an inevitable trend for analysis of building energy-saving potential, which can speed up traditional, labor-intensive, time-consuming building energy audit.

The proposed framework of energy-saving potential for existing public buildings can be utilized in three aspects: building thermal load, the scheme selection of chillers, maintenance and operation (M&O).

In terms of data, the data package, basic building and equipment information and the time series data of total electricity consumption of plant, is utilized to analyze the energy-saving potential for thermal load, scheme selection of chillers and M&O. Meanwhile, detailed analysis for M&O requires submetering data with basic equipment information.

The analysis for thermal load is based on the researches about key variables affecting building electricity consumption. Detecting true value of key variables is the core in this part. First, the databased of building-consumption is established with Grasshopper and eppy which provide technology of batch modeling and simulation. And then, the data-driven model would be trained on the database mentioned, and the inputs and output are key variables and total electricity consumption of plant respectively. Specifically, the tree-based approach to hyperparameter-tuning is used to tap into the potential of data-driven model. At the end, random global search algorithms are used to detect the true value of key variables for qualitative analysis for thermal load.

In the scheme selection of chillers, lightweight data-driven models are taken advantaged of to detect the distribution of thermal load. And then, the random global search algorithm is used to maximize the average part load ratio to optimize the rated capacities of chillers. According to the performance requirements of chillers in rated code, the actual chiller models are selected to analyze for this part.

Considering the huge gap about building informatization, two types of analysis are proposed in M&O, simple and detailed version. Simple version means, total electricity consumption of chiller group are utilized to tell whether the chiller degradation has happened or not, which is based on detection of key variables.

Detailed version means: statistical method is utilized to tell if there are some common problems about equipment efficiency. The intrinsic property of submetering data, multiple conditions, multi-variables, results in various problems about data quality, which means data-cleansing is the indispensable step before detailed version. The paper proposes an data-cleansing approach based on cross validation.

At the end, two distinct cases are selected to verify the accuracy and usability of the proposed framework. The first case is less informative, which are used to verify three parts, building thermal load, the scheme selection of chillers, and simple version of M&O. The second one is much informative, which means it is a good example to verify the detailed version of M&O.

The core of the paper is surrounded by how to analyze energy-saving potential in three parts. Two distinct cases can demonstrate that no matter what kind of building in informatization, it can find starting point to analyze in the proposed framework.

Key words: analysis of energy-saving potential, analysis of data quality, hyper-parameters tuning, data-mining

目录

摘要.....	I
ABSTRACT.....	III
符号注释表.....	VIII
第 1 章 引言.....	1
1.1 概述.....	1
1.1.1 碳中和政策驱动.....	1
1.1.2 建筑大数据平台现状.....	2
1.2 国内外相关研究.....	4
1.2.1 基于白箱模型的建筑节能潜力分析.....	5
1.2.2 基于黑箱模型的建筑节能潜力分析.....	6
1.2.3 现状与挑战.....	6
1.3 本课题的主要研究内容.....	7
1.3.1 本课题的研究对象和重要概念说明.....	7
1.3.2 本文的主要架构.....	9
1.4 基于数据挖掘的建筑节能潜力分析的流程搭建.....	9
1.4.1 整体技术路线.....	9
1.4.2 基于冷站总能耗的建筑节能潜力分析.....	9
1.4.3 基于分项数据的建筑节能潜力分析.....	11
1.5 本章小结.....	11
第 2 章 基于关键变量推断的建筑内热负荷分析.....	12
2.1 本章技术路线.....	12
2.2 建筑-能耗算例设计.....	13
2.2.1 建筑能耗数据库构建.....	13
2.2.2 建筑-能耗算例的批量生成.....	14
2.3 用于推测关键变量的黑箱模型形成.....	16
2.3.1 黑箱模型选择.....	16
2.3.2 黑箱模型的超参数优化.....	18
2.4 基于关键变量推测的建筑节能潜力分析.....	19
2.4.1 基于全局搜索算法的关键变量推测.....	19
2.4.2 衡量推测偏差的指标.....	21
2.5 本章小结.....	22
第 3 章 基于轻量化黑箱模型的设备选型方案节能潜力分析.....	23

3.1 本章技术路线.....	23
3.2 建筑负荷算例生成.....	23
3.3 基于树模型的黑箱模型超参数调优方法.....	25
3.3.1 黑箱模型超参数调优框架.....	25
3.3.2 动态超参数空间抽样方法.....	26
3.3.3 剪枝算法.....	28
3.4 基于全局搜索优化选型方案的建筑节能潜力分析.....	29
3.4.1 全局搜索算法.....	29
3.4.2 衡量方案优化的指标.....	30
3.4.3 以真实冷机为模板分析节能潜力.....	31
3.5 本章小结.....	32
第 4 章 基于实测分项计量数据的冷站运维节能潜力分析.....	33
4.1 本章技术路线.....	33
4.2 浅诊断.....	34
4.2.1 浅诊断中关键变量的选择.....	34
4.2.2 辅助推断关键变量的最优黑箱模型形成.....	35
4.2.3 分项能耗的数据预处理.....	35
4.2.4 基于关键变量推测的浅诊断能效诊断分析.....	38
4.3 深诊断.....	39
4.3.1 分项计量数据的数据质量问题概述.....	39
4.3.2 分项计量数据的缺失特性分析.....	40
4.3.3 基于交叉验证的分项计量数据的数据清洗框架.....	45
4.3.4 常见能效问题诊断.....	46
4.4 本章小结.....	47
第 5 章 基于实际建筑案例的节能潜力分析验证.....	48
5.1 案例一.....	48
5.1.1 案例描述与数据介绍.....	48
5.1.2 案例验证内容.....	50
5.1.3 基于关键变量推断的建筑内热负荷分析方法验证.....	50
5.1.4 基于轻量化黑箱模型的设备选型方案节能潜力分析方法验证.....	57
5.1.5 基于分项数据的冷站运维节能潜力分析——浅诊断方法验证.....	68
5.1.6 案例节能潜力分析总结.....	73
5.2 案例二.....	74
5.2.1 建筑描述与数据介绍.....	74

5.2.2 案例验证内容.....	76
5.2.3 深诊断方法验证——全楼宇分项计量数据数据清洗流程实例.....	76
5.2.4 深诊断方法验证——能效问题诊断.....	85
5.2.5 案例节能潜力分析总结.....	88
5.3 本章小结.....	89
第 6 章 总结与展望.....	90
6.1 本文总结.....	90
6.2 主要贡献.....	90
6.3 局限性与展望.....	91
参考文献.....	92
致谢.....	96
个人简历、在读期间发表的学术成果.....	97

符号注释表

缩写词	全称及含义
BAS	Building Automation Systems, 建筑自控系统
HVAC	Heating, Ventilation, Air Conditioning & Cooling, 供热、通风、空调以及制冷
SVM	Supporting Vector Machine, 支持向量机
SVR	Supporting Vector Regression, 支持向量机回归
ANN	Artificial Neural Network, 人工神经网络
CNN	Convolutional Neural Network, 卷积神经网络
RL	Reinforcement Learning, 强化学习
DOE	Department of Energy, 美国能源部
DoE	Design of Experiments, 实验设计方法
DT	Decision Tree, 决策树
EFB	Exclusive Feature Bundling, 互斥特征合并
SA	Sensitivity Analysis, 敏感性分析
GA	Genetic Algorithm, 遗传算法
PCA	Principal Component Analysis, 主成分分析法
MBE	Mean Bias Error, 平均误差
CV-RMSE	Coefficient of Variation of Root Mean Square Error, 变异系数均方根误差
DTW	Dynamic Time Warping, 动态时间规划
LHS	Latin Hypercube Sampling, 超拉丁抽样
LHD	Latin Hypercube Design, 超拉丁法设计
PSO	Particle Swarm Optimization, 粒子群算法
CAD	Computer-aided Design, 计算机辅助设计
ML	Machine Learning, 机器学习
EBNILM	Event-Based Non-Intrusive Load Monitoring, 非侵入式负荷拆分方法
NIM	Non-Intrusive Monitoring, 非侵入式监测法
NILM	Non-Intrusive Load Monitoring, 非侵入负荷监测法
IM	Intrusive Monitoring, 侵入式监测法
GOSS	Gradient-based One-Side Sampling, 单边梯度采样
GBDT	Gradient Boosting Decision Tree, 梯度提升决策树

SRRC	Standardized Rank Regression Coefficient, 标准化秩回归系数
PRCC	Partial Rank Correlation Coefficient, 偏相关系数

变量符号	含义
B	支管
M	干管
T	温度
Q	流量
P	压力
C	电流百分比
E	能耗
F	水泵频率
FF	冷却塔内风扇频率
ON/OFF	启停状态
PLANT	冷站
CT(S)	冷却塔 (群)
CH(S)	冷机 (群)
CHWP(S)	冷冻水泵 (群)
CWP(S)	冷却水泵 (群)
LOAD	间接测得瞬时冷负荷
TIN	室内温度
TOUT	室外干球温度
HOUT	室外相对湿度
TSET	冷冻水供水温度设定点
TCHRW	冷冻水回水温度
TCHSW	冷冻水供水温度
TCRW	冷却水回水温度
TCSW	冷却水供水温度

下标符号	含义
<i>ce</i>	冷凝器入口
<i>ee</i>	蒸发器入口
<i>el</i>	蒸发器出口
<i>LV</i>	限制
<i>chw</i>	冷冻水

<i>s</i>	模拟条件下
<i>rated</i>	额定/设计条件下
<i>b</i>	支管
<i>cw</i>	冷却水
<i>ctf</i>	冷却塔内风扇

第 1 章 引言

1.1 概述

由于全球城镇化快速扩张，全球气候转暖的因素下^[1]，全球建筑能耗其在能源系统中占比约为 31%，部分区域可高达 57%。其总量从 1970 年折合 14 亿吨油量（TOE）发展到 2010 年 36 亿吨油当量，预计到 2040 年可到达 55 亿油当量^[2]。

在 2015 年，我国建筑总能耗占总体能源消耗的 28.6%。根据发达国家城市变迁发展，可以得出建筑总能耗占总能源消耗的比重将会提升至 33%^[66]。相较工业建筑和交通运输，大型公共建筑的能源消耗比重较大，其中国内的大型公共建筑存在较为明显的问题，比如高能耗，低能效，其表现在：2020 年公共建筑和城镇住宅建筑均占全国建筑总能耗的 38%，但公共建筑的建筑面积总量仅为城镇及农村住宅建筑的总建筑面积的 0.23^[3]，针对既有公共建筑的节能潜力分析势在必行。

1.1.1 碳中和政策驱动

Nan Zhou 等^[4]学者在 2018 年时认为在中国建筑碳排放在 2045 年达到饱和，这种推测的假设前提是建筑持续保持高能耗需求，直到 2030 年开始进入技术驱动节能的用能场景，虽然目前的建筑节能技术可以较为有效将建筑用能需求缩小，但这仍然需要强有力的政策驱动，才能突破建筑高能耗困境的瓶颈。

在此背景下，诸多建筑节能方面的工作在中国政府推动下大面积展开，其中包括针对新建建筑不断修订建筑用能限额规范，对既有建筑进行节能评估与改造以及在建筑领域推广可再生能源应用。这些举措都成为实现 2020 年中国政府正式提出“2030 年实现碳达峰，2060 年碳中和”的战略目标的关键环节^[5]。

针对既有建筑进行节能评估与改造中，最广泛形式为建筑节能审计。建筑能源审计是通过对建筑能源利用效率、消耗水平、经济效益和环境效果进行监测、诊断和评价，从而发现节能潜力并提出节能运行调试和改造建议^[6]。原住建部早在 2007 年便发布了《国家机关办公建筑和大型公共建筑能源审计导则》，并在日后广泛征求意见，对此进行修订，促进我国建筑能源审计不断发展。

建筑能源审计目前涉及的范围主要包括：建筑围护结构、动力系统、暖通空调系统、可再生能源系统、水资源利用情况及建筑室内环境的检测。因为建筑能源审计的目标繁多，在传统做法中，需要收集并统计异构化数据，耗时耗

力且流程复杂，如图 1.1 所示。这并不利于快速判断建筑节能潜力，同时在漫长审计流程中，建筑依旧会保持高用能状态，不利于实现“碳达峰”阶段性目标。

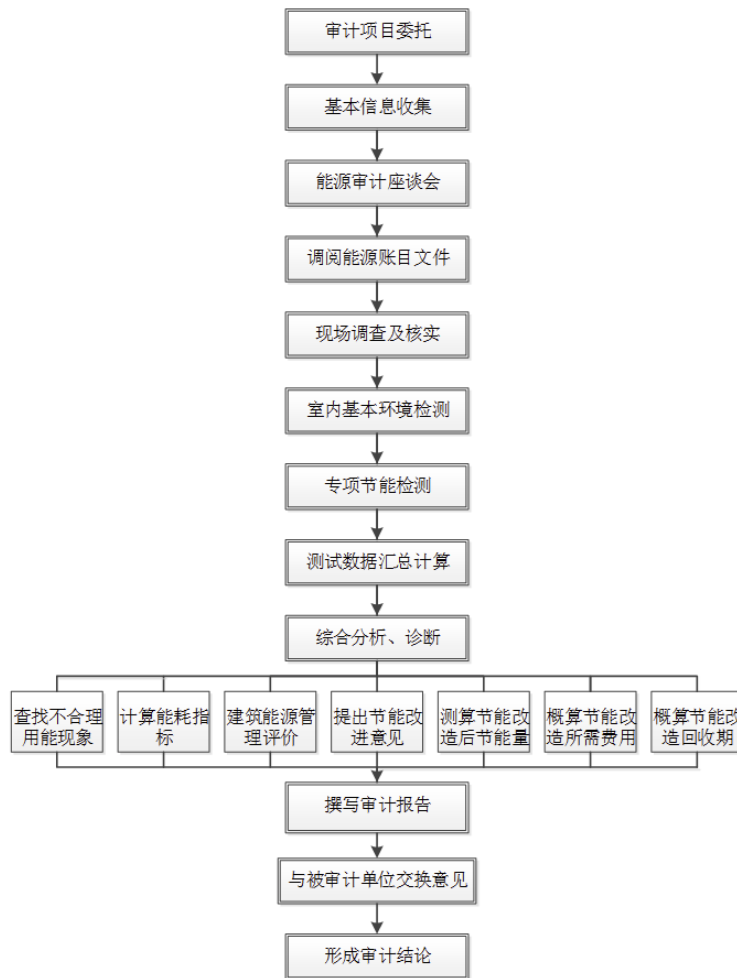


图 1.1 建筑能源审计流程

1.1.2 建筑大数据平台现状

自 2007 年住建部、财政部印发了《关于加强国家机关办公建筑和大型公共建筑节能管理工作的实施意见》后，先后已有 33 个省市开展公共建筑能耗检测平台建设，全国实施能耗建筑数量已经超过 10000 栋；以上海市为例，如图 1.2，至 2020 年 12 月 31 日，全市累计共有 2017 栋公共建筑完成用能分项计量装置的安装并实现与能耗监测平台联网，覆盖建筑面积 9208.3 万 m^2 ^[7]；在深圳市，截止 2021 年 1 月，深圳市接入能耗监测平台的公共建筑累计 702 栋，监测建筑总面积为 3456 万平方米，基本覆盖了各类公共建筑^[60]。这为建筑能耗的科学管理和控制奠定坚实的数据基础^[8]，并有效减轻能源审计员的工作复杂度：审计人员不必长时间蹲点检测数据以此指标评价建筑用能系统的效率。

建筑能耗数据作为时序数据，在不断积累的过程中，相关数据的密度、维度和复杂度均会不断提高，这会导致“数据灾难”和“维度灾难”等问题，这对传统的行业人员提出更高的数据分析和挖掘的能力，但这也为本课题搭建一套较为全面的、针对既有公共建筑节能潜力分析的流程提供了数据依据。

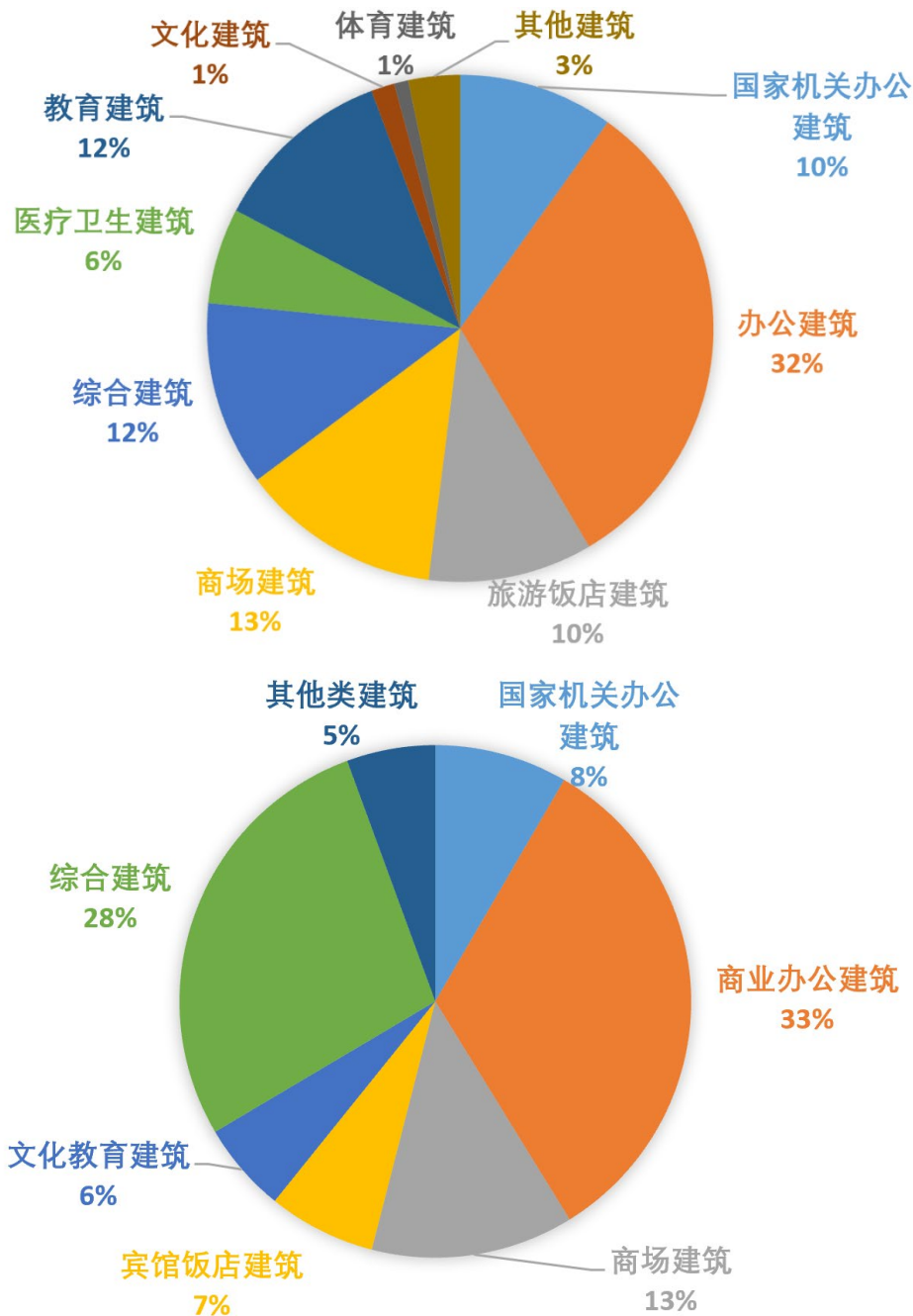


图 1.2 2020 年上海市（上）和深圳市（下）接入能耗监测平台不同类型建筑面积占比饼图

国内建筑能源检测平台起步较晚，依旧存在由于监测服务商不同，建筑间的数据形式不统一的问题，这会导致大量不同时间、空间维度的数据不能融合^[9]

较难形成完整的建筑用能画像，这对所有的既有公共建筑均进行无差别的节能潜力分析造成了很大的障碍，尤其当分析方法需要统一的数据输入格式时。

另外，无法最大化利用海量累计的数据无疑是一种资源浪费。为此很多学者采用仿真模拟方式补全既有建筑的用能画像，但这存在一个问题，真实的能耗监测平台数据可以反应出的天气、人流等带来的不确定性，而这是难以用模拟数据替代的，而在实际情况中，因为各类软硬件问题，能耗数据会存在诸如死值、缺失值异常值等问题，这无疑又进一步加剧了研究人员区别数据异常和不确定性带来的数据波动的难度，相信在不久未来，对建筑能源监测平台或建筑自控系统（Building Automation Systems, BAS）也会纳入建筑能源审计的评判范畴中。

1.2 国内外相关研究

目前针对建筑节能潜力分析可以分为基于模型的方法和免模型的方法，基于模型的方法又可以细分为白箱模型、黑箱模型、灰箱模型，免模型的方法又可以分为简化工程算法和数学统计方法，这里会更侧重基于模型的方法归纳梳理。

1) 白箱模型：白箱模型又称为正演模型法，目前较为主流的建筑全能耗模拟分析软件主要有：EnergyPlus、DeST、ESP-r、DOE-2、TRNSYS、Dymola 等^[10]，但是这些软件的侧重点并不相同：其中以 EnergyPlus 为例，该软件更偏向于建筑负荷模拟，其中的建筑用能系统，如 HVAC（Heating, Ventilation, Air Conditioning & Cooling）系统为虚拟空调系统，以节点表示空调系统的结构；另外一类的典型代表是 TRYSNS 和 Dyloma，这类软件可以自定义系统中设备组件和连接形式，在设备层级的性能模拟上可能更具精准性。但是白箱的模型缺点在于不能精确考虑天气、人员等导致的不确定性。

2) 黑箱模型：随着计算机软硬件的飞速迭代和量子计算^[11]的兴起，上个世纪提出的诸多机器学习方法重新在工业界中活跃起来，以 SVM（Support Vector Machine）为首的机器学习方法在诸多领域大放异彩，随后更加复杂人工神经网络（Artificial Neural Network, ANN），计算视觉领域的 CNN（Convolutional Neural Network），强化学习（Reinforcement Learning, RL）理论逐渐在建筑用能系统中被应用和证明其有效性，如 SN, Qiu 等采用 Q-Learning Table 对冷热源系统中的设备采用无模型控制^[12-14]。虽然诸多黑箱模型在很多任务中都有较为优异的表现，但是其解释性仍有待进一步探索和挖掘，其迁移性也存在种种限制。支撑黑箱模型的一大支柱是海量的建筑

能耗数据，不断累积的时序数据为自动挖掘信息^[15]提供来源。一般的黑箱模型在暖通空调节能潜力分析方面的应用流程^[16]见图 1.3。

3) 灰箱模型则是在上述两种方法中寻求一种平衡，使模型既考虑人员、天气等因素带来的不确定又符合物理规律。一种应用较为广泛灰箱模型是采用白箱模型的流程，但是对于不容易测量获得的参数采用实测数据拟合^[17]。

4) 简化工程算法，该方法计算工作量小且较为简单，计算精度在可容忍范围内，适合实际工程中的快速计算，如建筑能源审计。计算的依据主要是采用 ISO 16346^[61]和 ISO 1634^[62]中关于建筑整体能耗和计算制冷供热能耗系统的简化算法。

5) 数学统计方法是总结数据分析的规律评价公共建筑的用能情况，简单规范法^[63]、关联规则分析^[32]、聚类方法等。简单规则法是一种相对性能指标（单一输入和单一输出之比），计算简单，但若仅采用单一性能指标作为评价，则会相对狭隘。SC, Zhan^[64]对八十多栋不同类型建筑的每日逐时能耗时间序列聚类，基于能耗特性对建筑进行重分类，打破传统的基于建筑使用类型的建筑分类模式，以类为单位评估每一类的建筑中是否存在节能潜力分析。

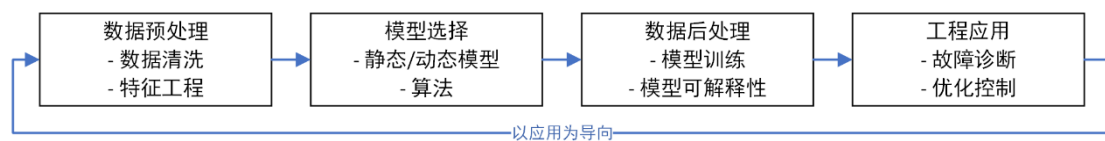


图 1.3 常见的黑箱模型应用的流程

1.2.1 基于白箱模型的建筑节能潜力分析

高力强等^[18]采用 eQUEST 对石家庄铁道大学基础教学楼的空调系统性能进行模拟（地源热泵+风机盘管），得出结论：应优先在后续改造中考虑优化机组性能。李欣欣^[19]以西安某商场为研究对象，采用 DesignBuilder 和 EnergyPlus 进行模拟分析对比，得出应该更换冷水机组，调整机组开机时间等具体措施来节能。付林乐^[20]用 DeST-c 对青岛某办公建筑进行全年能耗的模拟，从建筑本体和空调系统两方面判断节能潜力。初期设计的不合理，后期设备的运维不善，人员节能意识的淡薄均会造成实际使用偏离初期设计，从而带来节能潜力。ZH, Pang 等^[21]以美国能源部（Department of Energy, DOE）标准商业建筑模型中的中小型建筑为例，用 OpenStudio 和 EnergyPlus 模拟 occupant-centric 控制（OCC）分析建筑节能潜力：在保证人员基本热舒适的同时减少新风量带来的新风负荷。YX, Chen 等^[22]用 DeST 和 MATLAB 计算月、周、小时和实时负荷，并以此为

基础对比在四种控制方式下的冷机能耗，以实时负荷为基础控制的冷机能耗和以逐时负荷为基础控制的冷机能耗相差无几。Tatiana Alves 等^[23,24]根据巴西建筑法规利用 EnergyPlus 建立标准建筑能耗模型，采用敏感性分析（Sensitivity Analysis, SA）并结合巴西未来建筑建设速度分析六种节能改造方法的节能量。Chaham Alalouch 等^[25]对 Oman 地区三类气候带地区的建筑用 eQUEST 模拟，以当地法规最低能效的标准进行改造节能，发现在热带干燥地区节能改造的效果最为明显。

1.2.2 基于黑箱模型的建筑节能潜力分析

黑箱模型相较于白箱模型来说对专业知识门槛较低且很多工具开源且定期维护更新，这些因素都使得采用黑箱模型进行建筑节能潜力分析成为一种流行趋势。Y, Pan 等^[26]提出基于 CatBoost 的能耗预测方法，并在多源数据集上进行验证，分析影响建筑能耗的关键变量，这对节能改造提供了参考。Marjian Ilbeigi 等^[27]采用 ANN 及 MLP（Multilayer Perceptron）预测建筑能耗，并采用 EnergyPlus 检测影响能耗的关键变量还分析变量与建筑能耗之间的敏感度，最后采用遗传算法（Genetic Algorithm, GA）计算节能潜力。XY, Li 等^[28]对某栋教学楼的能耗模式采用 AdaBoost 和 Random Forest 分类，挖掘影响教学楼能耗的关键变量。但目前少有文献研究聚类算法在节能潜力方面的研究，YB, Chen 等^[29]采用模糊 c 均值聚类（Fuzzy c-means Clustering）和支持向量机回归（Support Vector Regression, SVR）预测逐时负荷，但未进一步采用聚类方法分析建筑节能潜力。Siriwarin Petcharat 等^[30]利用聚类算法（Clustering Algorithm）对建筑照明系统做节能潜力分析，比一般分析结果精度更高。

1.2.3 现状与挑战

目前建筑外形愈发复杂多变，且诸多建筑内 HVAC 系统不再是单一系统，而是由多类系统糅杂构成完成的空调系统。不论是建筑几何模型，还是空调模型，在模拟软件中构建这些模型都对建模人员提出了更高的要求。不仅如此，复杂模型会增加计算时间。很多模型输入参数也在这一过程中变得难以量化，例如人员时刻表及设备启停。此外，新冠疫情对建筑用能系统的冲击是有目共睹的，增加了建筑用能的不确定性因素，这些因素都会导致白箱模拟的结果和实际用能情况存在较大的差别。

但是，不同的因素对建筑用能的影响程度并不相同，这一点可以从不同学者的敏感性（SA）分析结果得出。从宏观层面，可以用相对少的关键变量预测

粗颗粒的能耗数据，例如建筑单位面积年制冷能耗（Cooling EUI），逐月的 HVAC 总能耗数据，但是细颗粒的数据可能目前仍是模拟数据无法代替的。

目前现有的研究大部分是着眼于某一个方面，也就是针对特定的建筑的用能系统进行节能潜力分析或者能效诊断，并不是针对大部分既有公共建筑做较为全面的节能潜力分析；例如有的研究针对 HVAC 系统设备的初期选型设计，有的针对 HVAC 系统实际运行策略的优化，而且对于数据颗粒度都有较为严格的要求。另外，还需要有详细建筑信息的楼宇方能诊断，这个前提准备缺一不可。

本文研究内容想打破这两点，提出一套有多切入口的公共建筑节能潜力分析框架和方法，不同信息化程度的建筑均可采用本文提出方法做节能潜力分析：例如老旧的公共建筑，仅能提供基本的建筑、设备信息和冷站总能耗；或者交付年限不久的建筑，出于对隐私的保护，仅能提供冷站分项计量数据，和基本设备信息，这些建筑虽然从用能画像均处于“缺失状态”，但也可以采用本方法进行部分节能潜力分析。

1.3 本课题的主要研究内容

1.3.1 本课题的研究对象和重要概念说明

1.3.1.1 本课题的研究对象和实现目标

本课题的题目是基于数据挖掘的公共建筑节能潜力分析，研究目标是分析建筑在建筑内热负荷、设备选型方案、冷站运维三个方面的节能潜力；在数据挖掘方面，实现目标是不同的建筑能耗信息均可以作为输入进行不同程度的建筑节能潜力分析；这里采用的思路并非多源异构数据的融合，而是建立多个“数据接口”，不同的能耗数据导入不同的切入口可以进行差异化建筑节能潜力分析。

节能潜力分析的差异化主要是由建筑能耗数据的差异性造成的，这也本文从多切入口出发的原因，这里具体建筑能耗数据的差异性做进一步的说明：

1. 数据颗粒度不同，有的建筑的实时数据为逐时数据，有的建筑的实时数据是每 15 分钟的数据，本文均采用逐时数据进行建筑的节能潜力分析；
2. 数据丰富度不同，不同的既有建筑因为建造年限的不同，能耗监测平台的设计有所差异，建成不久的建筑通常采集数据较为全面，除能耗以外，水温，流量等数据的监测均会涉及；一些年代较为久远的建筑，可能在进行能源审计前仅存在设备群大类的电表数据；

3. 数据质量不同：不同建筑的能耗监测可能由不同的三方平台管理，实测数据质量会存在明显的差异，数据质量的差异同样会对节能潜力分析的精确度和可信度带来一定的影响。

1.3.1.2 本文节能潜力分析与能效诊断之间的关系

能效诊断是本文研究的节能潜力分析的一部分，从狭义角度上说，能效诊断的主体为空调系统和设备，如何消耗更少的能源处理相同的建筑负荷，但节能潜力分析在此范围上有所扩充，还包含分析建筑负荷是否合理，是否可以降低建筑负荷来减少设备的运行“负担”。

1.3.1.3 本课题的重要概念说明

- 1) 多切入口：一般的节能潜力分项方法和流程只有一个“首”和一个“尾”，即只有一处导入数据，所有流程必须全部运行方可得出最后的分析结果。但是本文提到的多切入口是指：有不只一个“首”和不只一个“尾”，根据能耗数据的特点，选择不同的切入口（“首”）导入数据，并不需要走完全部也可以得到部分节能潜力分析结果。这样做既可以保证充分利用了建筑能耗数据，同时尽量避免采用模拟数据融合后可能会导致的建筑用能画像虽然“补全”但是“失真”的情况发生。
- 2) 差异化建筑节能潜力分析：其中差异化是用“宏观”数据进行“宏观”节能潜力分析，用“微观”数据进行“微观”节能潜力分析。
- 3) “宏观”数据：包含建筑基本信息，设备信息以及冷站总能耗时序数据。
- 4) “微观”数据：建筑冷站的分项计量数据。
- 5) “宏观”分析：包括关于建筑负荷的节能潜力分析，关于建筑设备选型的节能潜力分析以及浅诊断。
- 6) “微观”分析：深诊断。
- 7) 关于建筑负荷的节能潜力分析：本文仅分析建筑负荷中内热部分的组成是否合理。
- 8) 关于设备选型方案的节能潜力分析：本文仅分析设计阶段的冷机选型方案与实际建筑负荷分布的匹配程度。
- 9) 浅诊断：在分析冷站运维的节能潜力时，因为数据的不充分，只定性分析是否冷机自身性能劣化而带来了部分节能潜力。
- 10) 深诊断：同样是在分析冷站运维的节能潜力时，利用分项数据对空调系统及设备进行能效诊断。
- 11) 本文涉及的数据：本文节能潜力分析的对象是建筑冷站，所以本文所用数据限定在：建筑静态信息、设备信息、冷站总能耗数据，冷站的分项计量数据。

1.3.2 本文的主要架构

本文一共分为七个章节：

第一章介绍课题研究的现实意义、技术背景以及可行性；说明了本文的具体研究内容和特点，最后介绍节能潜力分析的技术路线和流程搭建。

第二章介绍基于关键变量推断的建筑内热负荷分析的技术路线：采用批量建模和模拟的方法搭建建筑-能耗模拟数据库，在此基础上训练黑箱模型，后接全局随机优化算法推测影响建筑内热负荷的关键变量，定性分析建筑内热负荷构成。

第三章介绍基于轻量化黑箱模型的冷站设备选型方案节能潜力分析的技术路线，该部分的节能潜力分析需要建议在第二章的基础上，根据推测的真实建筑负荷分布采用全局随机优化算法优化冷机名义制冷量。

第四章介绍基于实测分项数据的冷站运维节能潜力分析的技术路线，鉴于不同建筑用于的分项数据不同，该部分有两个切入口对应不同的分析路线——“浅诊断”和“深诊断”。

第五章对采用两个在信息化程度上截然不同的建筑案例验证该分析方法和框架的实用性和准确性。

第六章对本文研究工作进行总结和展望。

1.4 基于数据挖掘的建筑节能潜力分析的流程搭建

1.4.1 整体技术路线

图 1.4 是本文提出的公共建筑节能潜力分析的技术路线流程图，技术路线中主要分为三块，每块内容章节的对应关系在 1.3.2 本文主要架构中有所提及，在这里不过多赘述。

1.4.2 基于冷站总能耗的建筑节能潜力分析

本文的分析对象是冷站，不涉及末端部分，所以采用的实测数据和其他信息，均是冷站相关的数据和信息。这里冷站总能耗的输入可以看作是一个切入口。

图 1.4 中，蓝色方框表示数据的要求，其中包括建筑历史动态数据（包括附近气象站采集的室外天气参数）和其他对应的静态数据（地理位置、建筑面积、建筑层高、体形系数、窗墙比等）。

利用建筑冷站总能耗数据和静态数据，辅以其他方法对未知关键变量的推测，用于分析建筑内热构成：是否人员在室情况导致建筑负荷过高，或者建筑的照明、设备散热带来的建筑负荷超过了国家建筑节能规范的要求。

根据上述关键变量的推测结果，构建黑箱模型可以得到建筑负荷的分布，这里是为分析设备选型方案上是否具有节能潜力提供数据支撑，根据全局优化算法以平均部分负荷率（Partial Load Rate, PLR）为指标可以得出最优的选型方案，匹配真实的冷机数据库可计算理想情况下在选型方案优化上的节能潜力量。

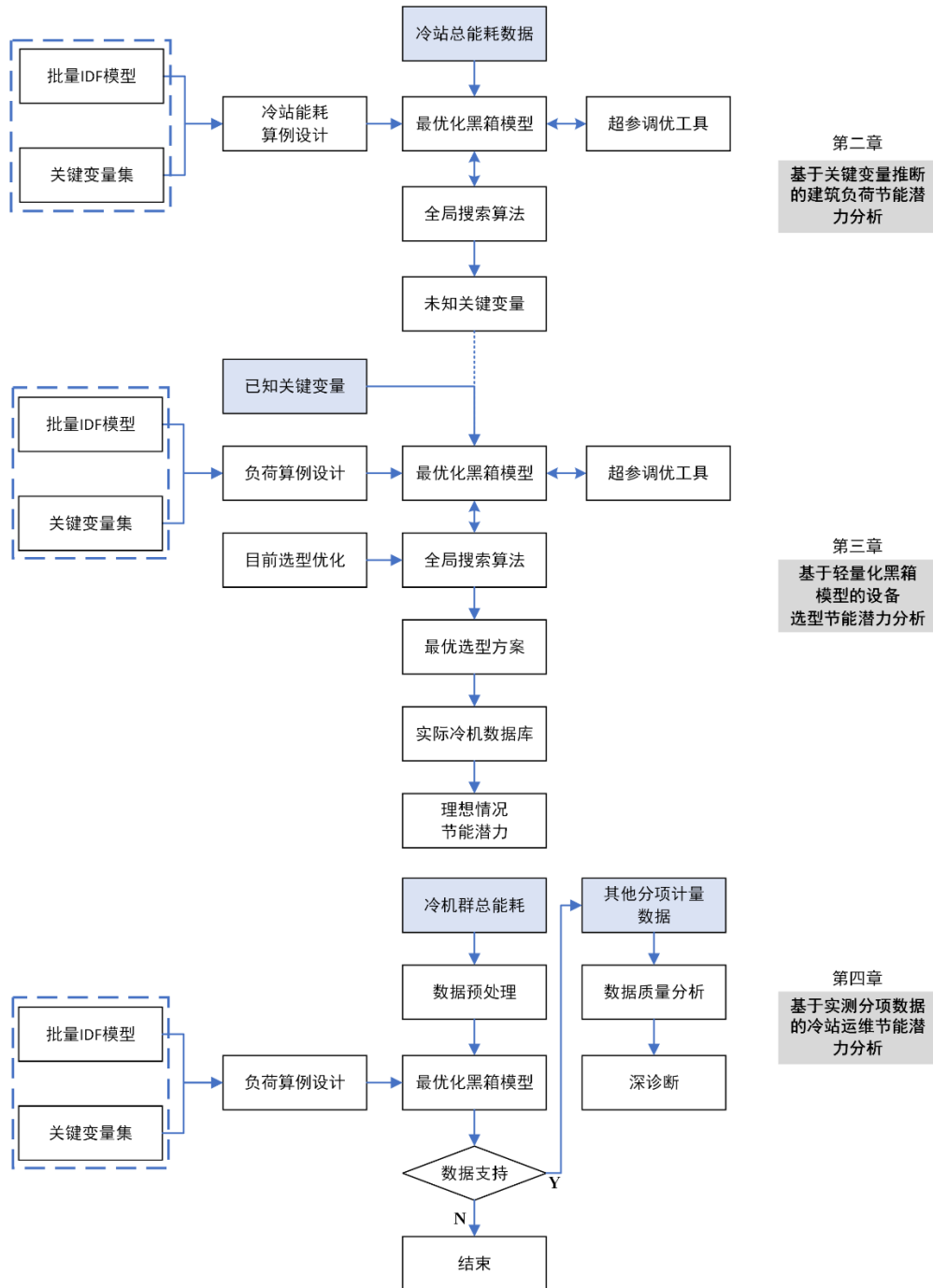


图 1.4 建筑节能潜力分析的流程搭建

1.4.3 基于分项数据的建筑节能潜力分析

这里内含**两个切入口**：冷机群总能耗和其他分项计量数据。冷站运维方面的节能潜力分析的差异性主要受分项数据的丰富度的影响。根据图 1.4，如果只有设备群的分项能耗则浅诊断：在数据预处理后，只用于判断节能潜力主要是否源于冷水机组自身的性能劣化；如果有其他分项计量数据则深诊断：进一步分析是否冷却塔等配套设备存在劣化问题抑或是其他能效问题。

1.5 本章小结

本章主要介绍了四部分内容，第一部分介绍了本文研究内容的背景，主要从国家政策角度和数据技术角度阐述本文研究内容的实际意义及可行性；第二部分主要介绍了本文研究的基础并总结了不同学者对于建筑节能潜力分析做出的贡献以及局限性；第三部分说明对本文的部分名词补充说明，并介绍本文的研究内容和特色；第四部分介绍了本文提出的公共建筑节能潜力分析的技术路线和流程框架，以不同的数据作为切入口分析建筑内热负荷、冷站设备选型方案和冷站运维情况。

第2章 基于关键变量推断的建筑内热负荷分析

2.1 本章技术路线

建筑负荷是多方共同参与的结果，如建筑形态，建筑功能，照明功耗，设备散热密度等。传统做法是参照建筑图纸构建逼近实际建筑的建筑能耗模型，随后利用能耗模拟软件分析建筑负荷在同类型建筑中的水平。这类精细工作有较高的专业门槛；而且随着建筑形态、建筑功能的多样化，越还原真实建筑的建筑能耗模型，在能耗模拟阶段所花费的计算时长就越长；此外，很多影响建筑能耗的参数无从知晓，或者实际值和设计值之间存在很大的差异，这些问题都会对构建一对一能耗模型的分析方法带来诸多困难。在实际节能潜力分析中，相较于围护结构保温性能的问题，工程师更关注建筑内热负荷的构成是否合理，因为投入产出比较高。

目前诸多学者对影响建筑负荷、能耗的因素进行深度分析，并归纳出对建筑负荷、能耗影响较大且不能忽视的因素；考虑到精细化白箱的难度以及建筑冷站总能耗数据的易获取性，因此本章以批量建模为基础，以推测影响建筑负荷的关键变量的方式对建筑内热负荷进行节能潜力，这样可以提高时间效率，降低节能潜力分析的专业门槛。

本章的技术路线如图 2.1，可以分为三大块，建筑-能耗算例生成，黑箱模型的超参数优化和关键变量的推测。建筑能耗算例生成主要是初期为黑箱模型的训练提供数据支撑，当数据库在积累过程中不断有新的真实数据纳入，由于训练数据不断变化，自动对黑箱模型的超参数优化必不可少。随后根据现有的且以获取的影响空调总能耗的关键参数和建筑冷站总能耗采用全局优化算法进行对不易获取的影响建筑负荷的关键变量进行推测，根据推测值和国标进行对比。

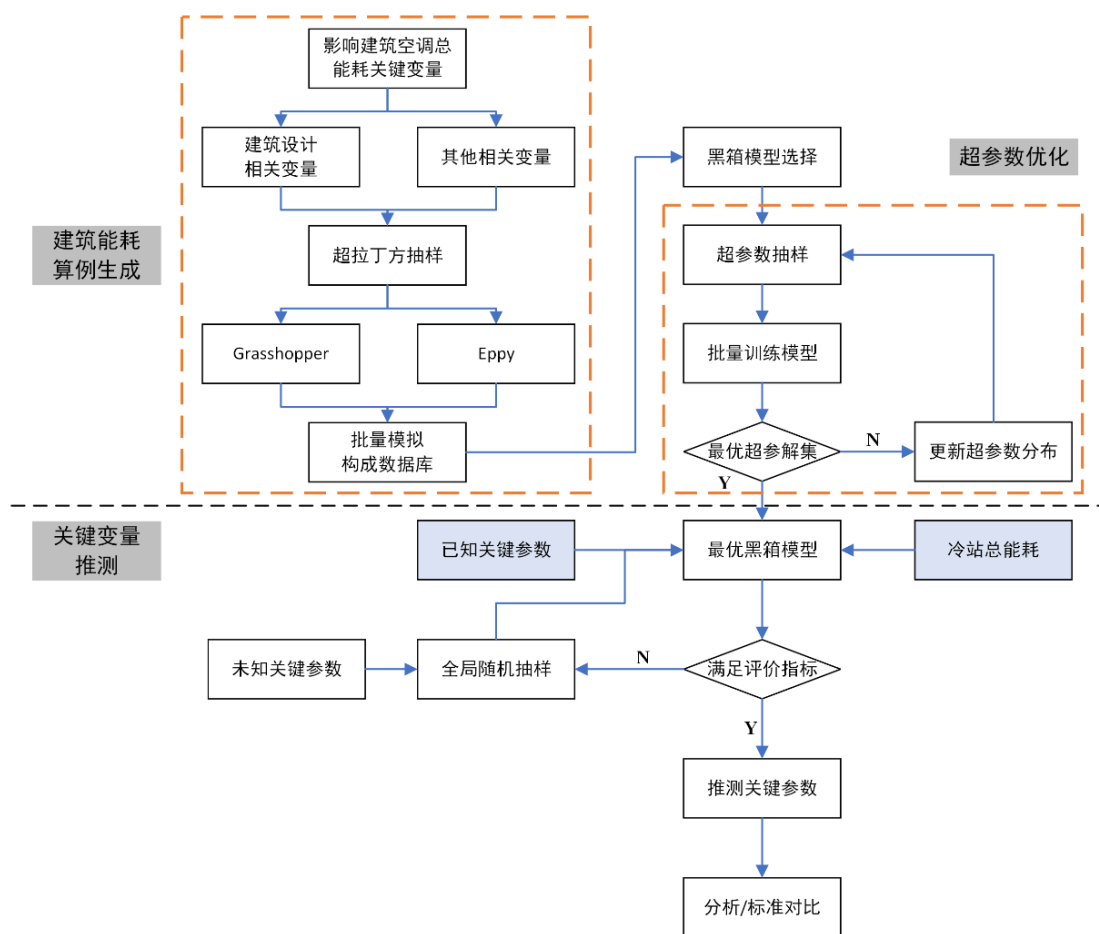


图 2.1 本章技术路线图

2.2 建筑-能耗算例设计

2.2.1 建筑能耗数据库构建

影响建筑能耗的因素很多，但是不同的因素重要程度有所差别，“抓住主要矛盾”是构建用于节能潜力分析的建筑能耗数据库的关键，用影响目标值的关键变量进行数据库是建筑能耗算例设计的基础。

朱明亚^[31]针对办公建筑在保证不同能耗预测精度情况下，用主成分分析法（Principal Component Analysis, PCA）分析建筑能耗预测的最小变量集，见表 2.1，沙华晶^[32]以五星级酒店为例采用敏感性分析（Sensitivity Analysis, SA）分析影响建筑负荷、不同冷站设备群分项能耗的关键变量群，其结论为：在夏热冬冷地区，对酒店类建筑，影响夏季冷负荷的关键变量为夏季空调设定温度、人员密度、新风渗透率、照明功率密度。从中不难发现，两位学者得出的结论

存在的较大的交集：与内热负荷相关的变量对建筑负荷的影响更大本章构建的建筑-能耗数据库的结构包括：关键变量、时间戳和冷站总能耗值。

表 2.1 最小变量集以 Cooling EUI 为目标在不同精度下最小变量集总结

模型方案	1	2	3
对应最小变量集名	CoolingEUI_Max	CoolingEUI_Med	CoolingEUI_Min
对应最小变量集包含变量	Plant 冷热源系统类型 HVAC 空调系统类型 trsp 水系统类型 bsc 体形系数 bldgA 建筑面积 sat 夏季平均温度 schd 建筑使用情景 epd 设备密度 wwr 窗墙比 sidt 夏季室内设计温度 ppd 人员密度 lpd 照明密度	Plant 冷热源系统类型 bldgA 建筑面积 sat 夏季平均温度 schd 建筑使用情景 epd 设备密度 wwr 窗墙比 sidt 夏季室内设计温度 ppd 人员密度	Plant 冷热源系统类型 bldgA 建筑面积 sat 夏季平均温度 schd 建筑使用情景 epd 设备密度 wwr 窗墙比
对应最小变量集的方法 1 变量选择依据	保留 78.5%方差解释度的全部变量	保留 71.4%方差解释度的最大荷载变量	保留 64.0%方差解释度的最大荷载变量
对应最小变量集的方法 2 变量选择依据	与 EUI 的相关系数 R>0.05, 不考虑冗余	与 EUI 的相关系数 R>0.05, 保证最小冗余	与 EUI 的相关系统 R>0.1, 保证最小冗余

2.2.2 建筑-能耗算例的批量生成

2.2.2.1 超拉丁抽样构成建筑算例

在确定了建筑-能耗数据库的结构之后，如何在保证数据丰富度的情况下填充这个数据库成为关键问题。理想情况下，用真实建筑的能耗数据填补较为合理，但是目前建筑行业信息化的一大现状是：研究人员较难获取较多细颗粒度、且高质量的能耗数据。所以本文采用计算机模拟作为基础，我们可以在由关键变量集组成的多维空间中均匀抽样生成多个建筑算例，然后通过 EnergyPlus 进行模型、填充建筑-能耗数据库，但是这会对计算机产生较大的计算负担，所以本文采用超拉丁方抽样代替均有抽样。

若用 χ 作为形容不同建筑在不同用能情况下的输入参数空间，若将其中一种建筑形态的一种用能情况视为一个高维向量 x_k ，其应保证 x_k 应均匀散布在 χ 中，即需要一种空间填充设计 (Space Filling Design)。记 $P = \{x_1, \dots, x_n\}$ 为一个有 n

个高维向量的空间填充设计。其中高维向量 $x_k = (x_{k1}, x_{k2}, \dots, x_{ks}) \in \mathcal{X}, k = 1, \dots, n$ 。计算 $y_k = f(x_k), k = 1, 2, \dots, n$ 。这是推测关键变量的基础。

空间填充设计往往采用实验设计方法 (Design of Experiments, DoE), 可以要因设计 (Factorial Design), 响应面设计 (Response-Surface Design) 以及随机设计 (Randomized Design) [31]。而随机设计中又包括完全随机抽样和分层随机抽样。本研究采用超拉丁抽样进行空间填充设计。

LHS 采用两部随机化, 其本质为分层抽样方法。取 s 个独立的 $\{1, 2, \dots, n\}$ 的随机置换 $\pi_j(1), \pi_j(2), \dots, \pi_j(n), j = 1, 2, \dots, s$, 组成一个 $n \times s$ 的矩阵, 其成为超拉丁方设计 (Latin Hypercube Design, LHD), 记为 $LHD(n, s)$, 它的第 k 行 j 列的元素记为 $\pi_j(k)$ 。取 $[0, 1]$ 上 ns 个均匀分布的独立样本, $U_{ij} \sim U(0, 1), i = 1, 2, \dots, n, j = 1, 2, \dots, s$, 记 $x_k = (x_{k1}, x_{k2}, \dots, x_{ks})^T$, 其中

$$x_{kj} = \frac{\pi_j(k) - U_{kj}}{n}, k = 1, 2, \dots, n, j = 1, 2, \dots, s \quad 2.1$$

则设计 $D = \{x_1, x_2, \dots, x_n\}$ 即为一个 LHS 设计, 并记为 $LHS(n, s)$ 。

2.2.2.2 基于 Grasshopper 软件快速建模

根据 HLS 方法抽样得到若干个样本点后, 如何根据样本点中的关键变量建立对应的建筑模型进行模拟成为亟待解决的问题。手工构建诸多建筑 idf 模型是一个极为耗时的过程, 为节省时间, 我们可以采用 Grasshopper 软件进行快速建模。

Grasshopper 是一种图形编程语言, 需要依赖 Rhinoceros 3D 计算辅助设计 (Computer-aided Design, CAD) 的环境运行, 通过在画布上拖拉组件 (components) 来进行变成, 可以不同组件之间环环相扣, 将其他组件的输出变为某一个组件的输入。Grasshopper 主要用来设计建筑生成算法 (Build Generative Algorithm), 其中很多的组件可能生成建筑几何的 3D 视图, 除此之外, Grasshopper 还包括其他种类的算法, 比如数字、文本、视听以及触觉应用 [33]。其非常适合参数化建模、光照分析以及建筑能耗分析。

本研究则是讲 LHS 抽样结果中与建筑形态相关的关键变量导入 Grasshopper 生成 idf 文件, 作为 EnergyPlus 模拟的基础。

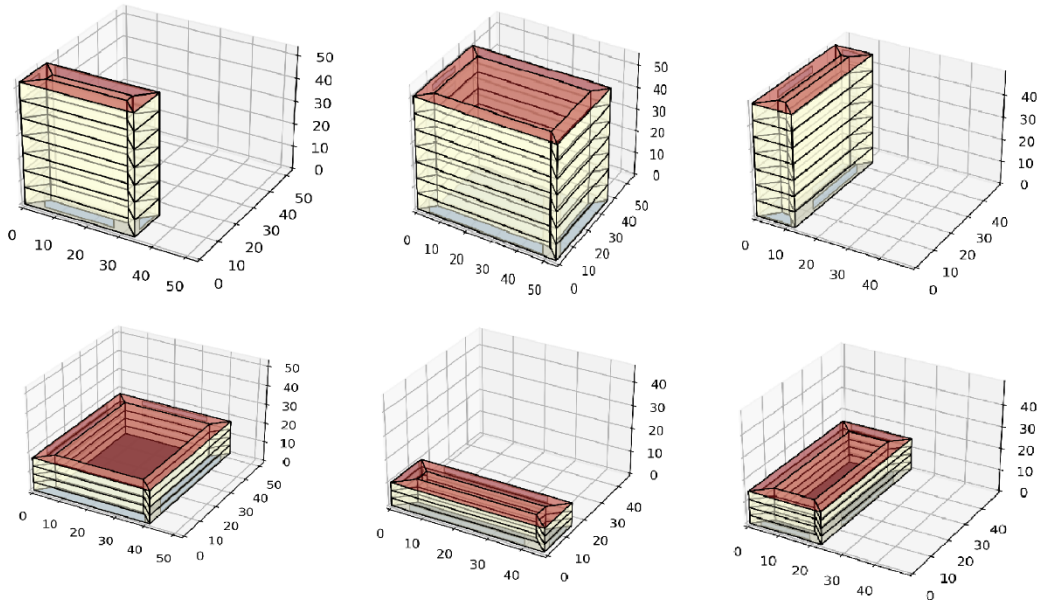


图 2.2 Grasshopper 生成的部分建筑外形简约可视化展示

2.2.2.3 基于 Eppy 对建筑-能耗算例批量模拟

Grasshopper 生成不同建筑外形的 idf 文件后，我们需要对与建筑外形无关的关键变量导入到建筑 idf 模型中，这里采用一种计算语言的脚本批量 Eppy 处理这一步工作。

Eppy 是一个专用于修改 EnergyPlus 的 idf 文件的 Python 脚本语言，用户可以根据自身的需要在 idf 文件里检索和修改信息。

2.3 用于推测关键变量的黑箱模型形成

2.3.1 黑箱模型选择

根据 EnergyPlus 批量模拟 LHS 抽样形成的算例填补完建筑-能耗数据库之后，选择合适的黑箱模型对建筑-能耗数据库中的数据进行拟合是下一步需要考虑的问题。

机器学习（Machine Learning, ML）的初期发展可以追溯到上个世纪，但是受限于当时的计算机硬件水平，机器学习并未得到大面的应用。目前随着计算机硬件以及超算水平的不断发展，机器学习的应用井喷式迸发。其中基于树的模型（Tree-based Models）是监督式机器学习（Supervised Machine Learning）的一种，其看起来是一棵倒立的树，其中它的节点表示一个特征变量，其节点之间的连接表示决策（Decision），每一个叶子节点表示一个输出值。决策树

(Decision Tree, DT) 可以分为分类树和回归树, 一般能耗预测常用的为回归模型, 在基于树的模型中则对应回归树模型。基于树的模型又可以分为单棵树模型和集成树模型 (Ensemble Method), 集成树模型是训练多个预测器

(Predictors), 集成树模型的泛化性和鲁棒性远胜过单棵树模型。

基于本文的研究目标, 本研究中主要采用的是 XGBoost^[34]和 LightGBM^[35]两种方法。

XGBoost (eXtreme Gradient Boosting) 极度梯度提升, 是一种基于 GBDT (Gradient Boosting Decision Tree, GBDT) 的算法, GBDT 是一种基于 Boosting 集成思想的加法模型, 训练时采用前向分布算法进行学习, 每次学习迭代都是学习 CART (Classification and Regression Trees) 拟合之前 $t-1$ 棵树的预测结果和训练样本之间的真实残差。

若训练数据集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, 损失函数 $l(y_i, \hat{y}_i)$, 正则化项 $\Omega(f_k)$, 整体目标函数为:

$$\mathcal{L}(\phi) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad 2.2$$

其中, $\mathcal{L}(\phi)$ 是线性空间上的表达, i 是第 i 个样本, k 是第 k 棵树, \hat{y}_i 是第 i 个样本 x_i 的预测值。

由于在 GBDT 梯度提升树表达方式为:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad 2.3$$

则 $\mathcal{L}(\phi)$ 转化成如下表达形式

$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \sum_k \Omega(f_k) \quad 2.4$$

XGBoost 将目标函数进行优化: 用二阶泰勒展开, 去除常数项, 优化损失函数; 随后正则化项展开, 去除常数项, 优化正则化项; 最后合并一次项系数, 二次项系数, 得到 XGBoost 最终目标函数:

$$\mathcal{L}^{(t)} = \sum_{j=1}^T \left[G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \lambda T \quad 2.5$$

其中, G_j 是叶子节点 j 所包含样本的一阶偏导数累加之和, 是一个常数; H_j 是叶子节点 j 所包含样本的二阶偏导数累加之和, 是一个常数; T 表示这棵树有 T 个叶子节点, w 表示第 t 棵树的权重向量 w 。

XGBoost 的改进点在于: 1) 相较于 GBDT 的一阶求导优化, XGBoost 采用泰勒二阶导展开优化; 2) XGBoost 在目标函数中引入正则化; 3) XGBoost 可以自动处理缺省值, 但是 GBDT 不允许出现缺省值; 4) 在 XGBoost 中寻找最佳分裂点的方法有贪心算法 (Greedy Algorithm)、近似算法 (Approximate

Algorithm)、加权分位数草图法 (Weighted Quantile Sketch) 以及稀疏感知法 (Sparsity-aware Split Finding) [32]。

虽然 XGBoost 轻量化、便捷、高效且灵活, 并在 Kaggle 竞赛中获得亮眼的成绩, 但 XGBoost 采用预排序算法选择分割点的缺点也非常明显, 在面对海量数据是空间消耗量巨大, 其原因是预排序算法不仅需要保存数据的特征值, 还需要保存特征排序的结果, 这样需要消耗训练数据两倍的内存。从时间角度上思考, 在遍历分割点的时候, 都需要进行分裂增益的计算, 代价巨大。

LightGBM 的提出是基于在不损害准确的情况下对上述 XGBoost 缺陷进行优化, LightGBM 采用下列思想进行优化:

- 1) 采用基于 Histogram 的决策树算法;
- 2) 采用单边梯度采样 Gradient-based One-Side Sampling (GOSS), 这样就可以仅仅只使用高梯度的数据去计算数据增益, 从而节省了时间复杂度和空间复杂度;
- 3) 采用 Exclusive Feature Bundling (EFB), 绑定互斥特征达到特征降维的目的;
- 4) 采用带深度的 Leaf-wise 的叶子生长策略, 对于分裂增益较低的叶子限制其叶子的生长;
- 5) 提高 Cache 命中率优化, 减少 Cache Miss 出现。

这两个集成模型均为轻量化模型, 因为目前笔者尚未由大量的实测数据, 需从模拟数据入手, 在数据数据尚未出现“维度灾难”时, 采用 XGBoost 模型和 LightGBM 模型训练数据的时间代价尚且不大, 所以两种集成模型均可以采用。

2.3.2 黑箱模型的超参数优化

数据驱动模型对超参数高度敏感, 默认的超参数集可能并不能将黑箱模型的精度提升到最大; 并且我们可以假设在后期数据积累的过程中, 当训练黑箱模型的建筑-能耗数据库的数据发生变化: 即真实建筑数据的加入, 黑箱模型需要对增量数据部分进行学习, 原先的超参数集可能不在适用, 需要重新优化超参数集, 见图 2.3。

所以未雨绸缪, 一种能够适应训练数据不断变化, 能够自动对黑箱模型的超参数进行调优的方法是十分必要的, 可以增加节能潜力分析的准确性和置信度, 考虑到本章撰写中章节间的匀称, 本节提到的自动超参数调优的技术细节将放在第三章中重点阐述。

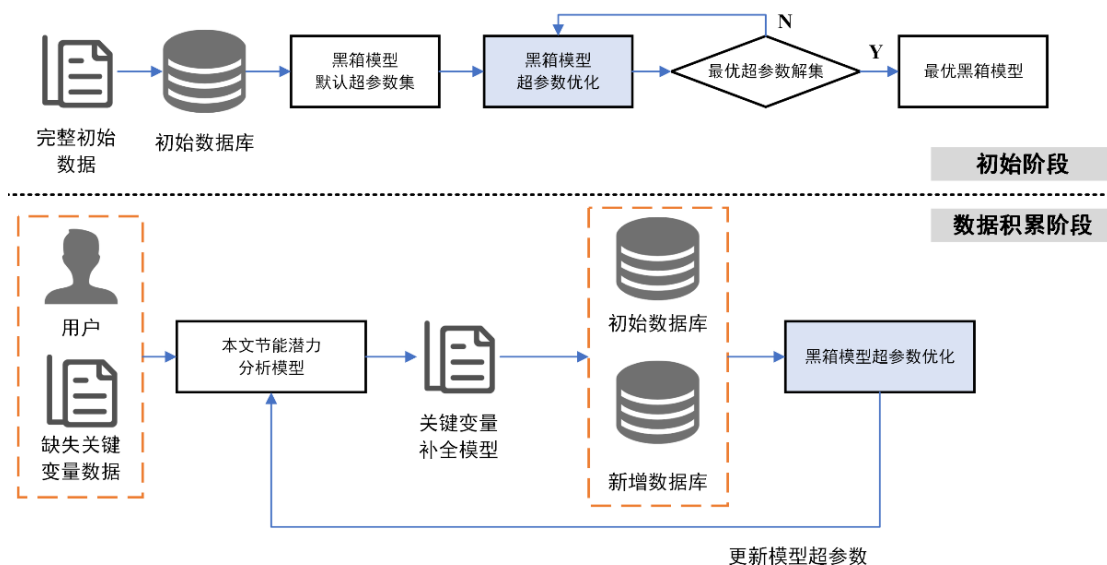


图 2.3 黑箱模型超参数优化与节能潜力模型分析之间的关系

2.4 基于关键变量推测的建筑节能潜力分析

2.4.1 基于全局搜索算法的关键变量推测

推测关键变量的黑箱模型以训练完成，接下来推测关键变量就成为这一章最后需要考虑的事情。关键变量推测可以视为集合约束的优化问题，笔者将该问题用数学语言表达，见公式 2.6：

$$\text{minimize } f(x) \text{ subject to } x \in \Omega \quad 2.6$$

其中，函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 成为目标函数或者价值函数，是一个实值函数，该优化问题的潜在含义便是我们需要找到合适的 x ，使得函数 $f(x)$ 达到最小。 x 是一个 n 维向量，表示为 $x = \{x_1, x_2, \dots, x_n\}^T \in \mathbb{R}^n$ ， x_1, x_2, \dots, x_n 相互独立，通常称之为决策变量。集合 Ω 是 n 维实数空间 \mathbb{R}^n 的一个子集，称之为约束集或者可行集 [37]。

在本研究中，上文的 x 指的是轻量化黑箱的输入特征向量，其维度 n 和黑箱的输入特征保持一致。其中 x 可能包含一个或者多个位置的关键变量，未知的关键变量如果仅有一个，可以保证其独立性，但若有多多个未知的关键变量需要推测，则难以完全保证：在任何一种缺失情况下的，缺失的关键变量之间是完全独立的。就本研究内容（建筑内热负荷构成是否合理），需要推测的关键变量有三：epd 设备密度、ppd 人员密度以及 lpd 照明密度（见表 2.1），这三个关键变量的真值难以获取，且三者之间确实存在一定的关系，但是抽象优化问题中，我们可以假设这三个变量之间相互独立，在后续的验证中笔者用案例论证，在

未严格遵循变量之间相互独立情况下，最后的预测结果的误差在合理可接受的范围内。

在确定了问题的性质后，解决该问题我们具象化的问题抽象为数学语言， $f(x)$ 为用关键变量真实值和抽样得到的关键变量值计算的能耗值之间的偏差，如果偏差越小，则说明抽样得到的关键变量值是接近关键变量真实值的。合集 Ω 中我们需要确定三个关键变量的取值范围，因为本研究中推测的关键变量是存在实际意义的，取值范围的约束可以一定减少出现陷入局部最优的困境。最后我们需要确定采用何种优化方法能够在较短时间推测出关键变量的值，且误差在可接受范围内。

优化方法中包括迭代算法和全局意义上的搜索算法^[37]，迭代算法包括梯度方法、牛顿法、共轭梯度法和拟牛顿法，能够从初始点出发开始，产生一个迭代序列。在很多时候迭代序列往往只能收敛到局部极小点，为了保证算法能够收敛到全局极小点，有时需要在全局极小点附近选择初始点，此外这些方法还需要计算目标函数的一阶导数，牛顿法中还需要计算二阶导数。而全局意义上的搜索算法可以对整个可行集中开展搜索，以找到极小点，这些方法只需要计算目标函数值，不需要对目标函数的求导，因此其泛化能力相对迭代算法更好。常用的全局搜索算法有单纯形法（Nelder-Mead），模拟退火法（Simulated Annealing），粒子群算法（Particle Swarm Optimization, PSO）以及遗传算法（Genetic Algorithm, GA）^[37]。

本研究采用更为通用的 PSO 和 GA 做推测，PSO 是一种随机搜索方法，由 James Kennedy 和 Russell C. Eberhart 提出^[37]，该方法是在社交互动原理的启发下得到的。在一次迭代中，PSO 并不是只更新单个迭代点，而是更新一群迭代点，称之为群，群中各点成为粒子。可将群视为一个无序的群体，其中每个成员都在移动，意在形成聚集，但移动方向时随机的，粒子群优化算法旨在模拟动物或者昆虫的社会行为，如蜂群、鸟群和羚羊群的形成过程。在介绍介绍 PSO 的算法步骤前^[37]，需要先定义每个粒子都是持续追踪其当前为之最好的位置，即截止到当前位置为止，它所经历的最好的位置（从目标函数值大小的意义而言）。称这种某个粒子相关的当前为止最好位置为个体最好位置 p_{best} ，全局最好位置为 g_{best} ；令 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 表示需要进行最小化的目标函数。 d 表示群体容量，群中各个粒子索引为 $i = 1, 2, \dots, d$ ，假设共有 d 个粒子。 $x_i \in \mathbb{R}^n$ 表示粒子 i 的位置，对应的速度为 $v_i \in \mathbb{R}^n$ 。 p_i 表示粒子 i 的 p_{best} ，相应的 g 表示 g_{best} ：

1) 令 $k := 0$ ，随机产生一个初始的粒子群，即产生 d 个粒子的位置 $x_i^{(0)}$ 及其对应的速度 $v_i^{(0)}$ ， $p_i^{(0)} = x_i^{(0)}$ ， $i = 1, 2, \dots, d$ ；令

$$g^{(0)} = \arg \min_{x \in \{x_1^{(0)}, \dots, x_d^{(0)}\}} f(x) \quad 2.7$$

2) 针对每个 $i = 1, 2, \dots, d$, 随机产生两个 n 维向量 $r_i^{(k)}$ 和 $s_i^{(k)}$, 按照均匀分布的原则抽取区间 $(0, 1)$ 中的随机数, 构成这两个向量的元素。令

$$v_i^{(k+1)} = \omega v_i^{(k)} + c_1 r_i^{(k)} \circ (p_i^{(k)} - x_i^{(k)}) + c_2 s_i^{(k)} \circ (g^{(k)} - x_i^{(k)}) \quad 2.8$$

$$x_i^{(k+1)} = x_i^{(k)} + v_i^{(k+1)} \quad 2.9$$

3) 针对每个 $i = 1, 2, \dots, d$, 如果 $f(x_i^{(k+1)}) < f(p_i^{(k)})$, 令 $p_i^{(k+1)} = x_i^{(k+1)}$; 否则, 令 $p_i^{(k+1)} = p_i^{(k)}$ 。

4) 如果存在 $i \in \{1, \dots, d\}$, 使得 $f(x_i^{(k+1)}) < f(g^{(k)})$, 则令 $i^* = \arg \min_i f(x_i^{(k+1)})$, $g^{(k+1)} = x_{i^*}^{(k+1)}$; 否则令 $g^{(k+1)} = g^{(k)}$ 。

5) 如果停止满足条件, 就停止迭代。

6) 令 $k := k + 1$, 回到第 2 步。

其中, 参数 ω 表示惯性参数, 参数 c_1 和 c_2 决定了粒子趋向于“好位置”的程度, 分别表示来自“认知”和“社会”部分的影响因素, 即粒子本身最好位置和全局最好位置对其运动的影响。在粒子群优化算法的不断完善后, 速度更新公式为:

$$v_i^{(k+1)} = \kappa \left(v_i^{(k)} + c_1 v_i^{(k)} \circ (p_i^{(k)} - x_i^{(k)}) + c_2 s_i^{(k)} \circ (g^{(k)} - x_i^{(k)}) \right) \quad 2.10$$

在实际应用过程中, 往往会指定一个速度上限 v_{max} , 这样算法中有关速度的部分应替换为:

$$\min\{v_{max}, \max\{-v_{max}, v\}\} \quad 2.11$$

遗传算法是另一种基于种群的随机搜索方法, 是在借鉴了遗传理论的基础上得到的^[37]。John Holland 首次提出遗传算法的基本理念。目前为止, 遗传算法已经在人工智能、最优化、神经网络训练有了广泛的应用。首先集合 Ω 中选定一组初始点, 用 $P(0)$ 表示, 代表初始种群。然后计算 $P(0)$ 中每个点对应的目标函数值, 基于计算结果, 产生一组新的点 $P(1)$ 。 $P(1)$ 的产生是源于对 $P(0)$ 的交叉和变异操作。循环展开这一过程, 产生种群序列 $P(2), P(3), \dots$, 直到达到停止条件。交叉和变异操作的目的在于创建一个新的种群, 使得种群目标函数的平均值能够大于上一代种群^[37]。

PSO 相较于 GA 迭代速度更快, 因为 PSO 不需要交叉遗传这些复杂的操作, 但是这两者均需要在效果和速度之中做权衡, 即需要去控制初始化粒子群和种群的数量。本文将比较 PSO 和 GA 这两种方法在推测关键变量中的效果。

2.4.2 衡量推测偏差的指标

上文中提到, 需要衡量根据关键变量真实值和抽样得到的关键变量值计算的能耗值之间的偏差, 那么如何衡量这两者之间的偏差呢?

一般采用模拟校准导则（Simulation Calibration Guidelines）中提出的指标来进行评价，例如 ASHRAE Guideline 14^[38]中的平均误差（Mean Bias Error, MBE）和变异系数均方根误差（Coefficient of Variation of Root Mean Square Error, CV-RMSE），这是专业研究人员常用的衡量模拟值与真实值之间差距的指标，计算公式见公式 2.12 和 2.13:

$$MBE = \frac{1}{m} \sum_{i=1}^m \frac{(I_i - y_i)}{y_i} \quad 2.12$$

$$CV - RMSE = \frac{\sqrt{m \sum_{i=1}^m (I_i - y_i)^2}}{\sum_{i=1}^m y_i} \quad 2.13$$

本文还考虑另外一种指标，因为建筑能耗数据本质为时间序列，判断不同时间序列相似度常用的另外一个指标为动态时间规整（Dynamic Time Warping, DTW）^[39]，尤其适用不同节奏的时间序列，DTW 使用归整路径距离（Warp Path Distance）来衡量两个时间序列之间的相似性。假设两个时间序列 X 和 Y ，长度分别为 $|X|$ 和 $|Y|$ 。规整路径的形式为 $W = w_1, w_2, \dots, w_k$ ，其中 $Max(|X|, |Y|) \leq k \leq |X| + |Y|$ 。 w_k 的形式为 (i, j) ，其中 i 表示为 X 中的 i 的坐标， j 表示为 Y 中的 j 的坐标。归整路径 W 必须从 $w_1 = (1, 1)$ 开始，到 $w_k = (|X|, |Y|)$ 结束，以保证时间序列 X 和 Y 中的每个点都在 W 中出现。另外 $w(i, j)$ 的 i 和 j 必须是单调增加的，见公式 2.14:

$$w_k = (i, j), w_{k+1} = (i', j'), i \leq i' \leq i + 1, j \leq j' \leq j + 1 \quad 2.14$$

点与点之间的计算距离常用的公式有曼哈顿距离（mahalanobis）、欧式距离（euclidean）、切比雪夫距离（chebyshev）等。

这两种衡量指标本文均会采用，和上述的两种全局随机算法进行两两组合。

2.5 本章小结

本章介绍了基于关键变量推测的建筑内热负荷分析的详细技术路线，包括每一环的技术细节，考虑到本文各章节内容的均衡性，本章仅介绍自动对黑箱模型超参数调优的必要理由，并未详细说明其技术细节，相关内容将会放至第三章。

第3章 基于轻量化黑箱模型的设备选型方案节能潜力分析

3.1 本章技术路线

在传统正向设计中，冷站设备的选型不仅需要考虑设计日的负荷还要需要考虑建筑负荷分布，这样能够简化实际运维过程中冷水机组的控制逻辑。但是在设计阶段，设计人员是无法得知建筑实际投入运行后的真实负荷分布，因为可能建筑的功能和用途有过改动，这会导致实际建筑负荷画像会和设计阶段的设计情况存在一定差距，这样会使设计阶段敲定的设备的装机容量和设备台数可能会与实际的建筑负荷不太匹配。

第二章使影响建筑负荷的关键变量均处于已知状态，在此基础上用黑箱模型推测建筑的真实负荷分布不是一件难事，实际建筑负荷分布有助于分析设计阶段的冷机选型方案是否合理。

图 3.1 是本章基于轻量化黑箱模型，对设备选型方案分析节能潜力的技术流程图，本章技术路线可以分为四大块，建筑-负荷算例生成，黑箱模型的训练，黑箱模型超参数调优，以及全局搜索分析设备选型方案。建筑负荷-算例生成的流程与第二章中建筑-能耗算例生成的流程类似；黑箱模型超参数调优的思路和第二章一致；最后根据推测的建筑负荷分布优化现有的设备选型方案，结合实际冷机模型库分析理想情况下的节能量。

此外需要提出的是，本章的冷水机组选型方案是针对既有建筑的，而非仍处于设计阶段的新建建筑。由于建筑冷站的设置不可以擅自更改，所以仅分析每台冷水机组的名义制冷量与建筑负荷的匹配程度，而并不更改冷水机组的台数。

3.2 建筑负荷算例生成

此处建筑=负荷算例生成的目的是填充建筑-负荷数据库，该数据的结构包括时间戳，影响建筑负荷的关键变量以及建筑负荷。

影响建筑负荷的因素可以分为与建筑形态结构相关的和与内热相关的两种类型的变量。这些关键变量在第二章中的建筑-能耗算例生成的章节中已做了详细的阐述，在这里就不过多赘述。

在确定了关键参数之后，即相关数据库的结构后，即采用和第二章中同样的思路：同样需要考虑采样的丰富性以及模拟计算的时间成本，本章同样采用 LHS 方法抽样。

随后根据 LHS 的抽样结果，分别采用 Grasshopper 和 Eppy 两个批量构建、处理 idf 文件的工具进行建筑负荷算例的批量生成和模拟，这里和第二章建筑-能耗算例不同的地方是：第二章中建筑能耗算例的模拟结果是建筑冷站的总能耗，而第三章中建筑-负荷算例的模拟结果是建筑负荷。

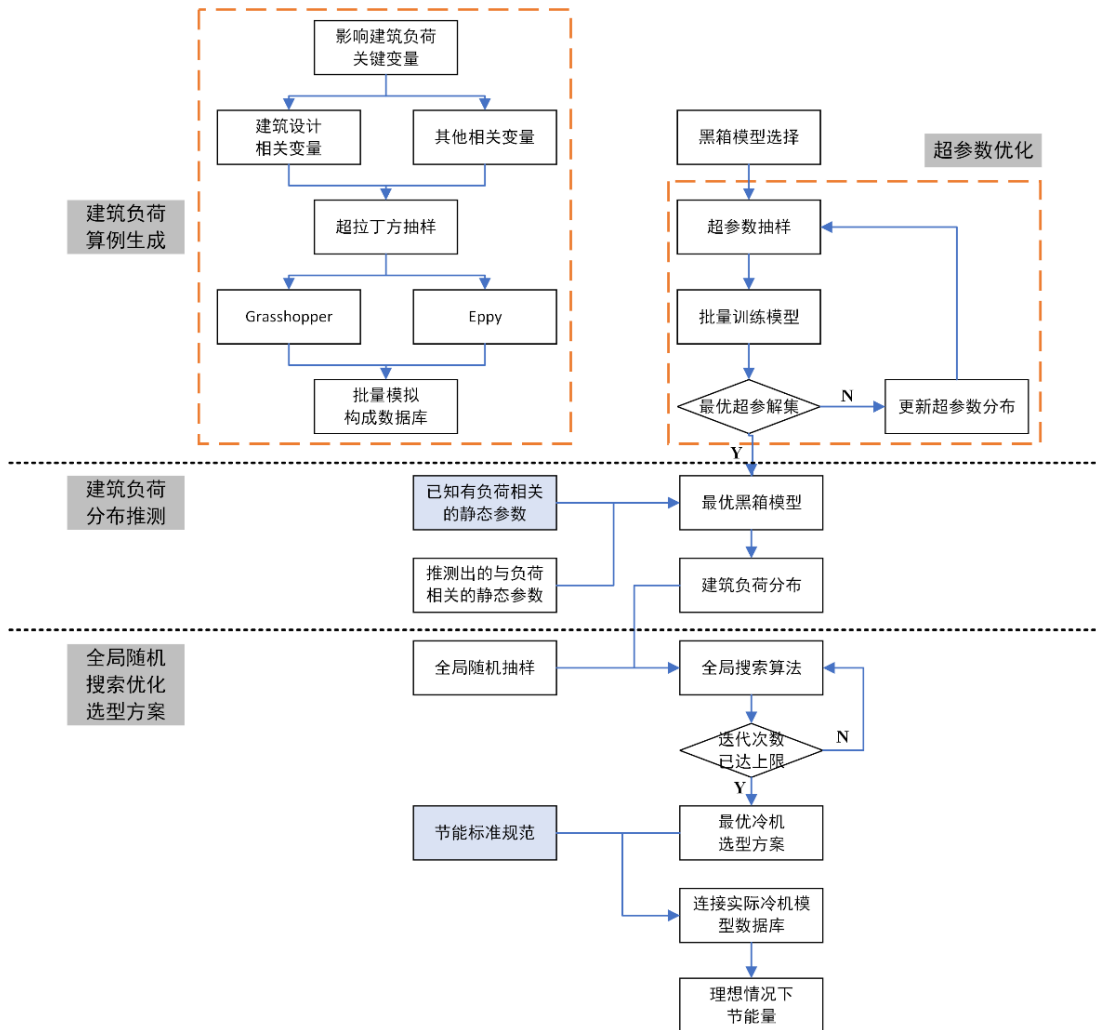


图 3.1 本章技术路线图

3.3 基于树模型的黑箱模型超参数调优方法

3.3.1 黑箱模型超参数调优框架

本章黑箱模型的选择沿用第二章选择的集成黑箱模型，因为该类模型在 Kaggle 竞赛和不同行业均有较为出色的性能。

黑箱模型的超参数调优一直是机器学习中的一个重要分支，当训练数据在不断变化和壮大的过程中，一组超参数是不足以使得模型在任何时候均达到最佳效果：假设当训练集从纯模拟数据变成真实模拟混杂的数据集后，初始模型的最优超参数可能并不能作为后续模型的最优超参数。网格搜索（Research Grid）是常用的超参数调优方法^[40]，当需要调节的超参数目不多时，其效率和时间复杂尚且可以接受，其调参的效率会随着需要调节的超参数数目增多而不断降低。

随着 ANN 的不断崛起，更高效的超参数调优方法也相继问世，Jasper S 等人^[41]采用高斯过程进行超参数调优，James B 等^[42]人采用基于树结构的 Parzen Estimator（TPE）对超参数调优，Google Vizier^[43]将剪枝（pruning）算法引入其中，监测每一轮迭代的结果，如果某个超参数集的黑箱模型的训练效果不及预期，则直接将该模型的训练中断，以加速整个调优的进程。

除了对超参数调优的方法不断优化外，还有其他学者利用分布式计算（distributed computing）的方法加速调参进程，例如 Philipp M 等人^[44]将分布式计算平台完成机器学习模型超参调优的过程。

超参数调优的效率主要有两个方面决定：1）搜索策略（searching strategy），2）性能预估策略（performance estimation strategy）。其中搜索策略决定了超参数的值，而性能预估策略则是预估当前超参数集的学习曲线（learning curve）然后决定是否需要取消用该对采用超参数的黑箱模型的训练，也是前文提到的剪枝。

根据前文笔者提到的抽样和剪枝的方法，展示超参数调优的技术路线图，见图 3.2。先采用独立抽样形成解空间，然后对解空间中每个解作评价，在此过程中采用剪枝算法对“不优秀”的解进行截断，分配更多资源给“优秀”的解，根据结果更新超参分布，根据新的超参数分布再抽样，再剪枝，按照上述步骤更新超参空间的分布然后循环。

后续小节中，笔者将详细介绍图 3.2 框架中两个主要部分：**抽样和剪枝**。

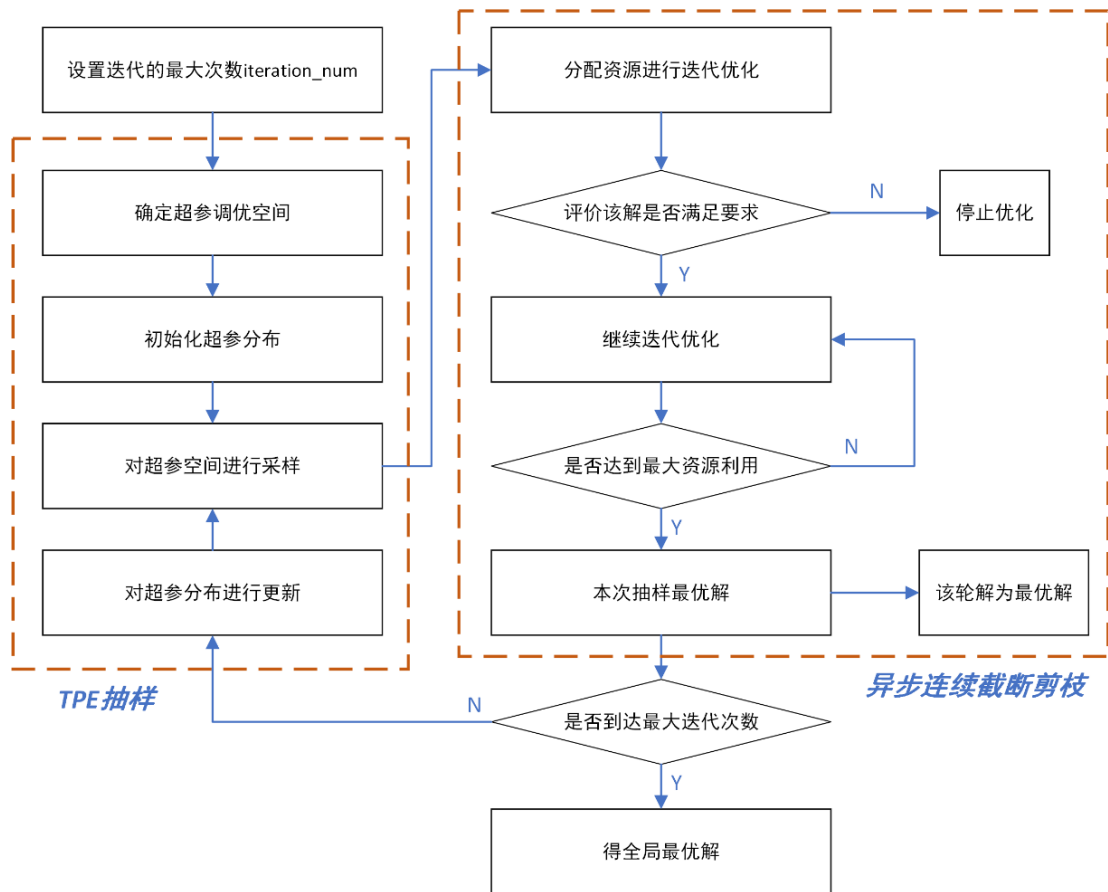


图 3.2 黑箱模型超参数调优框架

3.3.2 动态超参数空间抽样方法

此处提到的抽样方法 (Sampling Method) 与第二章提到的空间填充设计抽样并不相同, 这里提到的抽样方法主要可以分为 1) 相对抽样 (relative sampling), 2) 独立抽样 (independent sampling)。顾名思义, 相对抽样是探索超参数之间的相关性, 独立抽样则是不考虑超数参之间的相关性, 对每个超参进行独立抽样, 这里的独立抽样并不是一个简单概念, 例如 TPE^[42]在超参数相关性未知的情况下依旧可以表现优异。较为的热门的相对抽样方法是 CMA-ES^[45]。但是独立抽样和相对抽样的效率还是依赖于环境和任务。

3.3.2.1 TPE 抽样方法

超参数调优意味着在一个高维空间中寻优, 采用基于高斯过程的方法是直接模拟条件概率 $p(y|x)$, 但 TPE 是选择模拟全概率 $p(y)$ 以及 $p(x|y)$ 。TPE 选择把优化空间 (超参数构成的高维空间) 定义为树结构, 即只有在一个参数确定后, 才相继确定后续的参数。TPE 选择将优化空间中的先验分布用无参数分布 (non-parametric densities) 代替, 比如均匀分布 (uniform distribution), 对数均

匀分布 (log uniform distribution) 等。用这些无参数分布抽样形成 $\{x^{(1)}, x^{(2)}, \dots, x^{(k)}\}$, TPE 这样定义条件概率密度, 见公式 3.1:

$$p(x|y) = \begin{cases} \ell(x) & \text{if } y < y^* \\ g(x) & \text{if } y \geq y^* \end{cases}, \quad 3.1$$

其中分布 $\ell(x)$ 是根据 $\{x^{(1)}, x^{(2)}, \dots, x^{(k)}\}$ 对应的损失函数 (需要超参数调优的模型的损失函数) 的值 y 来定义的, 其中 $y < y^*$; 分布 $g(x)$ 则是根据 $\{x^{(1)}, x^{(2)}, \dots, x^{(k)}\}$ 来定义的。这里和基于高斯过程的方法有所不同的是: 基于高斯过程的方法对 y^* 的定义比较严苛, 一般 $y^* < \min \{y^{(1)}, y^{(2)}, \dots, y^{(k)}\}$ 。但是 TPE 对 y^* 的取值相对宽容, 可能 $y^* > \min \{y^{(1)}, y^{(2)}, \dots, y^{(k)}\}$ 且 $p(y < y^*) = \gamma$ 。

假设需要优化的函数的 EI, 其优化方向为使 EI 值就小化, TPE 中把 $p(x, y)$ 参数化为 $p(y)p(x|y)$, 这样加速对 EI 值的优化, 见公式 3.2:

$$EI_{y^*}(x) = \int_{-\infty}^{y^*} (y^* - y)p(y|x)dy = \int_{-\infty}^{y^*} (y^* - y) \frac{p(x|y)p(y)}{p(x)} dy \quad 3.2$$

通过构造 $\gamma = p(y < y^*)$ 和 $p(x) = \int_{\mathbb{R}} p(x|y)p(y) = \gamma \ell(x) - \ell(x) \int_{-\infty}^{y^*} p(y)dy$, 因此, 可得公式 3.3:

$$\int_{-\infty}^{y^*} (y^* - y)p(x|y)p(y)dy = \ell(x) \int_{-\infty}^{y^*} (y^* - y)p(y)dy = \gamma y^* \ell(x) - \ell(x) \int_{-\infty}^{y^*} p(y)dy \quad 3.3$$

所以最后可得公式 3.4:

$$EI^*(x) = \frac{\gamma y^* \ell(x) \int_{-\infty}^{y^*} p(y)dy}{\gamma y^* \ell(x) - (1 - \gamma)g(x)} \propto \left(\gamma + \frac{g(x)}{\ell(x)}(1 - \gamma) \right)^{-1} \quad 3.4$$

公式 3.4 表示对 x 的优化应该在 $\ell(x)$ 概率最大和 $g(x)$ 最小处, 树结构下的 g 和 ℓ 有利于根据 ℓ 提出很多和超参集, 然后根据 $g(x)/\ell(x)$ 来优化在每一次迭代下找出最佳 x^* 来优化 EI 值。

3.3.2.2 CMA-ES 相对抽样

CMA-ES 是全程协方差矩阵自适应进化策略 (Covariance Matrix Adaptation Evolution Strategy, CMA-ES), CMA-ES 从多变量高斯分布 (multivariate gaussian distribution) 抽样形成 $\{x^{(1)}, x^{(2)}, \dots, x^{(k)}\}$, 随后对每一组超参数进行评价, 这里的评价指标是指需要超参数调优的模型的 loss function, 根据每一组超参数的评价指标更新超参数空间的分布, 这样不断迭代的过程^[45]。但是 CMA-ES 抽样存在一定的限制:

- 1) CMA-ES 不支持类别参数 (标签参数), 如果超参数空间中存在类别参数, 建议使用 TPE 抽样。

2) CMA-ES 在高纬度空间抽样会有卓越的效果, 比如对于 ANN 或者 DL 中超过 100 个超参数需要调优的情况下, 会产生比较好的效果, 从另外一方面说, 当优化参数小于 100 的情况, 建议采用其他抽样方法进行超参数调优。

根据上述情况, 本研究中采用 TPE 抽样对 LightGBM 和 XGBoost 中超参数进行调优。

3.3.3 剪枝算法

3.3.3.1 连续剪枝算法

剪枝算法是否高效一定上决定了超参数调优的时间效率, 最初的剪枝算法包括两个部分:

- 1) 阶段性监测目标值的变化, 这里是指 loss function;
- 2) 中断没有满足预定义要求的尝试。

Jamieson 和 Talwalkar^[46]提出了连续截断算法 (Successive Halving), 见图 3.3。其算法背后的思想是先给每一个解配置一个很小资源, 在评价后保留最优的 $1/\eta$ 的解, 根据 η 给留下的解增配资源, 直到最后有解被分配到最大资源比例。在本研究中, 资源指的是随机梯度下降的迭代数。

```

1  Input 解数目  $n$ , 最小资源分配  $r$ , 最大资源分配  $R$ , 删减系数  $\eta$ , 最小停止迭代率  $s$ 
2   $s_{max} = \lceil \log_{\eta}(R/r) \rceil$ 
3  Assert  $n \geq \eta^{s_{max}-s}$  确保至少有一个解可以被分配  $R$ 
4   $T = \text{get\_hyperparameter\_configuration}(n)$ 
5  For  $i \in \{0, \dots, s_{max} - s\}$  do
6    // 在给定  $i$  的情况对每一个解进行训练, 并定义为一个 rung
7     $n_i = \lceil n\eta^{-i} \rceil$ 
8     $r_i = r\eta^{i+s}$ 
9     $L = \text{run\_then\_return\_val\_loss}(\theta, r_i): \theta \in T$ 
10    $T = \text{top\_k}(T, L, n_i/\eta)$ 
11 End
12 Return 在  $T$  中找到最优解

```

图 3.3 连续截断算法伪代码图^[46]

3.3.3.2 异步连续截断算法

在连续截断的算法基础上, Liam L^[47]提出了异步连续截断 (Asynchrony Successive Halving) 算法, 其算法伪代码见图 4.4, 其背后的思想是充分利用异

步性最大化并行计算能力，在异步连续截断中所有的解都被推进到下一个 run g，如果先前所有的解都没有达到评估要求，直接加入新的解，这样就能有更多的解进入到多轮 run g，这里并不要求用户预先定义需要评估的解的数量，但是其他的输入是与连续截断算法是一致的。 $run_then_return_val_loss$ 这块是采用并行计算。本研究中采用异步连续截断算法进行剪枝。

```

1  Input 最小资源分配 $r$ ，最大资源分配 $R$ ，删减系数 $\eta$ ，最小停止迭代率 $s$ 
2  Algorithm Asynchrony Successive Halving()
3    Repeat
4      For each free work do
5         $(\theta, k) = get\_job()$ 
6         $run\_then\_return\_val\_loss(\theta, r\eta^{s+k})$ 
7      End
8      For completed job $(\theta, k)$  with loss  $l$  do
9        更新解 $\theta$ 在第  $k$  轮 $run$ g以 loss  $l$ 
10     End
11  Procedure  $get\_job()$ 
12    // 一个解是可提升的是指在这个解还没有充分利用到资源 $R$ 是其评价指标就已经达到前 $1/\eta$ ，这也就是说明这个解还未被充分开发
13    假设 $\theta$ 就是最优潜力的解，目前已经训练到第  $k$  轮 $run$ g
14    If  $\theta$  exists then
15      Promote  $\theta$  to  $run$ g  $k + 1$ 
16      Return  $\theta, k + 1$ 
17    Else
18      增加一个新的解到目前当前 $run$ g
19      Return  $\theta, 0$ 
20    End

```

图 3.4 异步连续截断伪代码^[47]

3.4 基于全局搜索优化选型方案的建筑节能潜力分析

3.4.1 全局搜索算法

根据上述的黑箱模型可以得到目标分析建筑的建筑负荷分布，随后采用全局搜索算法分析冷机选型方案和建筑负荷分布的匹配程度是否可以进一步提高。

本章采用的全局搜索算法与第二章的算法一致，采用 GA 和 PSO 分别进行冷机选型方案的优化。但第二章不同之处在于本章的衡量方案的指标和第二章介绍的 DTW 和 CV-RSME 衡量冷站总能耗预测准确度的指标并不相同，在后续会详细介绍。

3.4.2 衡量方案优化的指标

本块采用平均 PLR (\overline{PLR}) 最高作为方案优化的指标，见公式 3.5:

$$\overline{PLR} = \frac{\sum_{i=1}^n load}{n}, n \in N^+ \quad 3.5$$

其中 $load$ 表示一年中的逐时负荷，单位 KW； n 表示时间，是一年中需要供冷的时长，以小时为单位。

逐时 PLR 计算涉及到每个小时的冷机启停表，本章研究是针对设计阶段的设备选型方案的评价和优化，所有采用理想的冷机启停表，即每个时刻开启冷机的名义制冷能力可以满足该时刻的冷负荷需求，但关闭或更换或添加其中任意一台开启的冷机将会制冷能力的浪费或不足，即 PLR 降低，见公式 3.6-3.8:

$$PLR = \frac{load_{actual}}{load_{chillers_small}} \quad 3.6$$

$$load_{chillers_small} = \min\{load_{chillers1}, load_{chillers2}, \dots, load_{chillersk}\} \quad 3.7$$

$$\sum load_{chillers_small} \geq load_{actual} \quad 3.8$$

其中 PLR 代表逐时部分负荷率； $load_{chillers_small}$ 表示为处理当前冷负荷所需要的最少冷机组合的名义制冷量之和； $load_{chillersk}$ 表示第 k 台冷水机组的名义制冷量，假设 $k = 3$ ，说明处理当前负荷所需要开启的冷水机组为 3 台，这是满足处理当前冷负荷的所有冷水机组启停的排列组合中的一种情况，那么针对每一个时刻建筑负荷的冷机启停的排列组合的数量 num_k 见公式 3.9:

$$num_k = C_3^1 + C_3^2 + C_3^3 \quad 3.9$$

以此类推可得，假设有 q 台冷机，则 num_k 的计算见公式 3.10:

$$num_k = \sum_{i=1}^q C_q^i, i = 1, 2, 3, \dots, q \quad 3.10$$

每个时刻的理想冷机启停状态，即是在这些排列组合中选择能够满足当前时刻负荷的最小冷机启停组合的名义制冷量之和。

3.4.3 以真实冷机为模板分析节能潜力

设计阶段的选型方案不涉及具体冷水机组设备的选择，即不知道具体冷机的性能；在缺少冷水机组性能模型的情况，是无法离线计算选型方案优化的节能潜力。这里采用 DOE-2.1 多项式拟合的冷机模型作为计算关于设备选型方案的节能潜力的基础。

DOE-2.1 软件中的冷机模型 EIR 是基于电压缩式冷机建立的^[48]，用来模拟冷机的制冷性能和压缩机电耗，EIR 中包括三条曲线：

- 1) 制冷能力-温度曲线
- 2) 能量输入与冷量输出比-温度曲线
- 3) 能量输入与冷量输出比-部分负荷率曲线

这三条曲线分别对应公式 3.11、3.12、3.13：

$$\begin{aligned} \text{ChillerCapFTemp} = \\ a + bT_{cw,l} + cT_{cw,l}^2 + dT_{cond,e} + eT_{cond,e}^2 + fT_{cw,l}T_{cond,e} \end{aligned} \quad 3.11$$

其中 ChillerCapFTemp 是指的制冷能力系数，当 $\text{ChillerCapFTemp} = 1$ 则说明处于工况； $T_{cw,l}$ 指离开冷机的冷冻水温度，单位℃； $T_{cond,e}$ 指进入冷机的冷却水温度，单位℃。

$$\begin{aligned} \text{ChillerEIRFTemp} = \\ a + bT_{cw,l} + cT_{cw,l}^2 + dT_{cond,e} + eT_{cond,e}^2 + fT_{cw,l}T_{cond,e} \end{aligned} \quad 3.12$$

其中 ChillerEIRFTemp 指能量输入和冷量输出比系数，当 $\text{ChillerEIRFTemp} = 1$ 说明处于工况下； $T_{cw,l}$ 是离开冷机的冷冻水温度，单位℃； $T_{cond,e}$ 指进入冷机的冷却水温度，单位℃。

$$\begin{aligned} \text{ChillerEIRFPLR} = a + b\text{PLR} + c(\text{PLR})^2 \\ = \frac{P_{chiller}}{P_{ref}} (\text{ChillerCapFTemp})(\text{ChillerEIRFTemp}) \end{aligned} \quad 3.13$$

其中 ChillerEIRFPLR 指能量输入和冷量输出之间的比值，当 $\text{ChillerEIRFPLR} = 1$ 说明处于工况下； PLR 指部分负荷率，指实际负荷和冷机装机容量之间比值； $P_{chiller}$ 指在 PLR 工况下的冷机功耗； $P_{ref} = \dot{Q}_{ref}/\text{COP}_{ref}$ 。

从中不难看出， PLR 是影响冷水机组能耗的关键因素，冷水机组的装机容量和数量都会对 PLR 产生较大的影响，这也间接验证了采用平均 $\overline{\text{PLR}}$ 作为衡量指标的合理性。

另外一个确定冷机模型的关键参数是制冷系数 (COP, Coefficient of Performance)，在 GB 55015-2011 《建筑节能与可再生能源利用通用规范》^[65] 中对不同制冷能力的冷水机组根据风冷和水冷的形式分别规定了性能下限，由此在实际冷机库模型中匹配最接近标准下限的模型，以此分析现有的设备选型方案的节能潜力。。

3.5 本章小结

本章详细介绍了关于设备选型优化的节能潜力分析的具体技术路线和技术细节，着重介绍了黑箱模型超参数优化的技术路线。

第4章 基于实测分项计量数据的冷站运维节能潜力分析

4.1 本章技术路线

既有建筑信息化程度不同，若想定义用相同的数据结构对不同建筑进行全方面、无差异的建筑节能潜力分析是相对困难的，因为不同建筑的数据差异性大。用模拟数据填充真实数据的空白，用模拟数据进行颗粒度的转化，这类数据融合方法目前只能填补暖通空调建筑总能耗等“宏观”数据。笔者认为这类数据填充和混合是一种“治标不治本”的方法，这种方法在一定程度上纵容了目前建筑能耗计量平台或BAS数据良莠不齐的问题，倘若一味用填充方法补全本身可以实际监测得到的数据，这样可能会导致目前兴起的建筑运维相关的物联网平台没有发挥其应有的价值。

在另一方面，面对目前既有建筑没有分项电表的情况，由学者研究了拆分“宏观”数据进行详细的建筑运维的能效诊断，这样拆分方法可以主要分为两类：非侵入式监测法（Non-Intrusive Monitoring, NIM）和侵入式监测法

（Intrusive Monitoring, IM），根据拆分目标的不同，比如针对负荷拆分的方法又可以分为非侵入负荷监测法（Non-Intrusive Load Monitoring, NILM）^[49]和侵入式负荷监测法（Intrusive Load Monitoring, ILM）；对其他不同的拆分目标也是同样如此。IM顾名思义需要另外安装电表或其他监测设备到BAS或者监测系统单独监测目标数据，而学术界将更多的经历投入NIM，例如从总电耗中拆分出不同设备的能耗。或者从总的建筑热负荷中拆除新风负荷、外部得热、内部散热等。基于事件驱动的非侵入式负荷拆分方法（Event-Based Non-Intrusive Load Monitoring, EBNILM）受到较多学者的研究，其核心在于需要知道不同的设备的启停以及工作状态转换的电信号^[50]。但是NILM大多都集中在对于住宅建筑的研究，主要是因为住宅建筑中的主要用电设备的运行特性相对规律，例如电冰箱为间歇性工作，而商业建筑中相对少的采用NILM的方法进行拆分，主要是其用电设备复杂且多样，设备使用的不确定性会因为人员的不确定性而增加，Alan M等人^[51]研究了在商业建筑中使用NILM的局限性：

- 1) 商业建筑中存在若干小和类似的负荷（在用能上情况类似），会增加识别和量化的难度；
- 2) 能耗上微小的变化会提供模糊且难以辨认的信息，进一步增加识别的难度；
- 3) 很多分项电表对于当前的INLM是难以监测到的，那么解决这种“盲区”可能会进一步INLM的侵入性。

基于上述学者的研究笔者认为应当优化和维护现有的能耗数据监测平台，直接采用实测数据进行细致的节能潜力分析，本文提出一种关于冷站运维的分层节能潜力分析的框架，考虑到不同建筑能耗数据的差异性，见图 4.1。

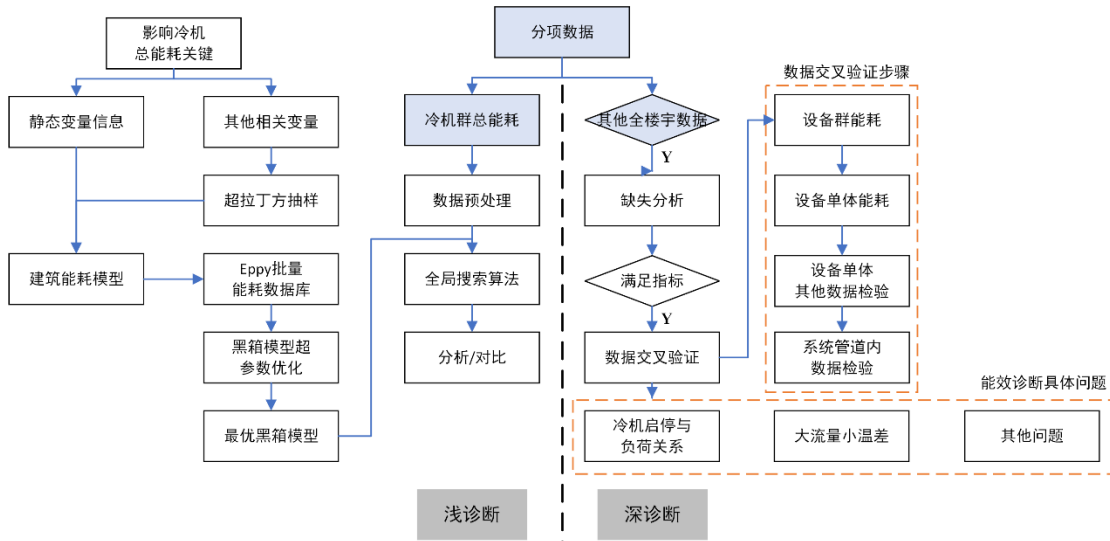


图 4.1 本章技术路线图

在本章中，分层分析可以分为图 4.1 中的“浅诊断”和“深诊断”：

- 1) “浅诊断”是用冷机群总能耗判断是否冷机自身性能存在劣化；
- 2) “深诊断”是用详细的分项能耗数据进行详细的能效诊断，也就是第一章中的“微观”分析。

“浅诊断”是依据关键变量真实值的推断进行分析，“深诊断”是对分项数据进行清洗，用免模型的统计学方法进行能效诊断。

“浅诊断”和“深诊断”既可以是并行关系也可以是先后关系，这取决于数据情况。

4.2 浅诊断

4.2.1 浅诊断中关键变量的选择

沙华晶^[32]根据标准化秩回归系数(Standardized Rank Regression Coefficient, SRRC)和偏相关系数(Partial Rank Correlation Coefficient, PRCC)的结果认为影响冷机能耗的关键参数是冷机 COP、风系统类型、水系统类型。后两者是无法更改的，本文将冷机 COP 分为冷机 COP 和冷机劣化系数，这两个参数是本章需要推断的关键参数。

冷机劣化的一部分原因是沉积物沉积在冷凝器表面，这通常是水质不佳造成^[48]。结垢会导致冷水机组内换热出现一定的问题，间接会导致排气温度，或者排气压力等参数指标达不到设定要求，从而进一步在不同的 PLR 下出现不同程度的性能劣化。而这样的结垢问题不会出现在不需要水为换热媒介的吸收式冷水机组内，冷机劣化系数会降低冷水机组的名义制冷量，当劣化问题特别严重时，冷水机组可能并不能处理设计建筑负荷。冷机劣化系数的影响见公式 4.1 和 4.2^[48]：

$$Q_{chiller,f} = Q_{chiller,ff} \times F_{chiller} \quad 4.1$$

$$COP_{chiller,f} = COP_{chiller,ff} \times F_{chiller} \quad 4.2$$

其中， $Q_{chiller,f}$ 是指在冷机在劣化情况下的名义制冷量；

$Q_{chiller,ff}$ 是指冷机在没有劣化或者是清洗保养之后的名义制冷量；

$COP_{chiller,f}$ 是指在冷机在劣化情况下的名义 COP；

$COP_{chiller,ff}$ 是指在冷机在没有劣化或者是清洗保养之后的名义 COP；

$F_{chiller}$ 是指冷机劣化系数，其中 $F_{chiller} \in [0,1]$ ，当 $F_{chiller} = 1$ 时，表示冷机没有存在劣化问题；当 $F_{chiller} = 0$ 时，表示冷机彻底报废。

4.2.2 辅助推断关键变量的最优黑箱模型形成

4.2.2.1 建筑-冷机能耗数据库形成

这一块的整体思路类似于第三章中建筑-能耗算例的构建以及第四章中建筑-负荷算例构建，用 LHS 方法对关键变量进行抽样，随后用 Grasshopper 和 Eppy 工具包批量处理 idf 文件，拟形成建筑-冷机能耗数据库。

4.2.2.2 最优黑箱模型的形成

根据建筑-冷机能耗数据库作为训练数据，训练黑箱模型，黑箱模型的选择见第二章，黑箱模型超参数的优化见第三章。

4.2.3 分项能耗的数据预处理

4.2.3.1 分项能耗的常见问题

在用冷机群能耗推测关键变量时，需要注意冷机群能耗严格意义上算分项计量数据，和总能耗相比，分项计量数据更易发现数据质量问题，如图 4.2，一般出现的问题如下：

- 1) 均值漂移；
- 2) 大量缺失值；
- 3) 信噪比较小。

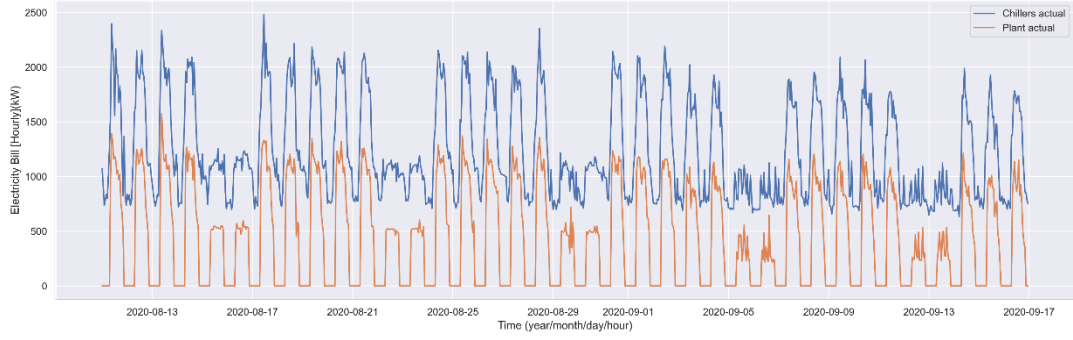


图 4.2 某真实建筑的冷机群能耗数据和冷站总能耗数据对比

建筑能耗数据曲线中的毛刺是因为传感器的质量问题，传感器的质量问题需要现场观察和调研，在这里采用信噪比（Signal-to-Noise Ratio, SNR）这一概念来衡量传感器传输的数据质量，SNR 用来比较信号强度和噪声强度，SNR 被定义为信号强度和噪声强度的比值（见公式 4.3），信号强度单位通常用分贝（dB）表示，见公式 4.4-4.8。当强度大于 0dB 且比值大于 1:1 时说明信号强度是大于噪声强度的。SNR 越低说明噪声强度和信号强度越相近，也就说明信号越不可用；SNR 数值越高，噪声强度相对越小，一般情况 SNR 在[5,20]区间内是较差的情况。

$$SNR = \frac{P_{signal}}{P_{noise}} = \left(\frac{A_{signal}}{A_{noise}}\right)^2 \quad 4.3$$

$$P_{signal,dB} = 10\log_{10}(P_{signal}) \quad 4.4$$

$$P_{noise,dB} = 10\log_{10}(P_{noise}) \quad 4.5$$

$$SNR_{dB} = 10\log_{10}(SNR) \quad 4.6$$

$$SNR_{dB} = 10\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) = 20\log_{10}\left(\frac{A_{signal}}{A_{noise}}\right) \quad 4.7$$

$$10\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) = 10\log_{10}(P_{signal}) - 10\log_{10}(P_{noise}) \quad 4.8$$

其中 P 为功率， A 为振幅。

4.2.3.2 基于 Savitzky-Golay 滤波的能耗时间序列降噪

在介绍 Savitzky-Golay 滤波器之前，现介绍高斯噪声（Gauss Noise），在后续章节，人为将高斯噪声加入真实冷机群总能耗数据中，验证该滤波器在极端情况下（传感器的传输质量很差，SNR 较低，噪声功率较高的情况下）可以有效降低 CV-RMSE，给数据降噪。

高斯噪声是指其概率密度函数服从高斯分布（即正态分布）的一类噪声，高斯噪声是一种常见的噪声。高斯噪声会在时间序列的每一个点上都产生噪声。

根据已知的 SNR 和真实信号的功率可以产生固定 SNR 的高斯噪声^[52]。假设信号功率 P_{signal} 和 P_{noise} 为公式 4.9 和 4.10:

$$P_{signal} = \frac{\sum_{i=1}^N (x_i)^2}{N} \quad 4.9$$

$$P_{noise} = \frac{\sum_{i=1}^N (n_i)^2}{N} \quad 4.10$$

其中 N 表示时间序列（信号或噪声）的长度。生成高斯噪声是已经确定了信噪比，即需要对噪声数据乘上一个数据 k ，其功率满足要求，见公式 4.11-5.14。

$$SNR = 10 \log_{10} \frac{P_{signal}}{P_{noise}} \quad 4.11$$

$$\begin{aligned} P_{noise} &= \frac{P_{signal}}{\frac{SNR}{10^{10}}} = \frac{\sum_{i=1}^N (kn_i)^2}{N} = \frac{(kn_1)^2 + (kn_2)^2 + \dots + (kn_N)^2}{N} \\ &= \frac{k^2(n_1^2 + n_2^2 + \dots + n_N^2)}{N} = k^2 P_{noise} \end{aligned} \quad 4.12$$

$$k^2 P_{noise} = \frac{P_{signal}}{\frac{SNR}{10^{10}}} \quad 4.13$$

$$k = \sqrt{\frac{P_{signal}}{\frac{SNR}{10^{10}} P_{noise}}} \quad 4.14$$

Savitzky-Golay 滤波器（S-G 滤波器）被广泛应用于数据流平滑除噪，是一种在时域内基于局域多项式最小二乘法拟合的滤波方法，这种滤波最大的特点在于在滤除噪声的同时可以保持信号的形状，和宽度不变。

设一个窗口内的一组数据为 $x[i], i = -m, \dots, 0, \dots, m$ ， i 的取值为 $2m + 1$ 个连续的整数值，现构造一个 n 阶多项式($n \leq 2m + 1$)来拟合这组数据，见公式 4.15:

$$f(i) = \sum_{k=0}^n b_{nk} i^k + b_{nn} i^n \quad 4.15$$

拟合数据点与原数据点的残差平方和见公式 4.16:

$$E = \sum_{i=-m}^m (f(i) - x[i])^2 = \sum_{i=-m}^m \left(\sum_{k=0}^n b_{nk} i^k - x[i] \right)^2 \quad 4.16$$

使用最小二乘法，要让拟合结果最好，则残差平方和应最小，也就是 E 对多项式的系数 b_{nk} 求偏导应为 0，见公式 4.17:

$$\frac{\partial E}{\partial b_{nk}} = 2 \sum_{i=-m}^m \left(\sum_{k=0}^n b_{nk} i^k - x[i] \right) i^k = 0 \quad 4.17$$

解得公式 4.18:

$$\sum_{i=-m}^m x[i]i^r = \sum_{k=0}^n b_{nk} \sum_{i=-m}^m i^{k+r} \quad 4.18$$

当需要拟合的单边点数 m ，多项式的阶次 n ，以待拟合的数据 $x[i]$ 已知后，即可求出该多项式。拟合出来的多项式用来求取该窗口内的中心点估计值，而对于后面的点，不断移动窗口重复操作即可。进一步求解后可以发现公式 4.19:

$$\begin{cases} f(i)|_{i=0} = 0! b_{n0} = a_{n0} \\ \frac{df(i)}{di} |_{i=0} = 1! b_{n1} = a_{n1} \\ \dots \dots \\ \frac{d^n f(i)}{di^n} = n! b_{nn} = a_{nn} \end{cases} \quad 4.19$$

可以看到计算 a_{nn} 相当于对原始数据进行一次 FIR 滤波，意味着输入与滤波器的单位冲击响应进行卷积就可以得到输出，即公式 4.20:

$$a_{nn} = \sum_{i=-m}^m h[i]x[n-i] \quad 4.20$$

Savitzky 与 Golay 根据这一特点制定了卷积系数表，根据系数使用离散卷积既可以快速得到该窗口内的中心点平滑值。平滑公式见公式 4.21:

$$x_{k,smooth} = \bar{x}_k = \frac{1}{H} \sum_{i=-w}^{+w} x_{k+1} h_i \quad 4.21$$

其中， $\frac{h_i}{H}$ 为平滑系数，由最小二乘法拟合多项式求得。

4.2.4 基于关键变量推测的浅诊断能效诊断分析

本部分的思路类似于第二章中的内容，根据上述**黑箱模型和降噪后的冷机群能耗数据推测关键变量**，全局搜索算法和衡量推测偏差的指标也均相同，不同之处在于推测的关键变量发生了变化，以及第二章中衡量推测偏差是以冷站总能耗为准，但是该部分以降噪后的冷机群总能耗为准，其他技术细节不在此处做过多的赘述。

4.3 深诊断

4.3.1 分项计量数据的数据质量问题概述

随着数字化浪潮、碳中和计划的提出，建筑行业信息化不断迅猛发展，不论是数字孪生技术的应用、还是元宇宙（Metaverse）在建筑行业的应用，均需要分项计量数据的支撑，这里的分项计量数据框定范围在建筑冷站内监测的数据，如冷冻水供回水温度、冷冻水流量、单台设备能耗等。

通过表 4.1 可以看出，当 BAS 或者建筑能耗监测平台提供的分项计量数据可靠，那么关于冷站运维的节能潜力分析的专业门槛会有所降低。但是，目前建筑用能系统，以建筑内 HVAC 系统为例，分项计量数据在传输和储存上存在一系列不容忽视的问题：

- 1) 时间序列中存在大面积缺失，时间序列中连续大面积出现-9999；
- 2) 时间序列在不合理的范围内波动；
- 3) 数据之间物理上强耦合性得不到满足，例如冷冻侧和冷却测数据无法满足能量守恒定律；
- 4) 数据和对应标签存在错位，例如冷却塔#1 的频率的监测点的数据标签是冷却塔#2 的频率；

孟欣^[53]总结建筑能源监测系统的数据特征：变量多、状态多、耦合性强、长期处于变工况。目前基于的真实分项计量数据的研究多侧重于数据应用，而对数据本身存在的问题的分析不多，而且数据的修补也仅针对数据的缺失^[53,54]，默认不缺失的、并且波动看似正常的的数据就视为能最大程度客观反映系统真实运作的的数据。在有关建筑能耗、冷负荷预测的数据比赛中，参赛者也依赖专家知识去识别建筑能耗监测系统中的异常值^[55]，由此可见分项数据的数据清洗还未形成相对通用的框架。

在工程中，并不是所有的分项数据同等重要，如设备群的电耗数据可能相对更加重要，因为这是缴费的重要依据。由此不难看出，数据质量可能会因为重要性的不同而不同，重要数据中的缺失值较少，相对次要的数据质量可能相对较差。本文数据清洗思路是：选出质量上乘的数据作为标定，利用物理规律修正其他数据。每栋楼宇的分项数据的质量问题可能五花八门，是否能在“**特性中找出共性**”成为搭建通用的分项数据清洗流程的关键。

本章立足于对百栋商业综合体的分项数据进行**缺失性分析**，提出一套相对**通用**的针对冷站的分项计量数据的**数据清洗流程**，清洗后的数据能够反应暖通空调系统的物理规律，可以直接用于深诊断。

4.3.2 分项计量数据的缺失特性分析

4.3.2.1 数据描述

通过对某集团百栋大型商业综合体 2020 年下半年的建筑运维数据作质量分析对发现当前分项计量数据的通病有一定的帮助。图 4.3 说明本研究中的某集团百栋大型商业综合体数量分布图，除海南、贵州、西藏、青海、甘肃及河北（内陆地区）的商业建筑未收录外，其余地区的商业综合体均有所涉及。图 4.4 说明建筑面积在不同地区的分布情况，大部分集中在江苏、四川、辽宁、山东和广东地区。收集的分项计量数据中的变量见表 4.1。

表 4.1 某集团大型商业综合体的分项计量数据中的变量汇总表

层级	变量名		
系统	冷冻水泵功率	冷却水泵功率	干管冷冻水供水温度
	干管冷冻水回水温度	干管冷却水供水温度	干管冷却水回水温度
	冷机群功率	冷却塔总功率	
冷机	冷冻水供水温度	冷冻水回水温度	冷却水供水温度
	冷却水回水温度	冷冻水流量	冷却水流量
	冷机启停状态	电流百分比	冷机功率
冷冻水泵	冷冻水泵功率	冷冻水泵频率	
冷却水泵	冷却水泵功率	冷却水泵频率	
冷却塔	风机功率	风机频率	
其他	室外干球温度	室外相对湿度	室内温度

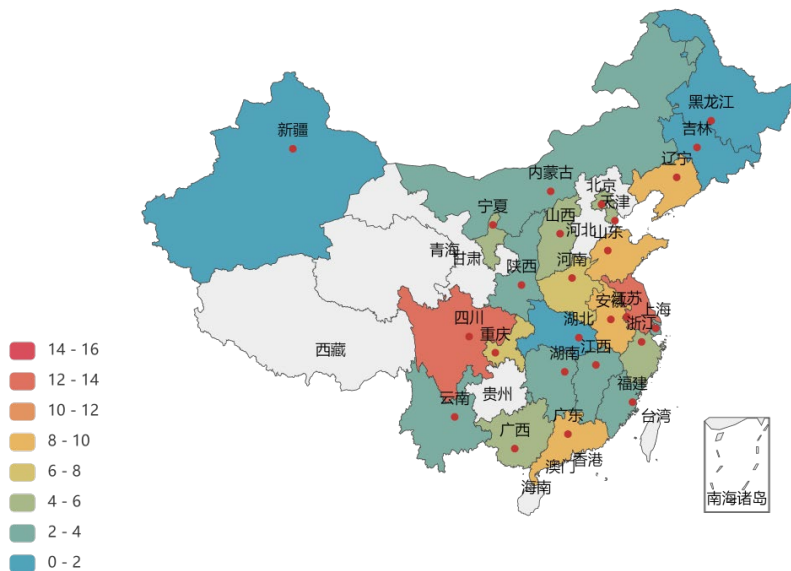


图 4.3 某集团大型商业综合体数量分布图

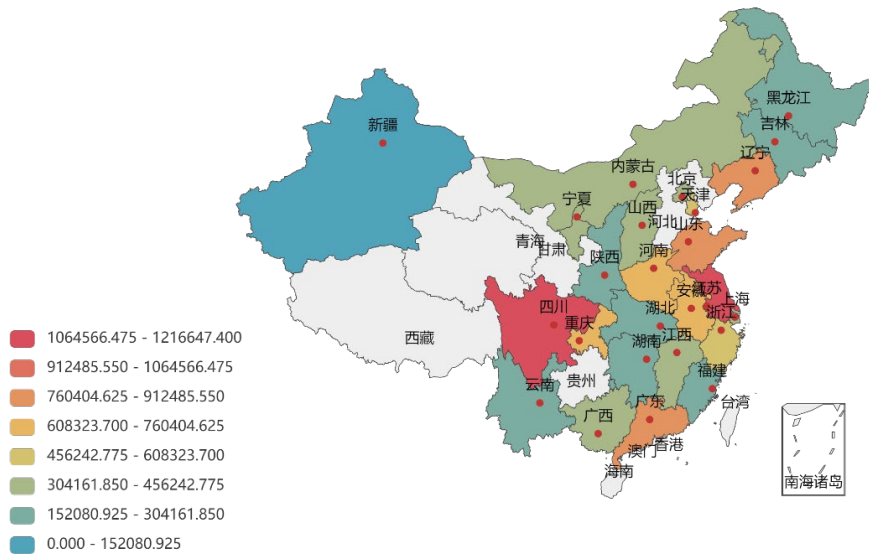


图 4.4 某集团大型商业综合体建筑面积分布图

4.3.2.2 缺失性分析-单栋

图 4.5-4.8 为某栋综合体中所有采集分项数据的缺失情况分布图，从中不难看出，冷却塔风机频率出现大规模缺失，并且有存在一定的关联性；单个冷冻水泵及冷却水泵的功耗在全年不同时间段均存在缺失；单个冷机功耗的缺失也存在一定的关联性，在 2020 年下半年初，除冷机、冷冻水泵、冷却水泵功耗以及系统层级的变量外，其余大部分变量均存在集体缺失的情况。

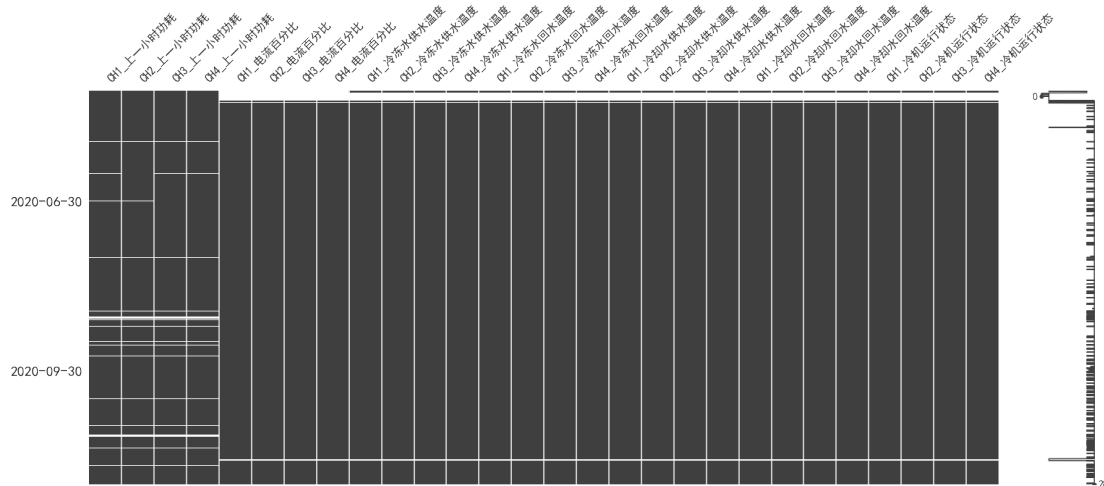


图 4.5 与冷机相关变量的缺失分布图，从左至右依次为每台冷机的功率、冷冻水供回水温度、冷却水供回水温度及冷机启停状态

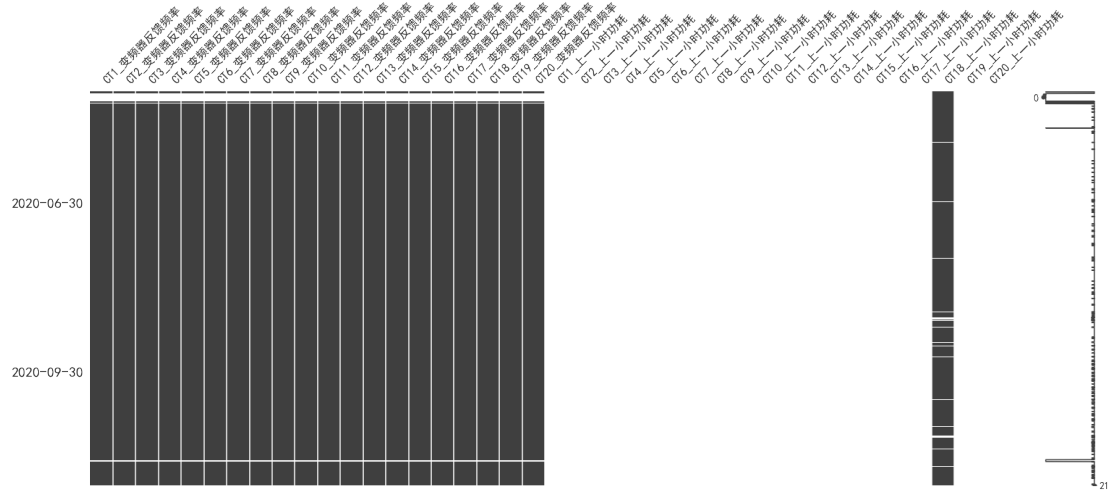


图 4.6 与冷却塔相关变量的缺失分布图，从左至右依次为每个风扇的实时频率

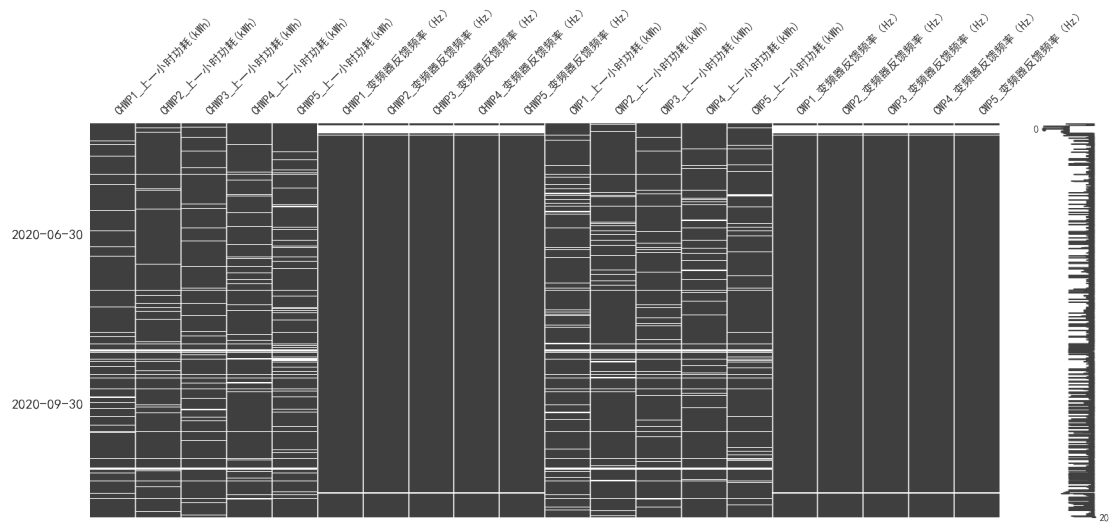


图 4.7 与冷却水泵和冷冻水泵相关变量的缺失分布图，从左至右依次为冷冻水泵功率、实时频率、冷却水泵功耗及实时频率

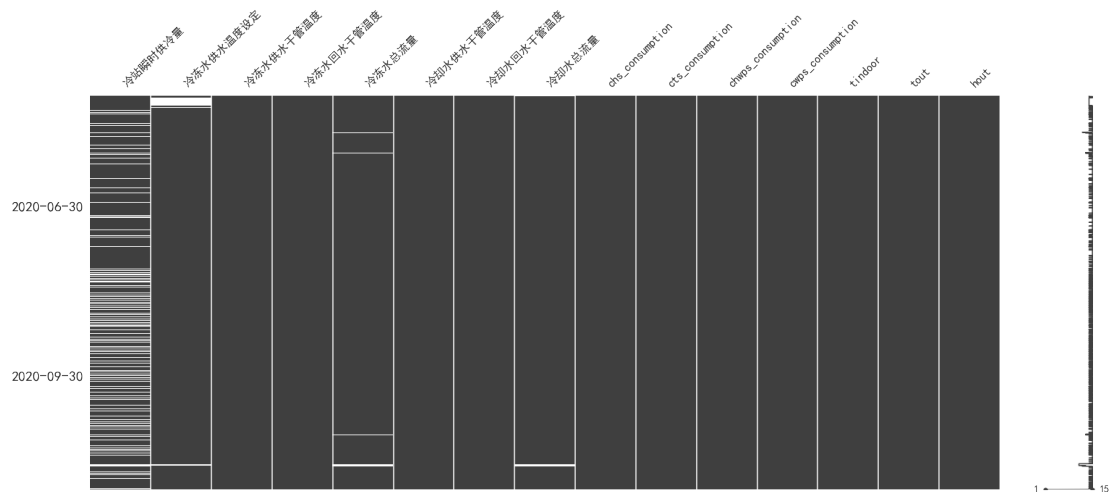


图 4.8 与系统和气象相关变量的缺失分布图，从左至右依次为间接测得冷负荷、干管冷冻水供回水温度、干管冷冻水水量、干管冷却水供回水温度、干管冷却水水流量、冷机群功耗，冷却塔风机总功耗、冷冻水泵群功耗、冷却水泵群功耗、室内温度、室外干球温度及室外相对湿度

上述是对单栋建筑的分项数据进行缺失性分析，不难发现，**设备群的总能耗、气象参数数据质量上乘**，有作为数据清洗框架的标定数据。

4.3.2.3 缺失性分析-整体

下面是对所有楼宇的分项计量数据作缺失性统计。图 4.9 表示不同变量在全部数据上的占比，与冷却塔直接相关的数据类型虽然较少，但是数据量在整体数据中占比大，主要因为冷却塔数量较多，并且每个冷却塔中的风机数量并不唯一，但是这样的点位布置存在一定的优化空间：因为仅仅只有冷却塔风机转速，并不能够有效同型号不同冷却塔在使用过程的劣化过程。与之相比，与冷机直接相关的变量比较丰富。从另外一个角度可以看出，工程上对冷机的关注度要高于对冷却塔的关注度。

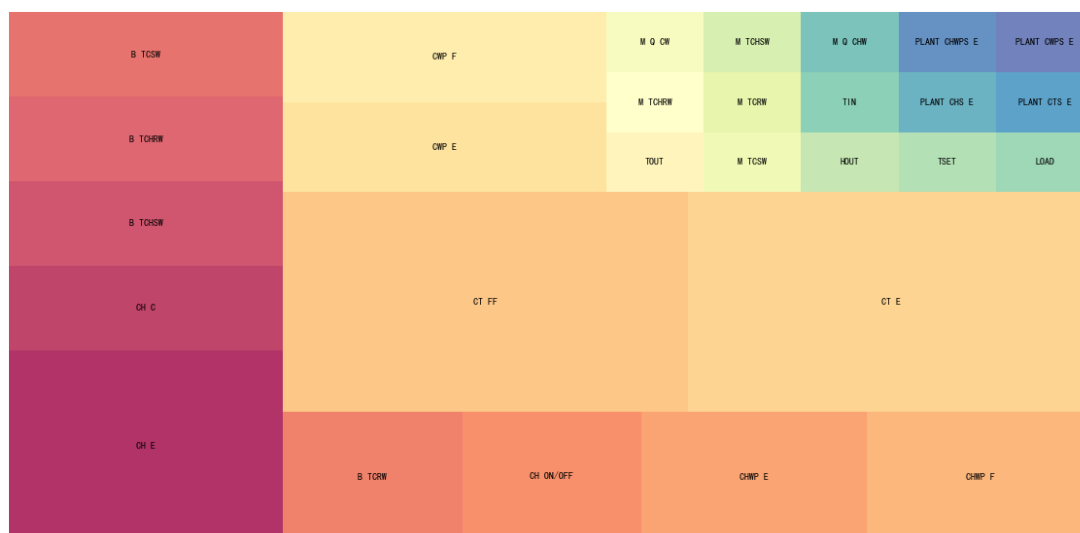


图 4.9 反映不同变量数量占比的矩形树状结构图

图 4.10 表示将所有分项计量数据进行分类归纳后，作缺失率统计，这里的分类是将每栋建筑中同一类型数据进行归纳，比如所有支管的冷冻水回水温度归纳为一类，所有冷却塔风扇的频率归纳为一类，依次类推。图 4.10(a)取值是所有商业综合体中每中变量缺失最严重的情况。各个设备群的总能耗的缺失值即使在糟糕的情况下也有较好的表现，这也印证了前文的结论，在数据缺失性分析中，**设备群的总能耗的数据质量优于其他分项数据**，且数据质量较为稳定，可以作为数据清洗中的标定数据。

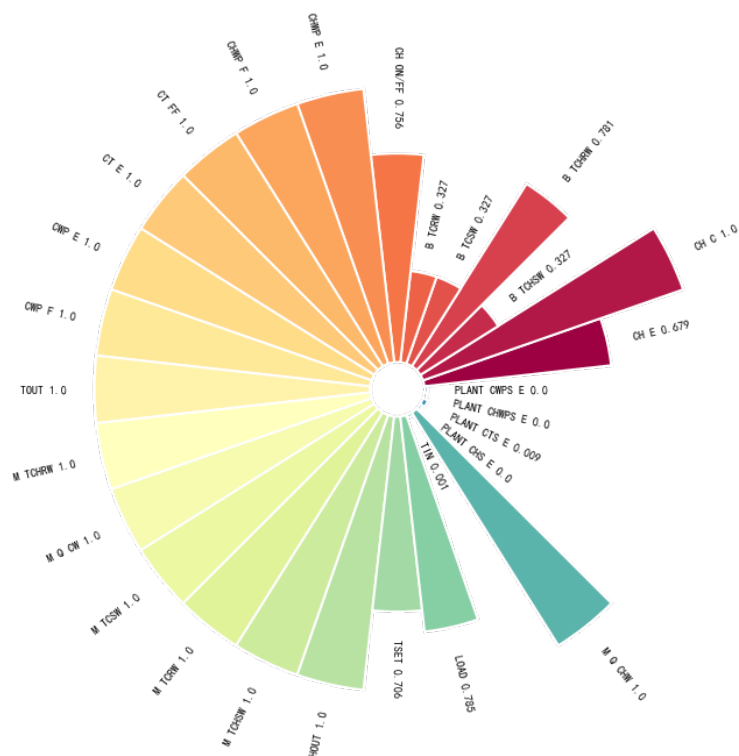


图 4.10 (a) 不同变量在数据缺失性分析中最差的情况

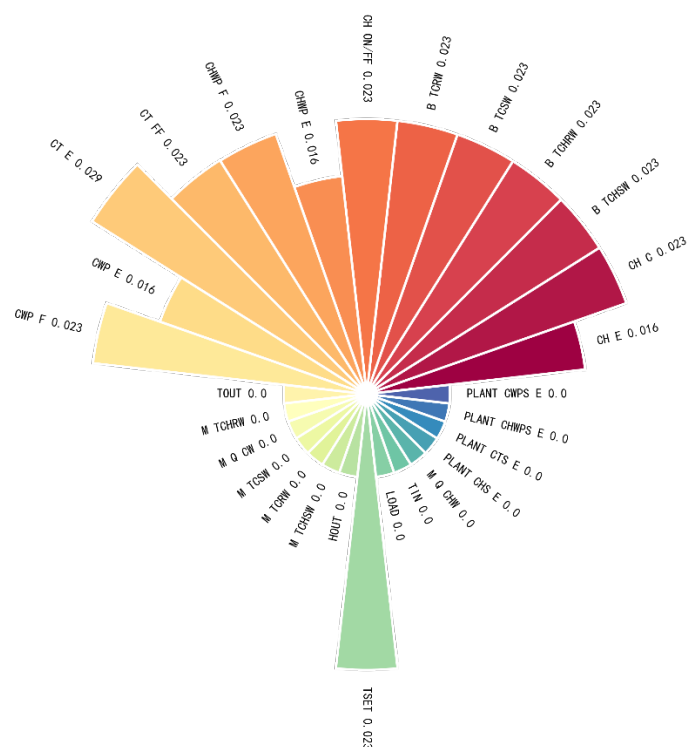


图 4.10 (b) 不同变量在数据缺失性分析中最好的情况

4.3.3 基于交叉验证的分项计量数据的数据清洗框架

根据上文分项计量数据的数据质量分析结果，本节提出了针对建筑冷站的分项计量数据的数据清洗框架（见图 4.11），这个框架是采用物理定律对数据进行交叉验证。以各个设备群的分项能耗数据为交叉验证的起点，随后分别验证与单个设备相关的数据，最后验证系统层级的数据。该数据清洗的流程为深诊断提供可靠的数据依据。图 4.11 中**虚线箭头**表示很可能因为数据的完全缺失而导致部分流程走不通。

下面以冷冻水泵总能耗——>单台冷冻水泵能耗——>单台冷冻水泵频率——>干管冷冻水流量——>干管水温度、冷负荷对本数据清洗流程作详细说明：

1) 冷冻水泵总能耗：除去零值、空值，根据专家经验判断冷冻水泵总能耗的直方图是否合理。

2) 单台冷冻水泵能耗：根据基尔霍夫定律和处理后的冷冻水泵总能耗进行数据清洗，即单台冷冻水泵能耗之和应等于冷冻水泵总能耗，见公式 4.22。

$$E_{all} = \sum E_i, i = 1, 2, \dots, n \quad 4.22$$

其中 E_{all} 为冷冻水泵总能耗， E_i 为单台冷冻水泵能耗， n 为冷冻水泵的总台数。

3) 单台冷冻水泵频率：根据相似准则进行数据清洗，修正数据和标签之间的不对应问题，见公式 4.23。

$$\frac{E_i}{E_{irated}} = \left(\frac{f_i}{f_{irated}} \right)^3, i = 1, 2, \dots, n \quad 4.23$$

其中 E_i 为单台冷冻水泵能耗， E_{irated} 为单台冷冻水泵额定能耗， f_i 为单台冷冻水泵频率， f_{irated} 为单台冷冻水泵额定频率， n 为冷冻水泵的总台数。这里需要说明的一点是：为保证水泵的安全运行，本文将变频水泵工作的最低频率设定在 30Hz，这样可以避免水泵低速情况是可能不满足相似准则的情况发生。

4) 干管冷冻水流量：根据相似准则和处理后的单台冷冻水泵频率进行数据清洗，见公式 4.24，后用专家经验比较计算的干管冷冻水流量和监测的干管冷冻水流量并判断是否合理。

$$Q_{all} = \sum \left(\frac{f_i}{f_{irated}} \right) Q_{irated}, i = 1, 2, \dots, n \quad 4.24$$

其中 Q_{all} 为计算的干管冷冻水流量， f_i 为单台冷冻水泵频率， f_{irated} 为单台冷冻水泵额定频率， Q_{irated} 为单台冷冻水泵额定流量， n 为冷冻水泵的总台数。

5) 干管水温度、冷负荷：根据能量守恒定律和修正的数据分析干管水温度、冷负荷是否合理，对偏差过大的数据剔除，见公式 4.25。

$$\rho c_p Q_{ch_{all}} (T_{chwr} - T_{chws}) + E_{chs} = \rho c_p Q_{c_{all}} (T_{cws} - T_{chwr}) \quad 4.25$$

其中 Q_{chall} 为修正的干管冷冻水流量， Q_{call} 为修正的干管冷却水流量， T_{chwr} 为干管冷冻水回水温度， T_{chws} 为干管冷冻水供水温度， T_{cws} 为干管冷却水供水温度， T_{cwr} 为干管冷却水回水温度， c_p 为水定压比热容， ρ 为水的密度。

水泵和冷却塔风机相关的数据清洗是类似，而冷机相关的变量略有区别：在图 4.11 中，与水泵频率和风机频率处于同一层级的是冷机启停状态，这里对冷机启停状态的清洗做额外说明：冷机的启停状态理论上应该和对应冷机的能耗对应，所以直接用单台冷机的能耗数据映射到对应冷机的启停状态中。

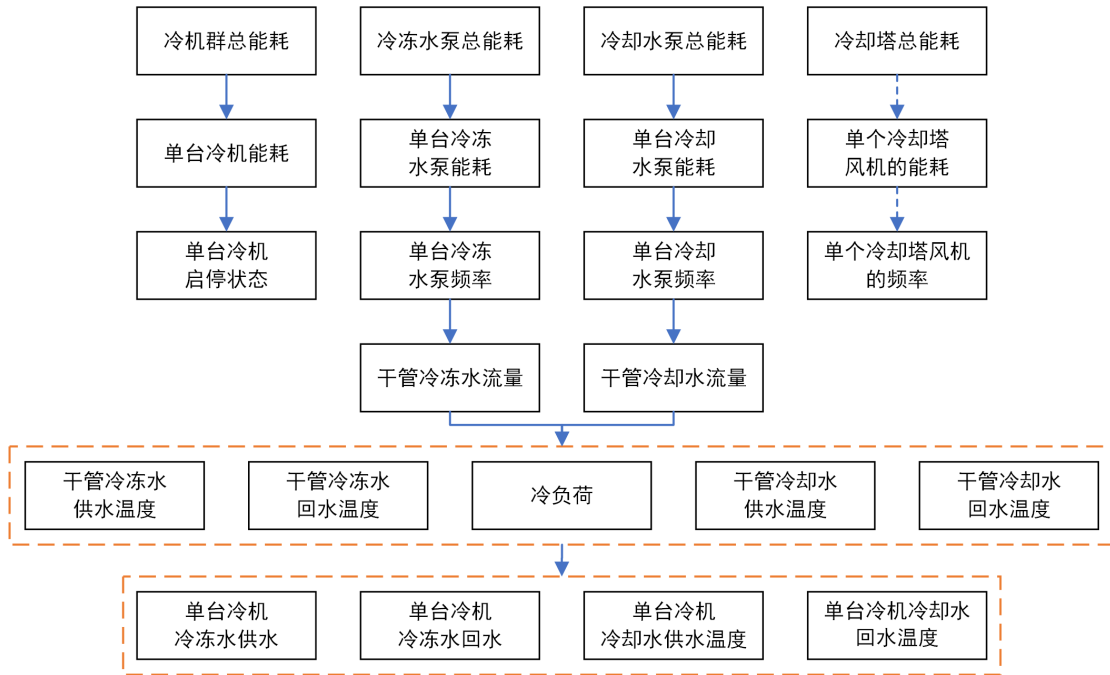


图 4.11 针对冷站能耗数据的数据清洗流程图

4.3.4 常见能效问题诊断

经过上述数据清洗流程处理的数据可以诊断常见的能效诊断问题^[56]，见表 4.2，该表同时说明了不同的问题诊断的数据要求，我们可以根据现有的数据去判断表 4.2 中的部分问题；值得注意的是，对于涉及压力数据的能效诊断问题，在部分楼宇的 BAS 或者能耗监测平台中并没有相关数据，相关问题得不到诊断。

表 4.2 冷站常见能效问题和以及对应所需的诊断数据^[56]

设备	能效问题	诊断数据
冷机	冷却水进水温度过高	$T_{ce} > T_{ce_LV}$
	冷冻水进水温度过高	$T_{ee} > T_{ee_LV}$
	冷冻水出水温度过高	$T_{ei} > T_{ei_LV}$
冷冻水泵	大流量小温差	$\Delta T_{chw} < \Delta T_{chw_rated} - 2$

	冷冻环路有阻塞迹象	$\frac{\Delta P_{b_chw}}{Q_{b_chw}^2} > \frac{\Delta P_{b_chw_rated}}{Q_{chw_rated}^2}$
冷却水泵	冷却环路有阻塞迹象	$\frac{\Delta P_{b_cw}}{Q_{b_cw}^2} > \frac{\Delta P_{b_cw_rated}}{Q_{cw_rated}^2}$
冷却塔	变频风机不变频或者多档调节风机不调速	$\Delta E_{ctf} \rightarrow 0 \cup \Delta f_{ctf} \rightarrow 0$
	冷却塔换热性能变差	$Q_{cw} \approx Q_{cw_rated} \cap \Delta T_{cw} > \Delta T_{cw_rated} + 1$

4.4 本章小结

本文介绍了基于实测分项计量数据的冷站运维节能潜力分析的细节，根据不同的数据情况进行“浅诊断”和“深诊断”：“浅诊断”仅需要建筑静态数据和冷机群总能耗，“深诊断”则是尽可能充分利用分项数据，在经过本文提出的相对通用的数据清洗流程后，直接用于诊断常见的冷站能效问题。

第5章 基于实际建筑案例的节能潜力分析验证

5.1 案例一

5.1.1 案例描述与数据介绍

5.1.1.1 案例建筑描述

本案例为上海某政府办公楼，该建筑位于上海，地上27层，办公建筑面积为49710.47m²，建筑外立面见图5.1。其工作时间为法定工作日8:30-18:00，建筑常驻人数为1671人，主要功能包括办公、会议室、厨房和网络机房等，建筑围护结构采用玻璃幕墙结构，屋顶采用保温层，玻璃幕墙贴膜，窗户玻璃采用中空玻璃，设有遮阳。

描述建筑外部形态和影响建筑负荷的关键参数真实值见表5.1，其中建筑的体型系数、窗墙比可以根据实际建筑图纸获取，建筑设计和实际施工之间的差距忽略不计；其余较为难以获取的参数，如人员密度（m²/Person）、照明功率（Watt/m²）以及设备功率（Watt/m²），根据搭建的EnergyPlus模型和实际冷站总能耗数据校核得出，实际冷站总能耗数据是从第三方能耗监测平台获取的2020年8月、9月的法定工作日工作时间的冷站总能耗数据。

根据ASHRAE Guideline 14中模型校核的指标为参考指标，从图5.2中可知，将表5.1中的关键变量值输入建筑能耗模型中，模拟能耗结果和实际能耗结果之间的偏差在标准规定的可接受范围内，其中 $CV-RMSE_{Hourly} = 0.141 < 0.15$ 且 $MBE_{Hourly} = 4.475e-5 < 0.5$ ；我们对能耗模型的模拟结果和实际数据之间的误差进行了一定分析：导致CV-RMSE较高的是因为九月份常驻人员出现一定的增长的时段。表5.1中的关键变量作为本文中需要推测关键变量的验证标准。

表5.1 建筑较难获取的关键参数真实值

数据类型	人员密度 m ² /Person	照明功率 Watt/m ²	设备功率 Watt/m ²
真实值	10	24	18.45
数据类型	建筑体形系数 Dimensionless	窗墙比 Dimensionless	
真实值	0.15	0.545	

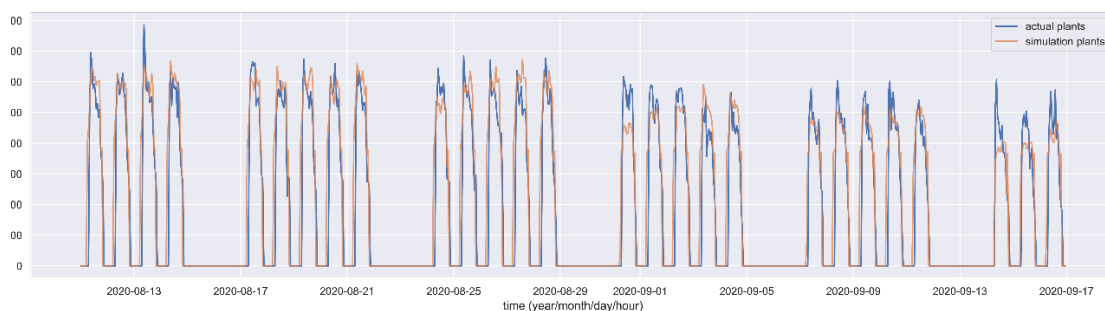


图 5.2 模拟能耗值和实测能耗值之间的误差

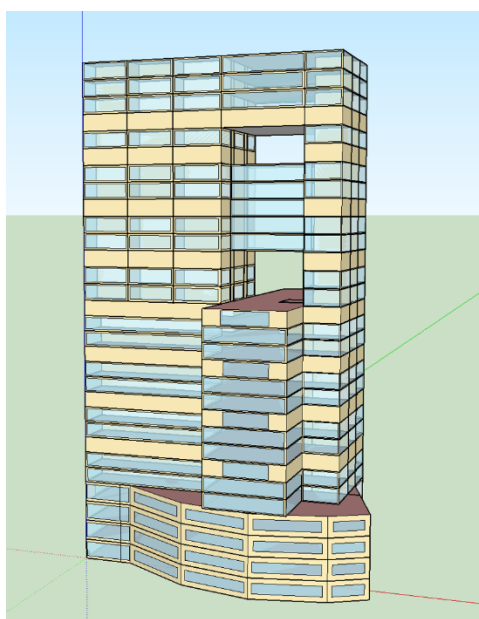


图 5.1 上海某政府办公建筑外立面视图

5.1.1.2 案例冷站信息描述

该案例中夏季供冷形式为冷水机组加风机盘管系统，其中冷站的核心设备为 3 台 650RT 离心冷水机组，其他配套设备见详细设备清单，空调末端采用风机盘管和新风机组合的形式，其详细 HVAC 系统设备清单见表 5.2，无铭牌参数，无建筑空调系统系统图。

表 5.2 上海市某政府办公楼 HVAC 系统（供冷）设备清单

名称	台数	参数 1	参数 2	参数 3
离心式冷水机组	3	单台制冷量:650RT	单台功率:429KW	
冷却水泵 1	2	额定功率:37KW	额定流量:400m ³ /s	额定扬程:23m
冷却水泵 2	2	额定功率:55KW	额定流量:550m ³ /s	额定扬程:28m
冷冻水泵 1	1	额定功率:37KW	额定流量:400m ³ /s	额定扬程:23m
冷冻水泵 2	1	额定功率:45KW	额定流量:320m ³ /s	额定扬程:34m
冷冻水泵 3	2	额定功率:55KW	额定流量:550m ³ /s	额定扬程:28m
冷却塔	3	电机功率:22.5KW	额定流量:600m ³ /s	
风机盘管	852	不详		

此外，该案例建筑刚结束建筑节能审计不久，合理推断该栋大楼的设备均经过一定的检查与养护，笔者根据实际能耗和该案例建筑能耗模型得出第四章中两个与冷机群能耗有关的关键变量（冷机综合 COP、冷机劣化系数）真实值见表 5.3。

表 5.3 该案例与冷机群能耗相关的关键变量

No.	变量名	单位	真实值	变量类型
1	冷机综合 COP	Dimensionless	5.41	float
2	冷机劣化系数	Dimensionless	0.98	float

5.1.1.3 案例实测数据描述

本案例的实测数据包括建筑供冷季冷站逐时总能耗、冷机群逐时总能耗及温度。本验证案例中并未采用实测温度数据：1) 因为温度数据标签不明确，并未说明温度数据来源的具体点位；2) 该数据为空值。

5.1.2 案例验证内容

本案例主要验证的内容有：1) 第二章的基于关键变量推测的建筑内热负荷分析；2) 第三章的基于轻量化黑箱模型的设备选型方案的节能潜力分析；3) 第四章的基于实测分项计量数据的冷站运维的节能潜力分析——浅诊断。根据上文的该案例实测数据描述可知，该案例无法进行第四章的深诊断部分，因为缺乏“微观”数据。

5.1.3 基于关键变量推断的建筑内热负荷分析方法验证

5.1.3.1 建筑-能耗算例生成

根据第二章中的技术路线，其抽样数据和先验取值范围见表 5.4，变量 9 建筑体形系数是根据变量 1-4 计算得出，因为 Grasshopper 的参数化设计仅需要参数 1-8，但是参数 9 的取值范围是满足大部分商业办公建筑的取值范围的，工业园区业态的群楼商业办公建筑不在考虑范围之内。

表 5.4 建筑形态数据采样指标及取值范围

No.	采样指标	单位	取值范围	变量类型
1	建筑边界长 Boundary_Length	m	[15,60]	浮点型 float
2	建筑边界宽 Boundary_Width	m	[10,60]	浮点型 float
3	建筑楼层数 Num_of_Floors	floor	[3,7.5]	浮点型 float
4	建筑层高 Height_each_Floor	m	[5,40]	浮点型 float
5	东窗墙比 EWWR	dimensionless	[0.1,0.9]	浮点型 float
6	南窗墙比 SWWR	dimensionless	[0.1,0.9]	浮点型 float
7	西窗墙比 WWWR	dimensionless	[0.1,0.9]	浮点型 float
8	北窗墙比 NWWR	dimensionless	[0.1,0.9]	浮点型 float

9	建筑体形系数 Coefficient Shape	dimensionless	[0.1,0.25]	浮点型 float
---	-----------------------------	---------------	------------	-----------

用雷达图（见图 5.3）可以更加直观其清晰表示对与建筑形态有关的参数采样的均匀性，其中一个色块表示一个采样点（一个建筑算例），其中取值超出 [0,1]均进行标准化处理，除 SC 参数，因为该参数的取值范围见[0.1, 0,25]。

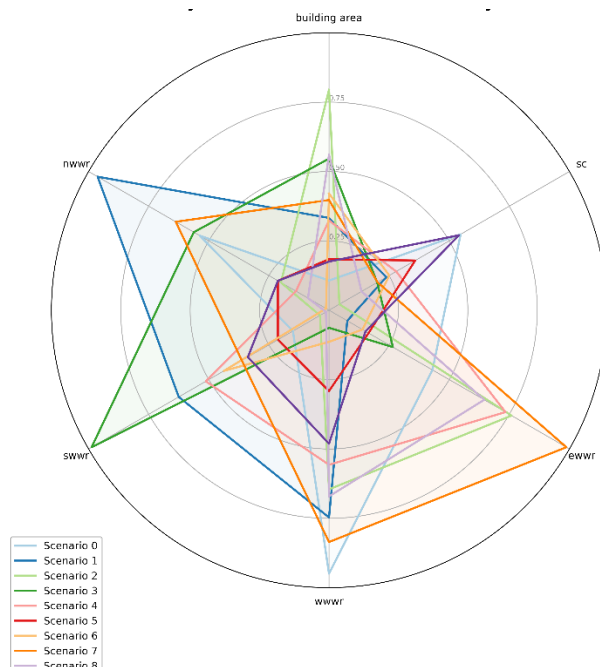


图 5.3 与建筑形态相关参数抽样结果雷达图（局部）

建筑形态无关的影响能耗的关键变量中，固定一部分关键变量，包括建筑 HVAC 系统的冷热源形式，水系统形式、空调末端形式，主要考虑到：1）大部分商业办公建筑的冷站系统和末端空调系统形式相同，2）缩短生成建筑-能耗数据的时间。仅对部分带有明显有不确定性的未知关键变量参数采用 LHS 抽样，抽样变量和取值范围见表 5.5。

表 5.5 待推测关键参数及取值范围

No.	关键变量	单位	取值范围	变量类型
1	人员密度 Occupant_Density	m ² /Person	[2,12]	浮点型 float
2	照明功率 Lighting_Density	Watt/m ²	[10,26]	浮点型 float
3	设备功率 Equipment_Density	Watt/m ²	[10,22]	浮点型 float

将表 5.5 中的抽样点用 eppy 批量写入 IDF 文件中，采用并行计算的方式形成建筑-能耗数据库。

5.1.3.2 黑箱模型选择及超参数调优

这里采用 XGBoost 作为黑箱模型，作为后续联动全局搜索算法推测照明功率、设备功率以及人员密度的基础；根据模拟数据训练 XGBoost 模型，XGBoost 模型的输入参数有室外气象参数（室外干球温度、相对湿度），日类型

(星期几、小时) 以及关键变量。训练结果见图 5.4。训练数据和测试数据是采用 scikit-learn 中 8:2 比例随机切分, 蓝线代表训练集中预测值和真实值之间的残差, 橙线代表测试集中预测值和真实值之间的残差, 横坐标表示训练迭代过程, 每训练 10 次显示一次误差, 即横坐标中的 1 表示第 10 次迭代完成后的残差, 总计迭代 100 次。

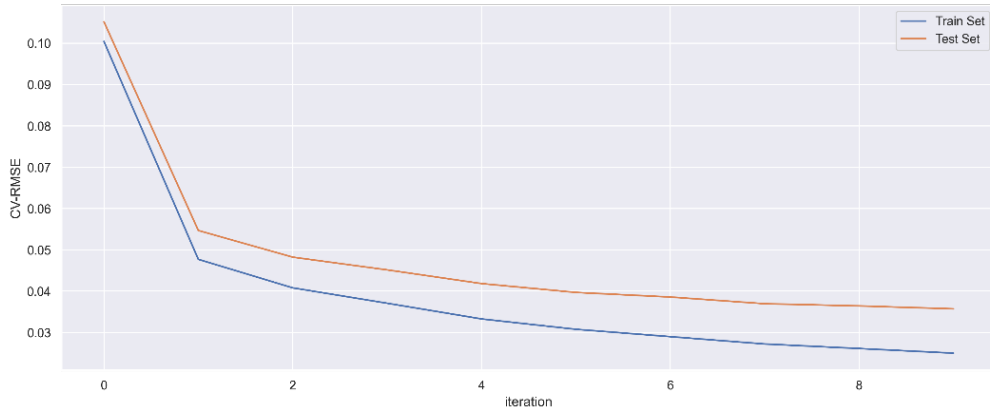


图 6.4 用于推断关键变量的 XGBoost 模型

从训练结果中可以发现, 在模型训练初期, 模型评价指标 CV-RMSE 就已经满足 ASHRAE Guideline 14 的要求, 因为 XGBoost 的超参数调优过。又因为训练时间在笔者可接受范围内, 所以并未采取 lightGBM 进行比较。XGBoost 模型中的需要调优的超参数见表 5.6。

表 5.6 XGBoost 超参数取值空间

No.	超参数	取值范围	变量类型
1	Learning_rate	[0.01,0.5]	浮点型 float
2	Max_depth	[1,10]	整型 integer
3	Gamma	[0,1]	浮点型 float
4	Min_child_weight	[1,10]	整型 integer
5	Subsample	[0.01,1.0]	浮点型 float
6	Colsample_bytree	[0.01,1.0]	浮点型 float

笔者将迭代数调整为 50, 训练得到的最优超参解见表 5.7。训练过程见图 6.5-6.8。

表 5.7 用于预测非设备层级关键变量的 XGBoost 模型的最优超参数

参数名	Learning_rate	Max_depth	Gamma
参数值	0.220	7	0.573
参数名	Min_child_weight	Subsample	Colsample_bytree
参数值	5	0.853	0.940

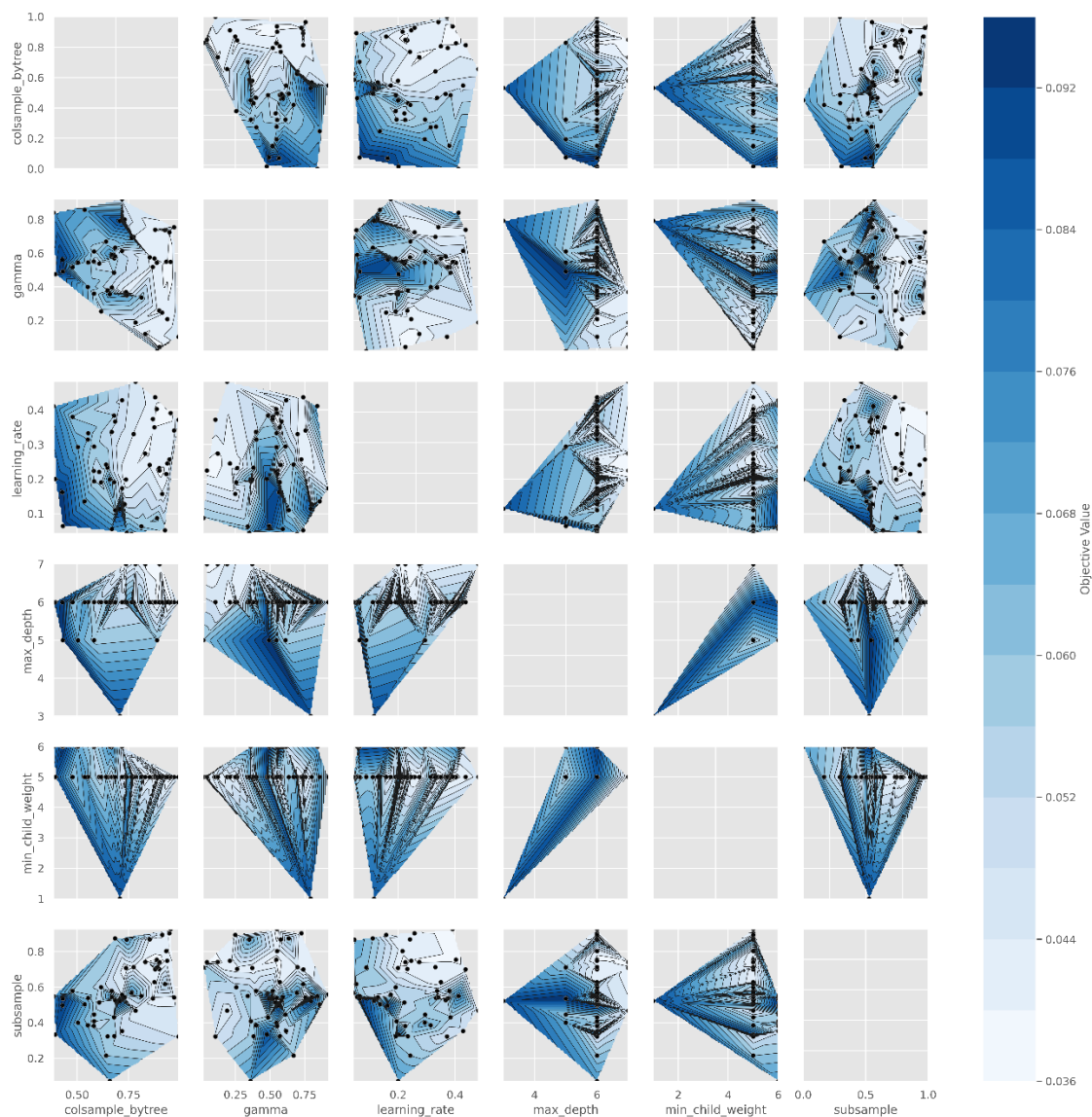


图 5.5 超参调优过程中超参数两两之间的关系

图 5.5 中展示在超参数调优过程，高维超参数解空间中不同超参数之间的关系，点表示调优过程中的抽样点，颜色表示目标值（此处目标值为 CV-RMSE），因为训练方向是最小值方向，所以超参空间中颜色越浅说明该区域越接近最优解，因为该超参空间是六维，为方便可视化，将高维解空间截开展示两两超参数的关系，从抽样点的分布可以看出，浅色区域的抽样点较为密集，从中可以看出该方法对于超参的优化并非简单的随机搜索而是存在优化方向，该结论也可以从图 5.6 中得出。

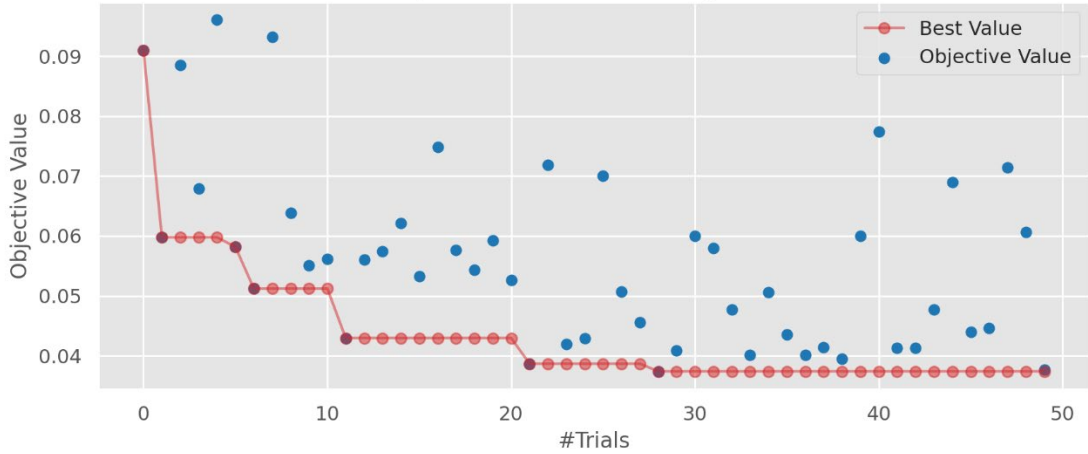


图 6.6 超参优化历史折线图

在图 5.6 中，红色的点表示本轮抽样中的最优超参数解，从每一轮迭代的最优超参数解的黑箱模型的训练指标 CV-RMSE 变化趋势可以看出，超参数空间抽样分布是不断更新变化的，以确保下一轮抽样中，解集会更趋向于最优解。图 5.7 可以更加清楚观察到：超参数在哪种情况可以使 CV-RMSE 达到最小。图 5.7 中的每个子图纵坐标代表的是 CV-RMSE 值，CV-RMSE 值越小说明模型精度越高，横坐标表示的是不同的超参数对应的取值，右边的 Color Bar 中颜色深浅表示训练迭代的次数，颜色越深表示迭代次数越高，根据散点图中的点疏密程度可以判断，随着迭代次数不断升高，目标确实朝着越小的方向前进。

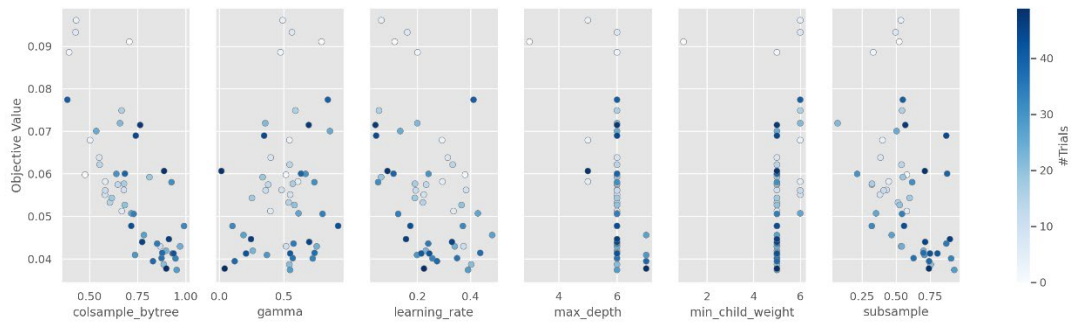


图 5.7 不同超参对应的目标值散点图

这里我们并不关心，最后是否得到全局最优解，我们当然可以把迭代数目设置到更大，得到所谓“全局最优解”，但是考虑到时间效率，将迭代数调整到 50，经过超参数调优后的 XGBoost 模型在测试集上的 CV-RMSE<0.4 已经远低于 ASHRAE Guideline 14 的指标。

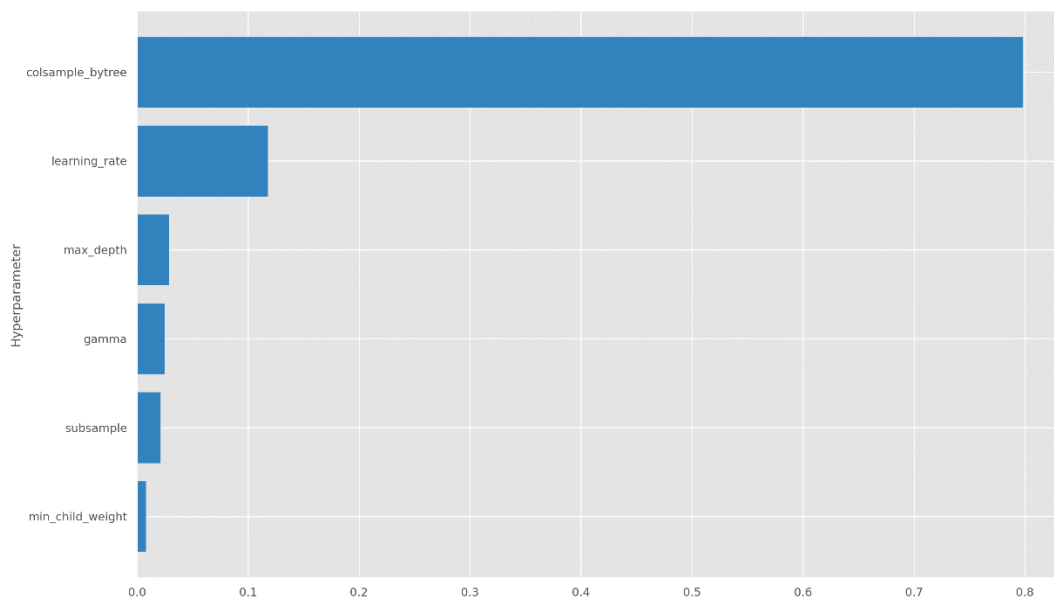


图 5.8 XGBoost 能耗模型超参重要性排序

虽然 XGBoost 对超参数非常敏感，但是不同超参数对模型的敏感性是不同，图 5.8 展示数据驱动的能耗模型中，不同超参数的重要性排序，与常规认知存在偏差的是：学习率（learning_rate）和最大深度（max_depth）在手工调参中常常被认为是主要调节的参数，而本文实验的结果显示 colsample_bytree 可能才是手工调参中需要有限调整的超参数。

5.1.3.3 基于全局搜索算法推测关键变量的准确性分析及验证

以含最优超参数的 XGBoost 为基础，结合不同算法和不同的指标推测关键变量，并从准确性和时间效率分析。为保证不同方法之间的可比性，统一将粒子群个数和种群数调整为 100，迭代次数均调整为 50 次。

最终不同组合下推测的关键变量结果见表 5.8，从表中可以看出采用传统指标 CV-RMSE 时间效率高且准确。但是四种方法中照明密度采用不同的方法推测的值存在一定波动。若将待推测的关键变量对 XGBoost 模型的重要性进行排序（见图 5.9），不难发现，在同一栋建筑中，人员密度的重要性远远大于其余两个关键参数，从这里也可以解释在每一种组合中，对人员密度的推测精度比剩余两个关键变量推测得更加精准。

表 5.8 不同全局优化方法推测的结果的比较

No.	类型	人员密度 m ² /Person	照明密度 Watt/m ²	设备密度 Watt/m ²	时间%H:%M:%S
	真实值	10	24	18.45	
1	GA + CV-RMSE	10.966	22.824	20.598	00:00:29.5
2	GA + DTW	9.252	19.089	20.053	00:01:19
3	PSO + CV-RMSE	10.755	22.600	20.309	00:00:17.7
4	PSO + DTW	9.768	19.058	20.077	00:01:53.8

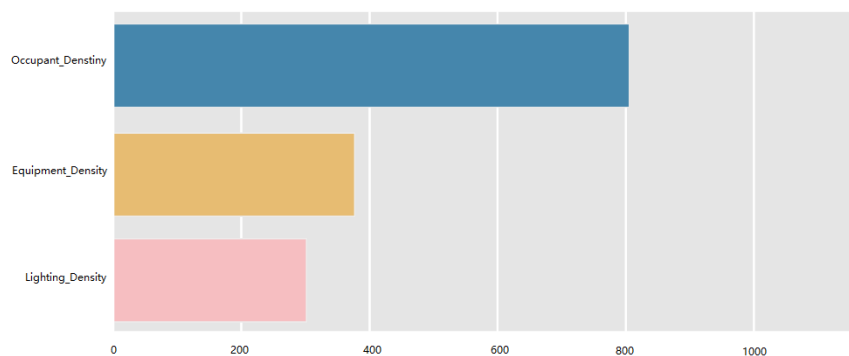


图 5.9 三个待推测关键变量模型中重要性排序

为确保推测结果的准确性，除了提高黑箱模型自身精度之外，还需保证推测结果收敛，因为全局搜索算法最后会有一些结果落入局部最优解，但是根据多轮实验发现每次的结果都大致相同，即本文研究的过程中陷入局部最优的困境发生的概率较小，GA 的推测结果见图 5.10，另外 PSO 不论配合 CV-RMSE 还是 DTW 指标，均在 10 次迭代内达到收敛。

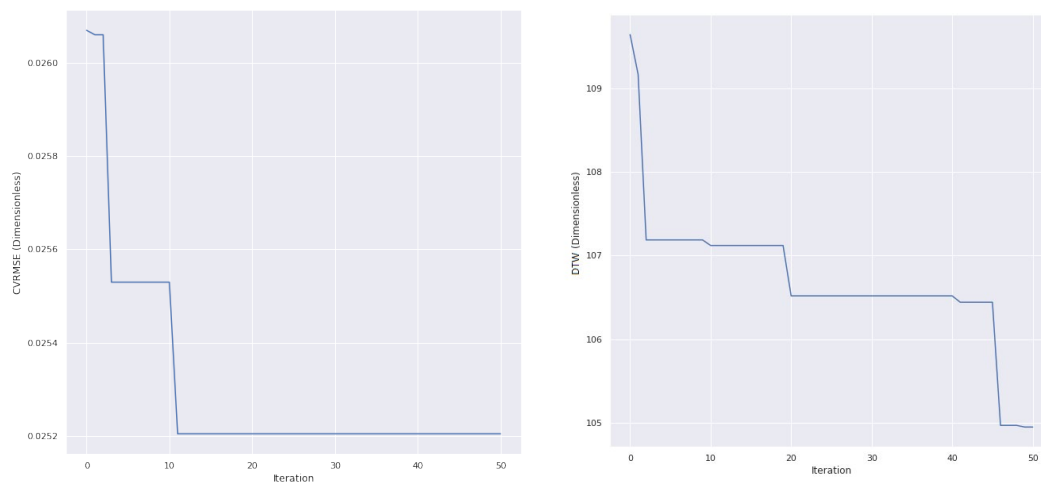


图 5.10 GA+CV-RMSE 迭代收敛情况（左）GA+DTW 迭代收敛情况（右）

根据上述推断的结果，在内热构成上，该建筑案例的照明系统中存在较大的节能潜力，《GB55015-2021 建筑节能与可再生能源利用通用规范》附录 C（见表 5.9）中办公类建筑的照明功率的要求 $8\text{Watt}/\text{m}^2$ ，该建筑的照明系统存在较大的节能潜力，标准中办公建筑的电器设备功率要求为 $15\text{Watt}/\text{m}^2$ ，该建筑中常用设备的散热较大，但这不意味在此方面，建筑存在较大的节能潜力，因为这大概率是建筑自身使用功能性质所导致的。将该政府办公楼的人员密度和标准中办公建筑的人员平均占有建筑面积进行对比，该建筑在此方面是贴近节能标准下限值的。

表 6.9 GB55015-2021 附录 C 不同建筑类型照明功率、电器设备功率密度和人均占有的建筑
面积要求

No.	建筑类型	照明功率密度 Watt/m ²	电器设备功率密度 Watt/m ²	人均占有的建筑面积 m ² /Person
1	办公建筑	8.0	15	10
2	旅馆建筑	6.0	15	25
3	商场建筑	9.0	13	8
4	医院建筑-门诊楼	8.0	20	8
5	医院建筑-住院部	6.0	15	25
6	学校建筑-教学楼	8.0	5	6
7	居住建筑	5.0	3.8	25
8	工业建筑	6.0	15	10

从外部得热角度看，该办公建筑的窗墙比过高，但这是初期设计保留的决定，但该建筑的遮阳又一定程度上削减了外部得热量。虽然本建筑在冬冷夏热地区，虽然 GB55015-2021 中仅对严寒和寒冷地区新建公共建筑体形系数作出限制，但也可以类比参考，该建筑本身的体形系数约为 0.15，尚在规范要求范围内。

5.1.4 基于轻量化黑箱模型的设备选型方案节能潜力分析方法验证

5.1.4.1 建筑-负荷算例生成

本节能耗-负荷数据库的生成直接依托用于建筑-能耗算例的 idf 模型并同样采用 LHS 方法进行抽样，非建筑形态相关的变量取值见表 5.10。

表 5.10 非建筑外形的影响负荷的关键变量抽样取值表

No.	变量名	单位	取值范围	变量类型
1	冬季室内设计温度 winter_desined_indoor_temperature	Celsius	[15,21]	浮点型 float
2	夏季室内设计温度 summer_desined_indoor_temperature	Celsius	[21,26]	浮点型 float
3	建筑使用情景 building_function	Dimensionless	[1,3]	整型 Integer
4	遮阳情况 overshadowing_or_not	Dimensionless	[0,1]	整型 Integer
5	人员密度 occupant_density	m ² /person	[2,10]	浮点型 float
6	设备功率 equipment_density	Watt/m ²	[15,25]	浮点型 float
7	照明功率 lighting_density	Watt/m ²	[10,20]	浮点型 float

在表 5.10 中建筑使用情景包括三种状态，分别是低频、标准以及高频^[31]。因为案例建筑在上海，并且采用 2020 年数据校准模型，所以选用 2020 年上海气象参数，见图 5.11、5.12。

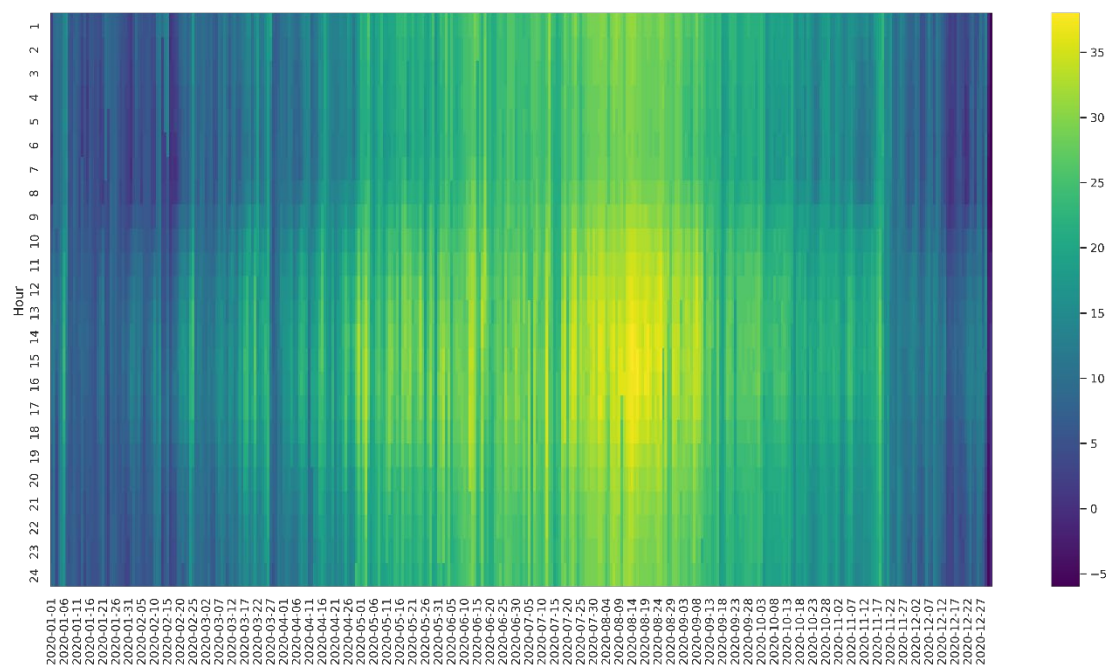


图 5.11 上海 2020 年全年室外干球温度热力图

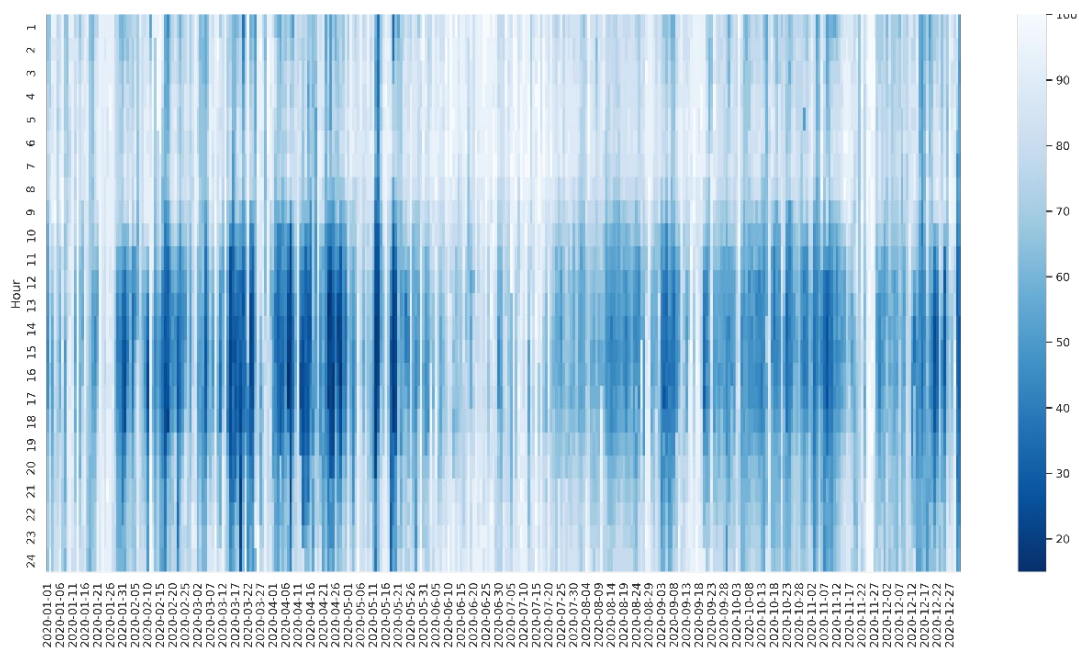


图 5.12 上海 2020 年全年室外相对湿度热力图

图 5.13 是表 5.10 中的关键参数的抽样的局部结果的展示，其中取值范围超过 $[0,1]$ 的参数，在可视化过程均采用均一化处理。从图 5.13 中可以看出数据采样的均匀性。笔者将对建筑设计的变量采样和非建筑设计方面的关键变量采样进行结合，可形成图 5.14，同样对取值范围非 $[0,1]$ 的变量进行归一化。

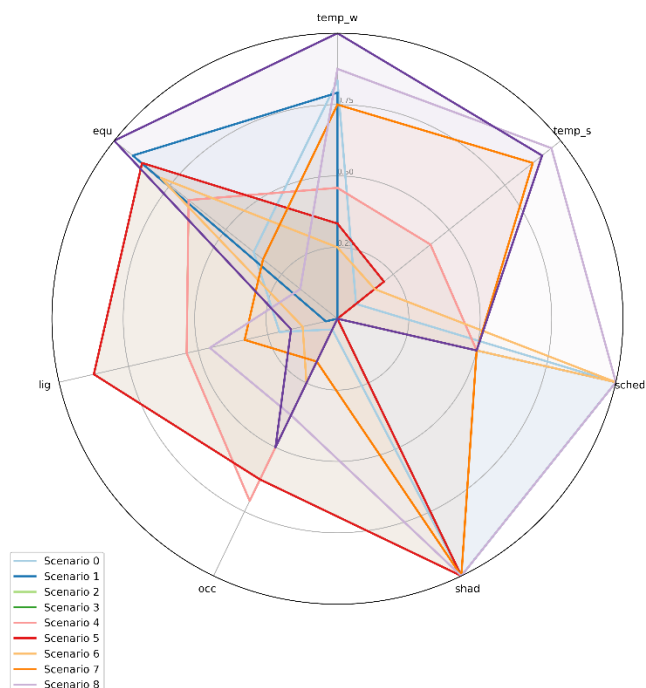


图 5.13 非建筑设计的影响建筑热负荷的关键变量的采样结果雷达图

图 5.14 中每一个色块表示某一个建筑-负荷算例中的一个关键参数，其中包括建筑面积、体形系数、窗墙比、冬夏室内设计温度、建筑使用频率、遮阳情况、人员密度、设备功率以及照明功率。每一行色条代表一个建筑-负荷算例，图 5.14 中仅对局部负荷算例的展示。

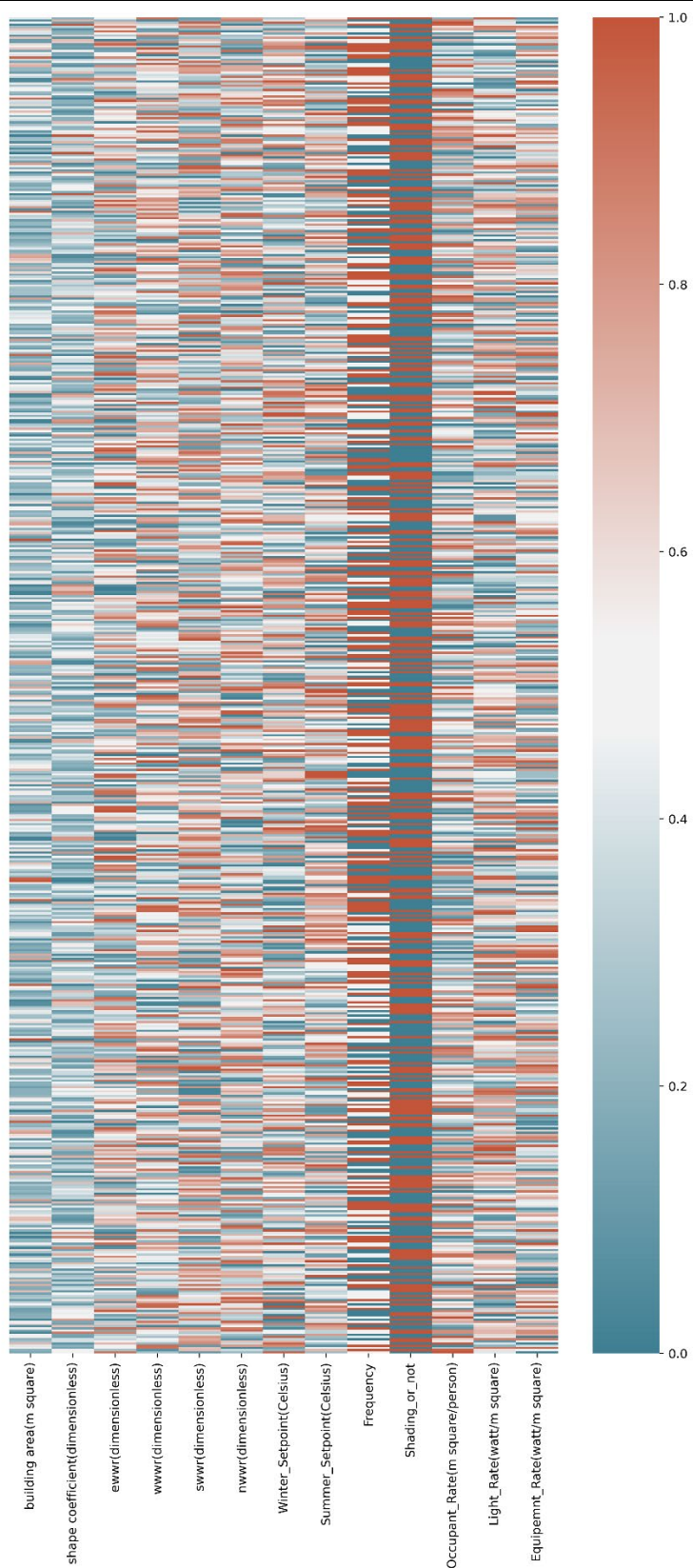


图 5.14 建筑负荷算例热力图（部分）

5.1.4.2 黑箱模型选择及超参数优化

当训练数据规模量不大的情况下，基于 XGBoost 模型改进的 LightGBM 的轻量化特点在应用上的优势并不能得到充分的体现，在建筑内热负荷的分析中，仅采用 XGBoost 辅助全局搜索优化算法推测关键变量而不采用 LightGBM 原因在于：虽然 XGBoost 采用全排列的方式搜索分裂点，在时间效率上略逊于 LightGBM，但是整个模型的训练的时间也在笔者的可接受范围内。根据第四章介绍的超参数调优方法，50 轮迭代的时间效率尚且在可接受度的范围内，所以笔者没有对 XGBoost 和 LightGBM 时间效率做对比，因为两者均属于轻量化模型的范畴，两者的时间效率并不会拉开特别大的差距。

但在本章中，XGBoost 模型和 LightGBM 模型同时被采用。采用 XGBoost 模型目的在于和前文建筑-能耗的 XGBoost 模型进行对比，在数据同源，任务不同情况下，超参数对模型敏感度的变化情况；而采用 LightGBM 模型目的在于和 XGBoost 模型比较，在数据同源，任务相同的情况下，模型精度。

本部分采用控制变量，尽量使模型精度的差异仅来自模型本身而非其他因素。同样先采用 sciki-learn 划分测试集和训练集，默认测试数据占数据总量的 25%，随后超参调优选择最优超参解。虽然两个模型本身的超参数因为模型不同不具备可比性，但可以保持其迭代轮数一致，参与调优的超参数数量保持一致。LightGBM 中需要调优的超参数见表 5.11。

表 5.11 LightGBM 超参数取值空间

LightGBM			
No.	超参数	取值范围	变量类型
1	Learning_rate	[0.01,0.5]	浮点型 float
2	Num_leaves	[2,256]	整型 integer
3	Feature_fraction	[0.4,1.0]	浮点型 float
4	Bagging_fraction	[0.4,1.0]	浮点型 float
5	Bagging_freq	[1,7]	整型 integer
6	Min_child_samples	[5,100]	整型 integer

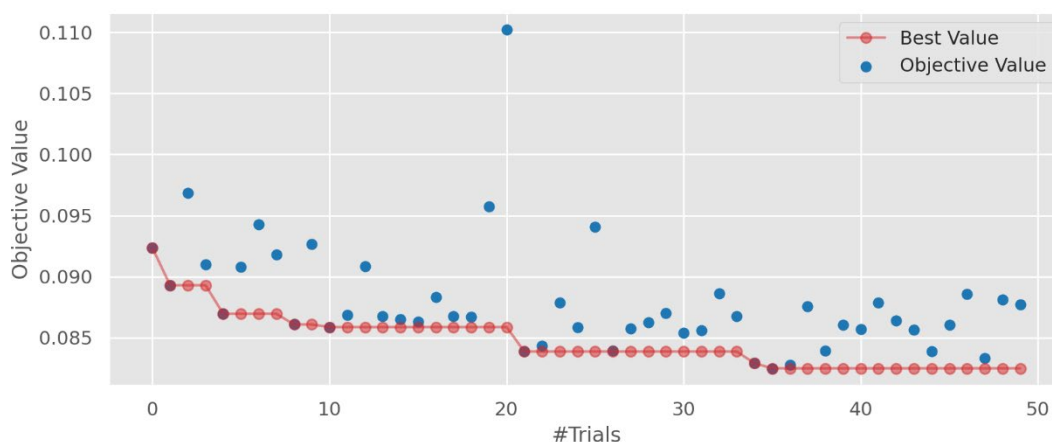


图 5.15 LightGBM 超参调优历史折线图

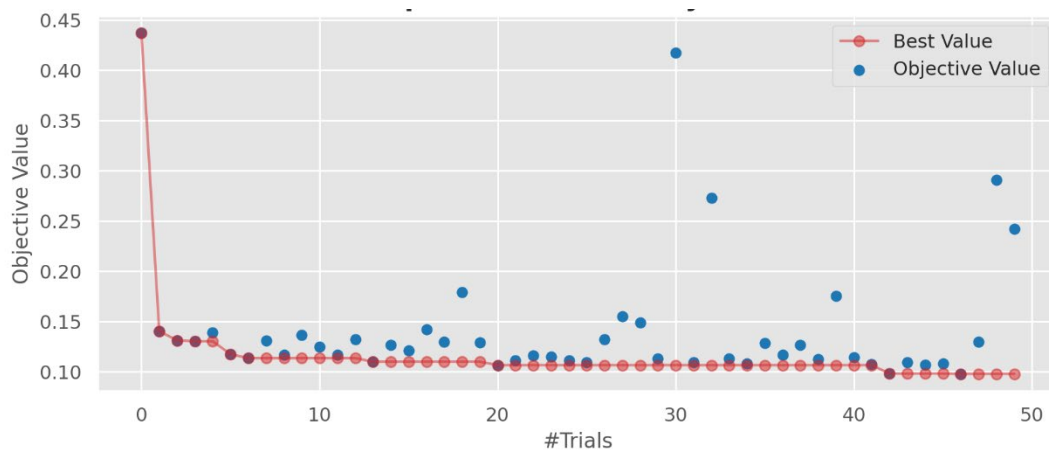


图 5.16 XGBoost 参调优历史折线图

通过图 5.15 和图 5.16 可以看出，从模型准确度上，XGBoost 和 LightGBM 在最优参数的情况均能满足 ASHRAE Guideline 14 的标准，但 LightGBM 模型的精度更胜一筹。

从图 5.17 展示，在超参调优的过程中，LightGBM 的超参数两两间的关系，颜色越浅代表该超参数的值越接近最优解。

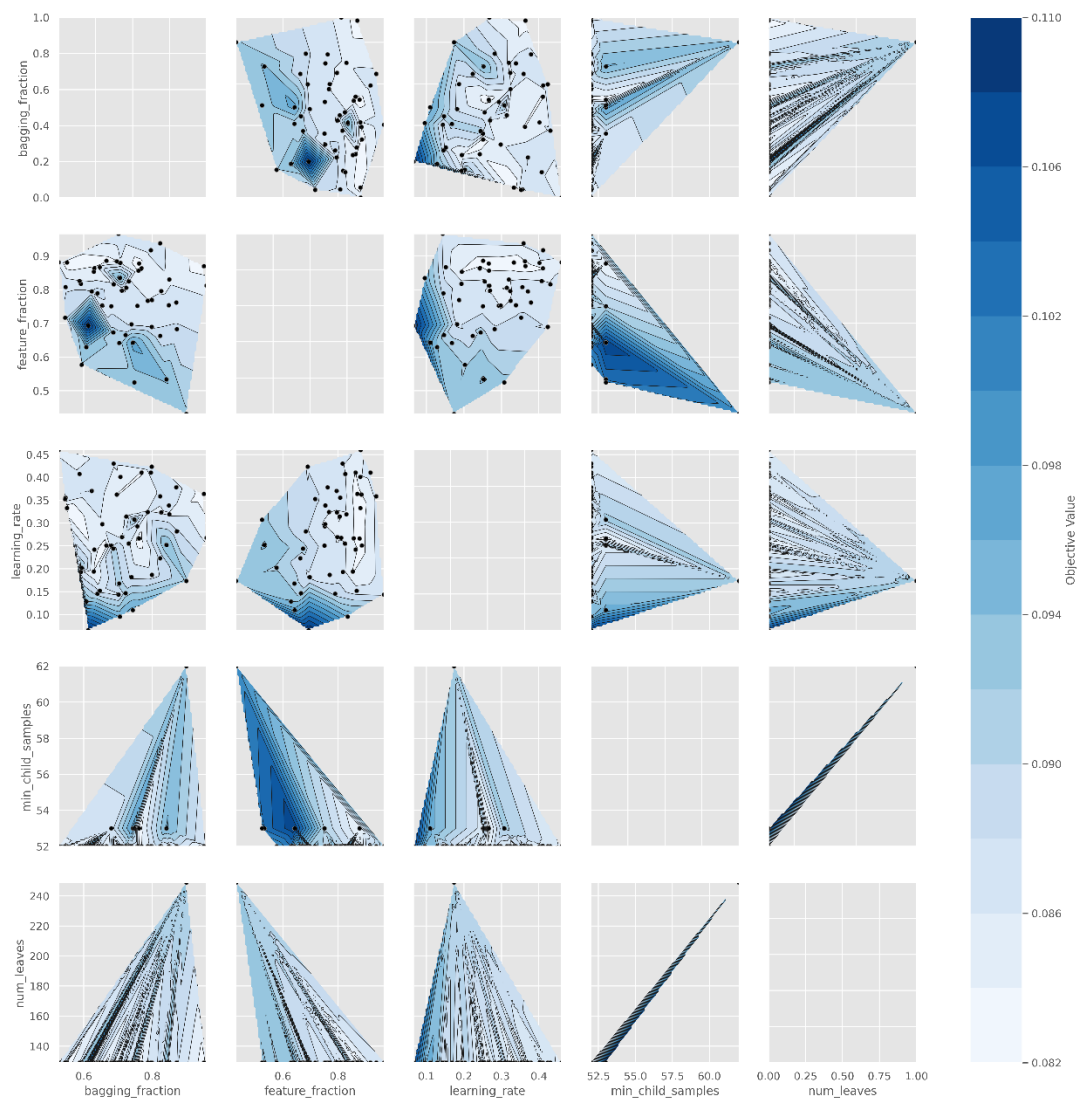


图 5.17 超参调优过程中超参数两两之间的关系 (LightGBM)

图 5.18 的展示内容与图 5.17 相似，区别在于黑箱模型不同，颜色越浅代表该超参数的取值越接近最优解，不论是图 5.17 还是图 5.18，根据抽样点的疏密程度可知，模型超参数的优化是向提高模型精度的方向进行的，而不是随机搜索。

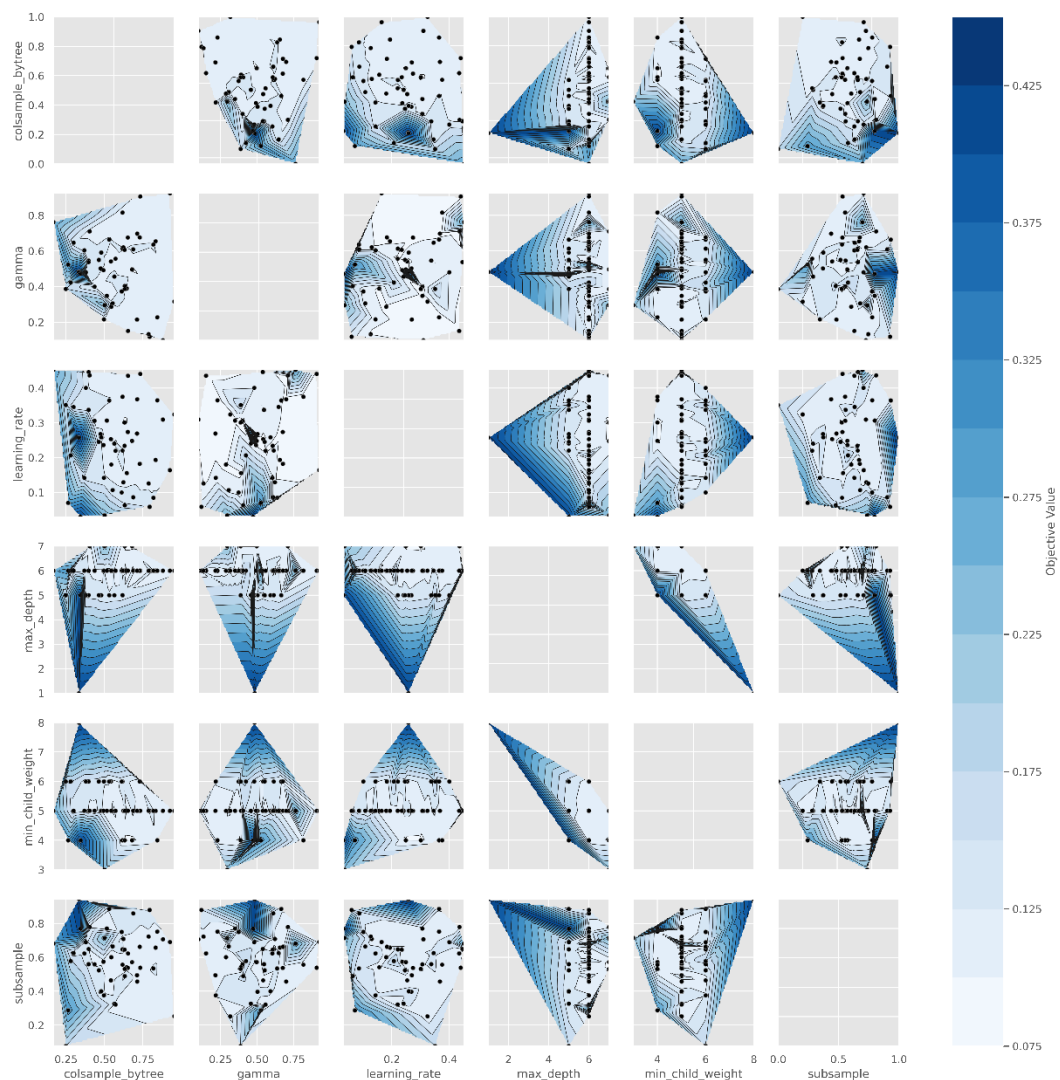


图 5.18 超参调优过程中超参数两两之间的关系 (XGBoost)

根据图 5.19 和图 5.20 的散点图可以更加清楚观察出超参数取值和模型优化目标 CV-RMSE 之间的关系，散点均集中在纵坐标值较低的底部区域，这样提高黑线模型的精度；散点的疏密程度（纵坐标值越低出抽样点越密集）再一次印证该超参数调优是有方向的。

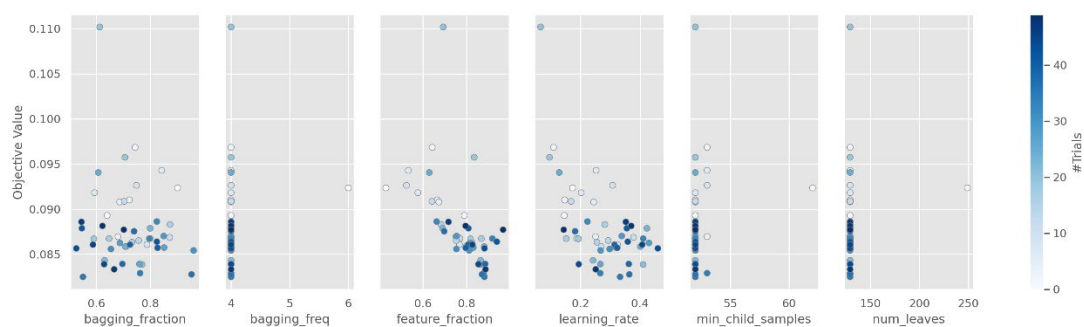


图 5.19 LightGBM 模型中不同超参数取值与目标函数 CV-RMSE 之间的关系

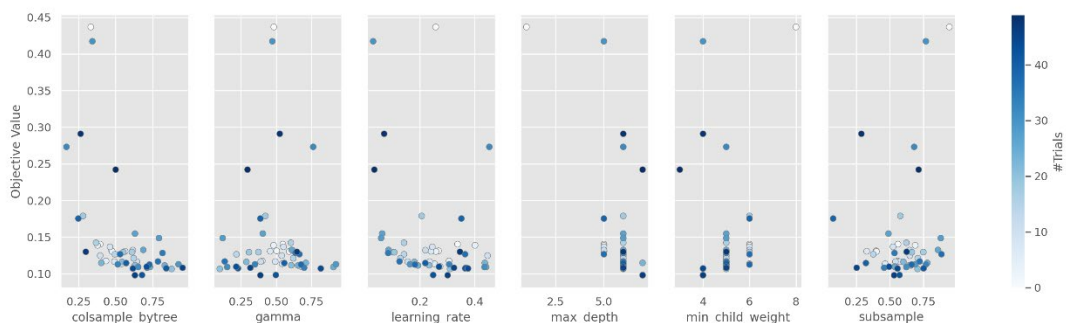


图 5.20 XGBoost 模型中不同超参数取值与目标函数 CV-RMSE 之间的关系

从图 5.21 和 5.22 中可以看出，在建筑负荷推测的任务中，模型对不同超参数敏感性并不相同，LightGBM 模型中 learning_rate 和 feature_fraction 是主要影响模型关键性的超参数，其他四个超参数对模型精确度的影响很小。在 XGBoost 模型中，colsample_bytree、learning_rate 和 subsample 分别是前三大影响模型精确度的超参数，这里得到的结论和图 5.8 得出的结论有略微的差别：图 5.8 分析的结论是，对 XGBoost 模型精度影响最大的是 colsample_bytree 和 learning_rate。但图 5.22 得出不同的结论：colsample_bytree, learning_rate, subsample 和 max_depth 是手动调参需要优先考虑的超参数。

即使采用同样的黑箱模型，当模型承担的预测任务并不相同时，超参数对模型的重要性排序也会与之发生变化，本章得出的结论同样对于手工调参有着重要的意义，往往机器调参得到的结果会与手工调参的固有认知会有一些偏差。

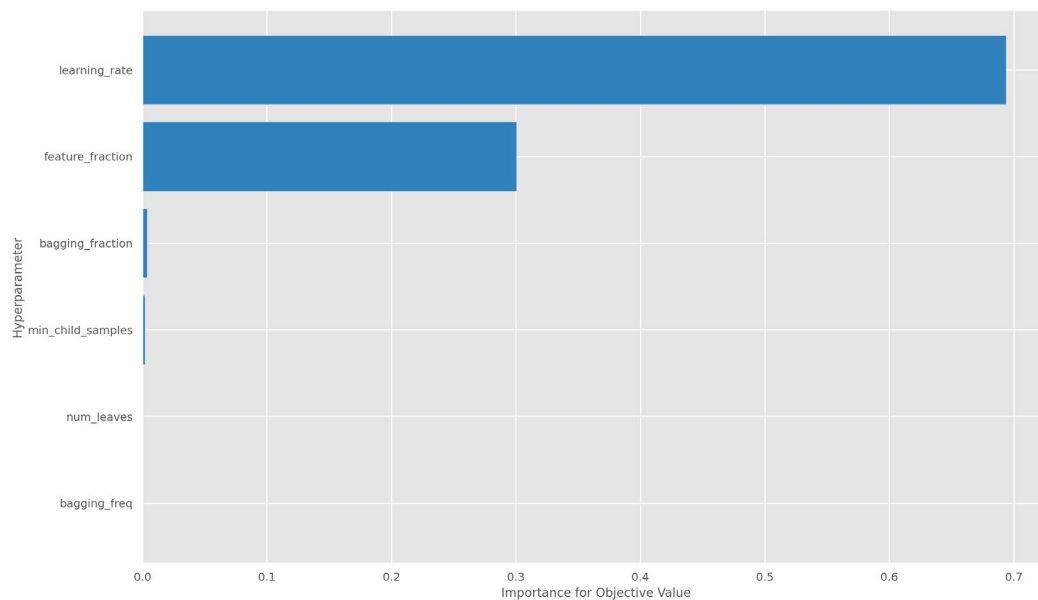


图 5.21 LightGBM 模型中超参数重要性排序

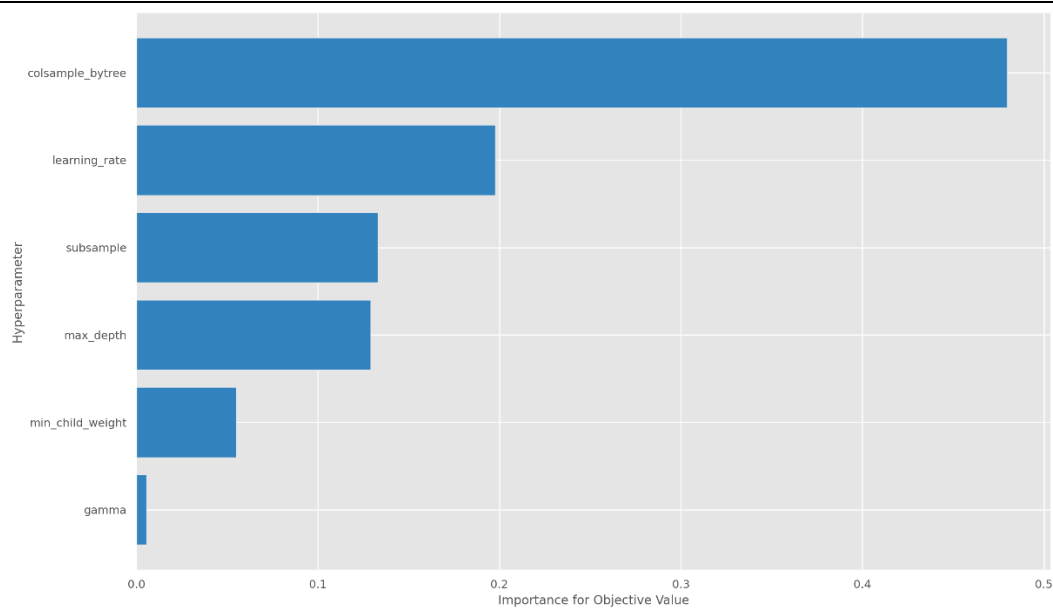


图 5.22 XGBoost 模型中超参数重要性排序

5.1.4.3 基于全局搜索分析设备选型方案

根据上述训练的模型推测逐时负荷，该栋建筑的冷热负荷见图 5.23 和图 5.24。

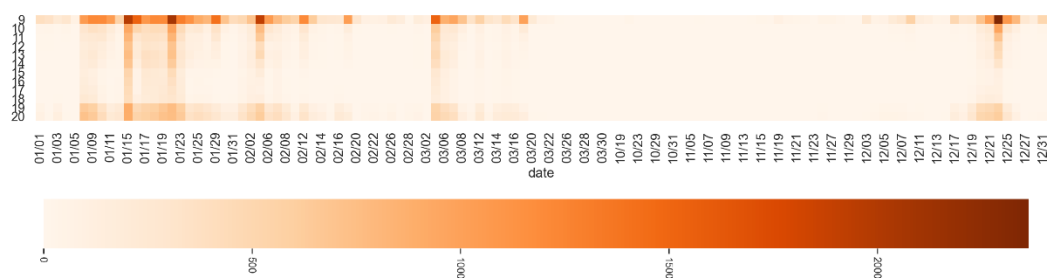


图 5.23 实例建筑热负荷热力图

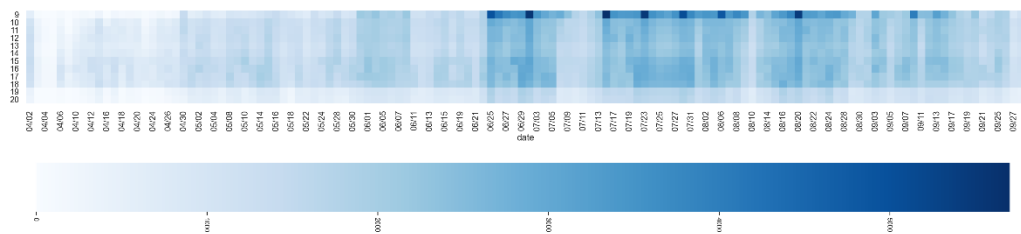


图 5.24 实例建筑冷负荷热力图

采用 PSO 优化该建筑案例的冷机选型方案，既有冷机选型方案是三台 650RT 的离心式冷水机组，造成现有设计冷机选型方案的原因可能为如下两点：

1) 该案例建筑的冷负荷分布相对均匀, 不同量级的逐时冷负荷出现的次数相对均匀;

2) 设计阶段的冷机配置并没有考虑到该栋建筑的冷负荷分布。

用平均部分负荷率 \overline{PLR} 分析造成当前冷机选型方案的是何种原因, 在理想冷机启停控制的情况下, 平均 \overline{PLR} 是 0.616。经过 PSO 优化冷机配置后, 平均 \overline{PLR} 提升至 0.777, 结果表明可能设计阶段的设备选型方案并未考虑建筑的实际负荷分布, 现有的冷机选型方案和建筑负荷的匹配程度有待进一步提高。



图 5.25 现有冷机配置在设计阶段的能耗情况

表 5.12 PSO 优化后冷机配置结果和现有的冷机配置情况

	Chiller 1	Chiller 2	Chiller 3
现有冷机配置情况	2321.15 KW	2321.15 KW	2321.15 KW
PSO 优化后冷机配置情况	930.963 KW	1697.923 KW	4047.364 KW

表 5.13 GB 50189-2015 中对冷水机组的性能要求 (节选)

类型	名义制冷量 CC(kW)	性能系数 COP(W/W)					
		严寒 A、B 区	严寒 C 区	温和 地区	寒冷 地区	夏热 冬冷 地区	夏热 冬暖 地区
水冷 离心 式	CC≤1163	4.70	4.70	4.74	4.84	4.93	5.02
	1163<CC≤2110	5.20	5.20	5.20	5.20	5.21	5.30
	CC>2110	5.30	5.30	5.30	5.39	5.49	5.49

对冷机选型方案的优化结果见表 5.12, 三台冷机的装机容量出现明显的分层, 结合 GB 55015-2021《建筑节能与可再生能源利用通用规范》中对离心式变频冷水机组的性能要求在现有的真实冷机库中按照标准下限进行选型 (表

5.13)。上海处于夏热冬冷地区，则按照下限搜寻的冷机配置应该是 930.963KW 的冷机对应的 COP 是 4.93，1697.923KW 的冷机对应的 COP 是 5.21，4047.364KW 的冷机对应的 COP 是 5.49，如果按照这样的配置是无法在真实冷机库里面配置一模一样的冷机（部分冷机模型能耗曲线在部分工况下不满足专家知识），只能去匹配相对合适的冷机模型去进行节能潜力分析，笔者分别选择了 932kW/5.09COP、1794kW/7.90COP、4396kW/6.63COP，这个选型确实过于激进但是由于现有冷机库内的匮乏，只能退而求其次。根据真实冷机性能模型在理想情况下（不考虑负荷在冷机间的分配），制冷季耗电量节省 13.38%（理想情况下，若实际更换冷机选型方案，在实际冷机启停控制不确定的情况下，节能效果可能远不及预期），PSO 优化前和 PSO 优化后的对比见图 5.26。



图 5.26 优化前和优化后冷机配置在耗能、PLR、COP 上对比

冷站中的设备除冷水机组以外，其他的设备均作为冷水机组的配套设备，所以机组改动必然会牵扯到其他设备的联动改造，包括冷冻水泵和冷却水泵以及冷却塔。主要是根据置换的冷水机组的额定水量去更换水泵和冷却塔。

5.1.5 基于分项数据的冷站运维节能潜力分析——浅诊断方法验证

5.1.5.1 分项数据的降噪处理

如图 5.27，蓝色的虚线表示没有噪声没有误差的冷机群总能耗数据，橙色虚线表示含有高斯噪声（SNR=5）的分项能耗数据，绿色点划线表示修正之后分项能耗数据值。当 SNR=5，信号传输质量较差的情况下，CV-RMSE 达到了 0.549，修正（采用 S-G 滤波器）之后降低到 0.3180。这是以传感器数据传输和监测质量很差的情况为实验，当传感器质量较好时，修正后的数据可能可以使 CV-RMSE 降至 0.15，但这是不具验证条件的，因为实际情况下，我们可能无法获取完全没有偏差的冷机群总能耗数据。

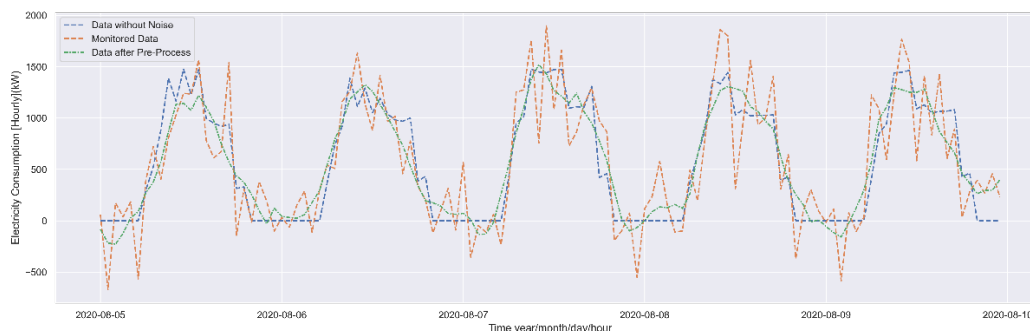


图 5.27 能耗数据监测值（人工加入噪声）、真实值、修正值

5.1.5.2 建筑-冷机能耗算例形成

本部分也是沿用上面的思路，为提高计算效率，本部分选择与案例建筑外形相仿的 idf 作为模板，固定已知参数，对未知的影响冷机能耗的待推测关键变量（表 5.14）参数采用 LHS 抽样。

表 5.14 待推测关键参数及取值范围

No.	关键变量	单位	取值范围	变量类型
1	内部劣化系统 Internal Deterioration Coefficient	Dimensionless	[0.75, 1.0]	浮点型 float
2	综合 COP General_COP	Dimensionless	[4.0, 7.5]	浮点型 float

5.1.5.3 黑箱模型选择和超参数优化

模型采用的是 XGBoost 模型，超参数调优结果见图 5.28-图 5.29。图 5.28 表达的意思和图 5.17、图 5.18 表达的含义相同，区别在于图 5.28 的模型输出为冷机群。

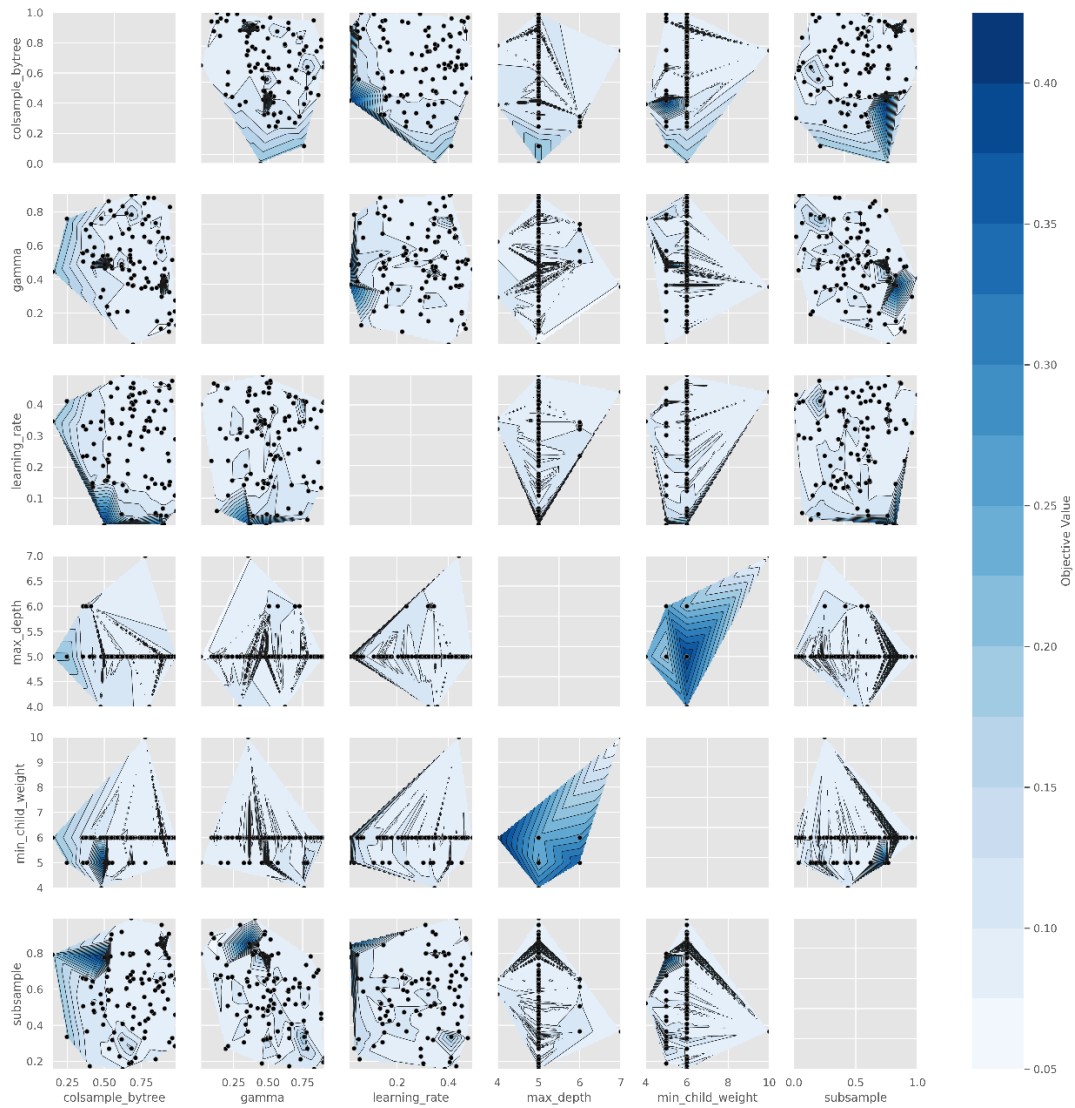


图 5.28 XGBoost 模型中超参数空间中 CV-RMSE 分布（模型目标：冷机群分项能耗）

图 5.29 表达的含义和图 5.19、图 5.20 表达的含义相同，再次验证第四章中基于树模型的超参数调优方法不同于 Random Search 和 GridSearch，该种方法是存在优化方向。

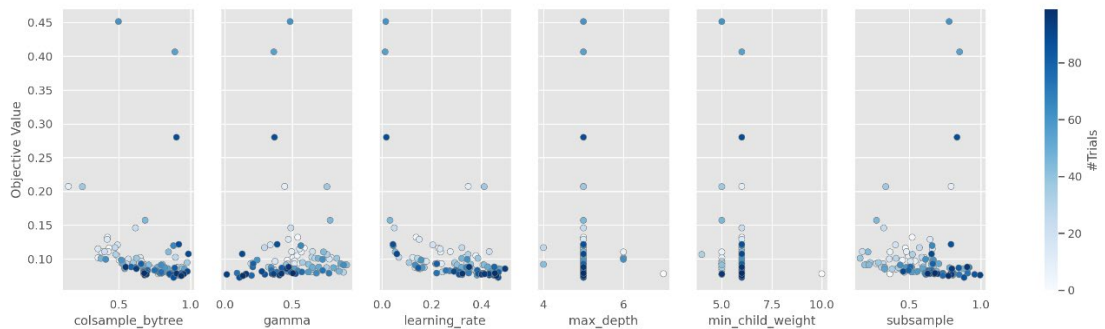


图 5.29 XGBoost 模型中不同超参值与目标函数之间的关系（模型目标：冷机群分项能耗）

从图 5.30 中可以看出可以看出超参调优的结果已经收敛，而且其训练结果已经远远小于 0.1 ($CV-RMSE < 0.1$)，最后模型的训练结果见表 5.15，其中 99 表示的是第 100 次迭代之后的全局最优解下的结果，eval_cvrmse 表示的是测试集数据的结果，train_cvrmse 表示的是训练集测试的结果，从训练结果上可以看出，超参数调优尽可能规避掉了模型输出带来的偏差。

表 5.15 预测冷机群分项能耗的 XGBoost 模型进行超参调优的结果值

No. Trial	eval_cvrmse	train_cvrmse
99	0.07264	0.06179

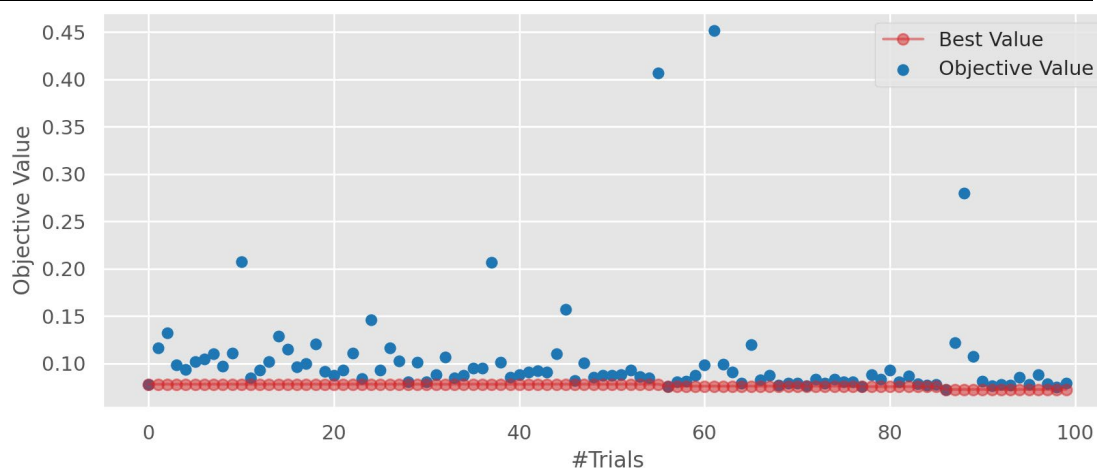


图 5.30 XGBoost 模型超参调优历史折线图（模型目标：冷机群分项能耗）

经过 100 轮迭代之后的最优化超参解见表 5.16，与前文的 XGBoost 模型的超参数相比，对应最优超参数与之间有很大差别，一方面是因为模型预测任务发生了改变，另一方是训练数据发生了改变，即使同样的模型争对同样的预测任务，如果训练数据发生变化，模型的超参数仍需要再做调整。

根据图 5.31 我们不难发现，影响预测冷机群分项能耗的模型的重要的超参数有 colsample、learning_rate、gamma 以及 subsample，与前文的 XGBoost 模型比较，发现重要性排序有一定的改变。

表 5.16 针对预测冷机群分项能耗的 XGBoost 模型的超参调优的结果

超参数	最优值	超参数	最优值
Learning_rate	0.4678	Min_child	6
Max_depth	5	Subsample	0.8828
Gamma	0.1079	Colsample_bytree	0.8823

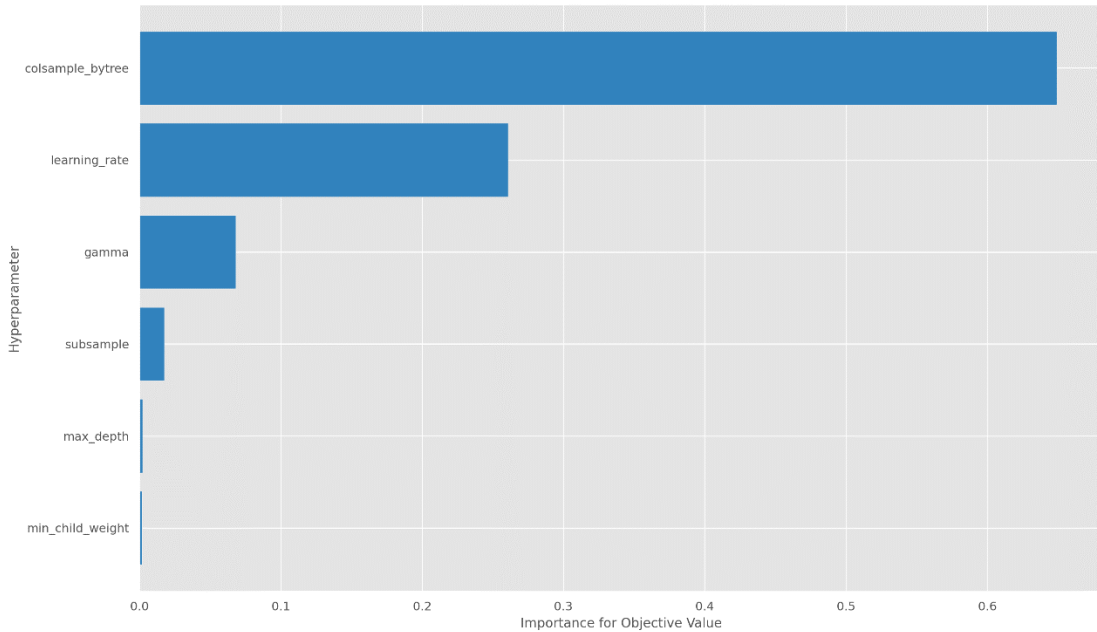


图 5.31 XGBoost 模型的超参重要性排序（模型输出：冷机群分项能耗）

5.1.5.4 基于关键变量推测的浅诊断

本部分将继续采用四种组合：PSO+CV-RMSE，PSO+DTW，GA+CV-RMSE 和 GA+DTW，推测变量并从结果准确性和时间效率角度进行分析，其中 PSO 采用 PYSWARM^[57]实现，GA 采用 GENETICALGORITHM^[58]实现。

为保证一定的比较性，笔者讲最大迭代次数进行调整：在 GENETICALGORITHM 中把表征迭代次数的 max_num_iteration 参数调成 50，在 PYSWARM 中把表征迭代数的 maxiter 参数调成 50（在保证结果达到收敛的情况下）。从图 5.32 和图 5.33 均可看出训练结果很快达到收敛，并且在图 5.32 中可以明显看出来以 CV-RMSE 为指标，其结果可以远小于 ASHRAE Guideline 14 的校准标准，在用 PSO 作为具体全局随机优化的算法中，无论是采用 CV-RMSE 还是采用 DTW 作指标，均在 10 次迭代内完成了收敛。其预测结果和时间效率见表 5.17，不同的方法组合的结果都相对靠近，CV-RMSE 和 DTW 相比时间效率较高，但采用 DTW 推测结果更接近真实值。

当推测变量不同时，如果追求极致准确性，可能需要采取不同的组合，在建筑内热负荷分析中推测照明密度、设备密度和人员密度时，CV-RMSE 相较于 DTW 准确性更高且时间效率更令人满意，但并不是在任何情况都能找到如此理想的组合，某些情况可能需要在时间效率和结果准确性上作一定的权衡和取舍。

表 5.17 用不同组合方法推测冷站设备关键变量的比较

No.	类型	冷机综合 COP	冷机劣化系数	时间%H:%M:%S
	真实值	5.41	0.98	
1	GA+CV-RMSE	5.669	0.937	00:00:27.332

2	GA+DTW	5.621	0.936	00:01:03.664
3	PSO+CV-RMSE	5.680	0.939	00:00:18.761
4	PSO+DTW	5.665	0.940	00:01:44.822

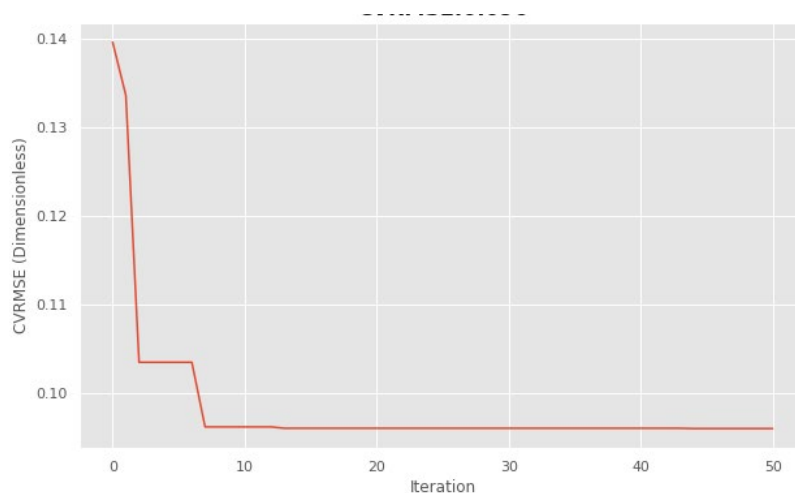


图 5.32 GA+CV-RMSE 预测冷站设备关键变量历史图

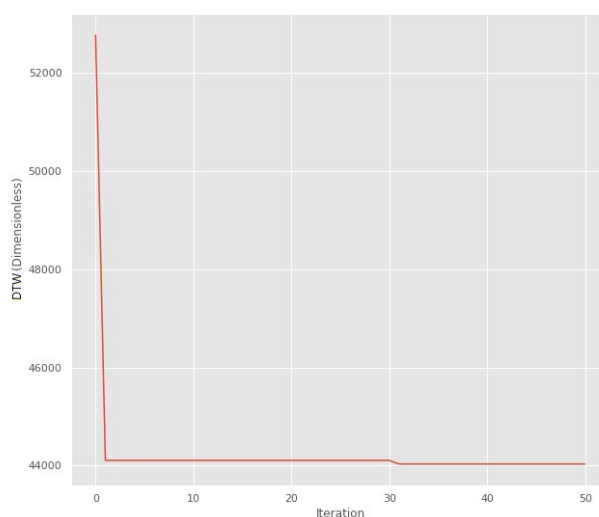


图 5.33 GA+DTW 预测冷站设备关键变量历史图

5.1.6 案例节能潜力分析总结

根据上述的节能潜力分析方法的分步验证，将节能潜力分析总结成表 5.18。

表 5.18 案例 1 节能潜力分析总结

案例信息	建筑类型	政府办公楼
	地理位置	中国上海
	建筑面积	49710.47m ²
	建筑楼层	27 层
	基本设备信息	√

		设备铭牌	×
		冷站总能耗数据	√
	数据信息	冷机群总能耗数据	√
		其他分项数据	×
	诊断内容	√ 建筑内热负荷分析	√ 设备选型优化
		√ 运维-浅诊断	× 运维-深诊断
		节能潜力分析依据	节能潜力分析建议
		窗墙比 0.545	无
		建筑体形系数 0.15	建筑体形系数在合理范围内
	建筑内热负荷分析	人员密度 10.966	无
		照明功率密度 22.824	照明功率密度较高，不利于建筑节能，建议检查照明系统
		设备功率密度 20.598	设备功率密度较高，不利于建筑节能，建议检查常用设备
	设备选型优化	冷机选型方法建议： 冷机#1: 930KW 冷机#2: 1700KW 冷机#3: 4050KW	
		预计节能目标	预计理想情况节电 13.38%
		节能潜力分析依据	节能潜力分析建议
	冷站运维	浅诊断	冷机劣化系数 0.937 冷机综合 COP 5.669
		深诊断	冷机保养良好
		节能潜力分析建议	无分项数据，不支持节能潜力分析
		1. 检查灯具	
		2. 检查设备	
		3. 建议优化冷机选型方案	

5.2 案例二

5.2.1 建筑描述与数据介绍

5.2.1.1 建筑描述

涉及隐私等相关因素，笔者并未获取全面的建筑信息，仅知道该建筑处于广西南宁地区，该建筑属于综合体（具体功能未知），建筑面积 92677m²。无建筑图纸、审计报告等三方资料。

在建筑暖通空调系统中，有具体设备较为详细的信息，见表 5.19。

表 5.19 广西南宁某综合体冷站核心设备关键信息

设备	数量	设备参数			
离心式冷水机组 1	3	制冷量	3868KW	消耗功率	721.9KW
		冷冻水供水温度	6℃	冷冻水回水温度	12℃

		冷却水供水温度	37℃	冷却水回水温度	32℃
离心式冷水机组 2	1	制冷量	1363.9KW	消耗功率	257.3KW
		冷冻水供水温度	6℃	冷冻水回水温度	12℃
		冷却水供水温度	37℃	冷却水回水温度	32℃
		额定流量	196m ³ /h	额定扬程	42m
冷冻水泵 1	2	额定功率	37KW	类型	变频
		额定流量	553m ³ /h	额定扬程	42m
冷冻水泵 2	3	额定功率	110KW	类型	变频
		额定流量	282m ³ /h	额定扬程	32m
冷却水泵 1	2	额定功率	45KW	类型	变频
		额定流量	789m ³ /h	额定扬程	32m
冷却水泵 2	3	额定功率	90KW	类型	变频
		冷却能力	2298KW	额定水量	395 m ³ /h
冷却塔 1	6	入口水温度	37℃	出口水温度	32℃
		风机功率	5.5KW	风机台数	3
		冷却能力	1686 KW	额定水量	290 m ³ /h
冷却塔 2	1	入口水温度	37℃	出口水温度	32℃
		风机功率	7.5KW	风机台数	2

5.2.1.2 数据描述

负责该综合的三方机构提供了 2020 年 5 月至 11 月的冷站运维的分项数据，数据清单见表 5.20。

表 5.20 广西南宁综合体分项数据清单

冷机系列	单个冷机逐时功耗	单个冷机冷冻水供水温度	单个冷机冷冻水回水温度
	单个冷机冷却水供水温度	单个冷机冷却水回水温度	
冷冻水泵系列	单个冷冻水泵逐时功耗	单个冷冻水泵频率	
冷却水泵系列	单个冷却水泵逐时功耗	单个冷却水泵频率	
冷却塔系列	单个风机逐时功耗	单个风机频率	
系统层级	冷量	冷冻水干管总流量	冷却水干管总流量
	冷冻水干管供水温度	冷冻水干管回水温度	冷却水干管供水温度
	冷却水干管回水温度	冷机群逐时总能耗	冷冻水泵逐时总能耗
	冷却水泵逐时总能耗	冷却塔逐时总能耗	

5.2.2 案例验证内容

从案例二的描述中不难发现，在信息化程度上，**案例二和案例一形成完美互补**，案例一知道比较完备的建筑信息但少有分项计量数据，但案例二因为隐私问题，建筑信息未知，但是有较为完整的描述系统运行的分项计量数据，所以案例二用来验证冷站运维节能潜力分析——深诊断部分的验证。

5.2.3 深诊断方法验证——全楼宇分项计量数据数据清洗流程实例

5.2.3.1 各设备群逐时能耗分析

在与工程师的沟通和数据分析后，可以看出各个设备群的总电表置信度较高，因为这是能耗账单的主要依据。图 5.34 中，第一排四个子图分别为冷机、冷却塔、冷冻水泵和冷却水泵逐时总能耗的直方图，未经任何处理。第二排为四个设备群在剔除零值后的逐时总能耗的直方图。设备群的逐时总能耗质量相对较优，可以作为分项数据数据清洗流程的开端，作为校准其他分项数据的标定。

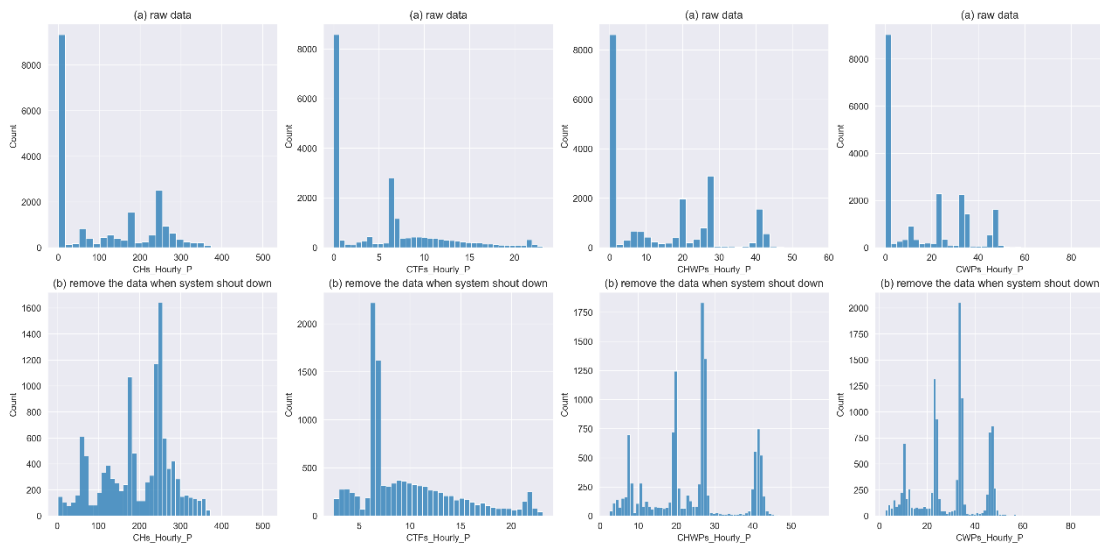


图 5.34 设备群逐时总能耗的直方图

5.2.3.2 冷机系列数据处理

根据基尔霍夫定律，理论上每个时刻单台冷机的逐时总能耗求和应与该时刻冷机群总电表的能耗值相同，为了能够充分利用分项数据，我们设置了可容忍误差范围为 $[0.75, 1.25]$ 。图 5.35 从左到右依次为 (a) 数据未处理时数据间的关系，(b) 仅剔除空值后两者之间的关系，(c) 根据基尔霍夫定律处理后两者之间关系。



图 5.35 基于冷机群总能耗对单台冷机逐时能耗的处理

在对能耗数据清洗后，处理冷机设备启停的数据，处理前与处理后的数据见图 5.36：左侧是未经处理的数据，从上至下四张子图分别代表四台冷水机组，能耗值和实际记录的冷机启停状态并不对应，违背专家经验。

根据清洗后的单台冷机能耗数据对冷机启停的数据进行修正，见右侧的四张子图。

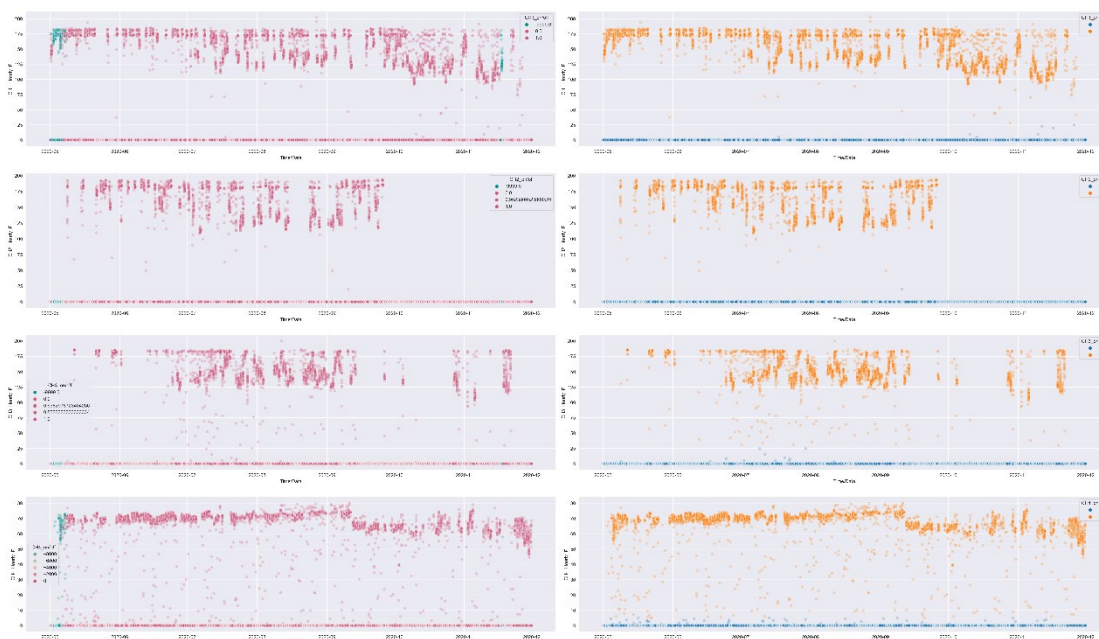


图 5.36 处理前与处理后的冷机启停与冷机能耗之间的关系

5.2.3.2 冷冻/冷却水泵系列数据处理

在处理完冷机能耗数据之后，对单个冷冻/冷却水泵能耗数据进行清洗，继续沿用基尔霍夫定律，见图 5.37 和图 5.38：从左至右依次 (a) 未经处理过数据，(b) 仅剔除空值后两者之间的关系，(c) 利用基尔霍夫定律处理后两者之间的关系。

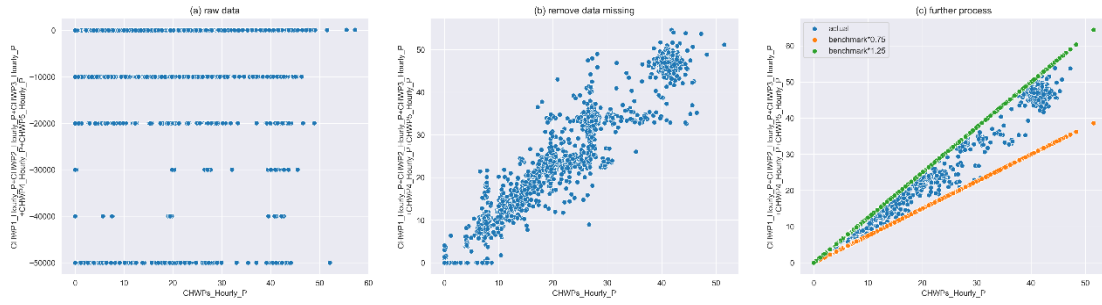


图 5.37 基于冷冻水泵群总能耗对单台冷冻水泵逐时能耗的处理

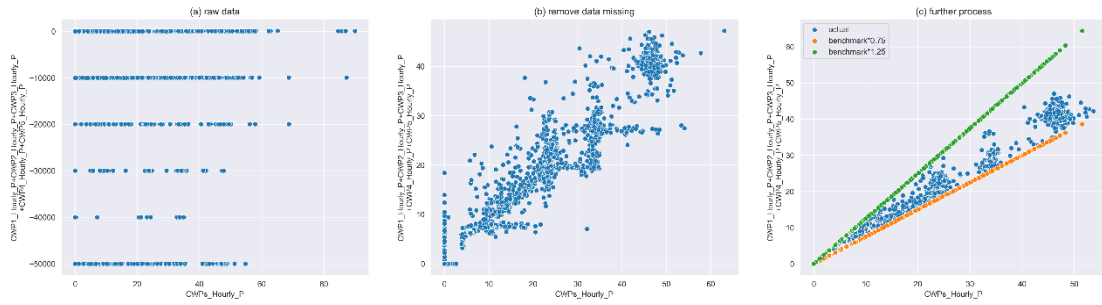
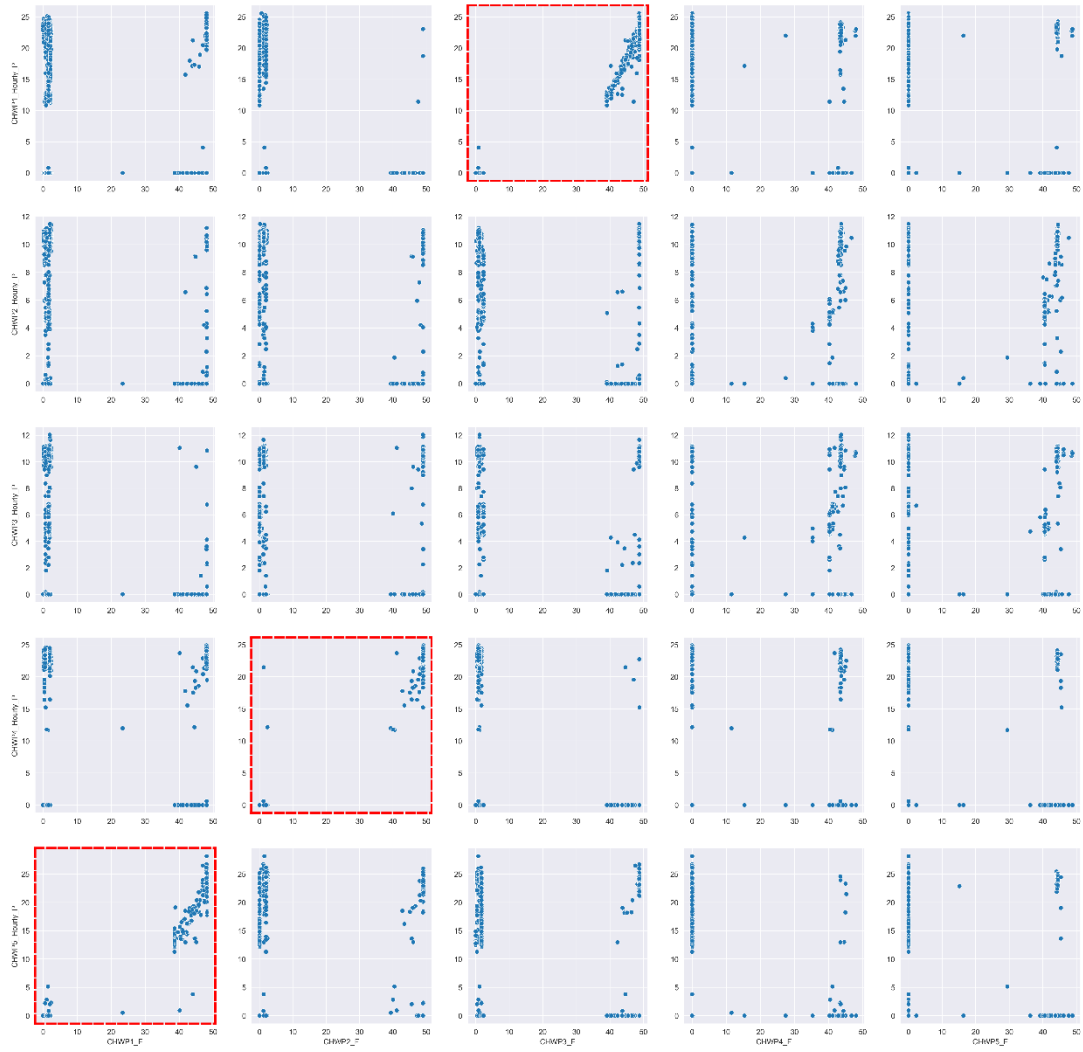


图 5.38 基于冷却水泵群总能耗对单台冷冻水泵逐时能耗的处理

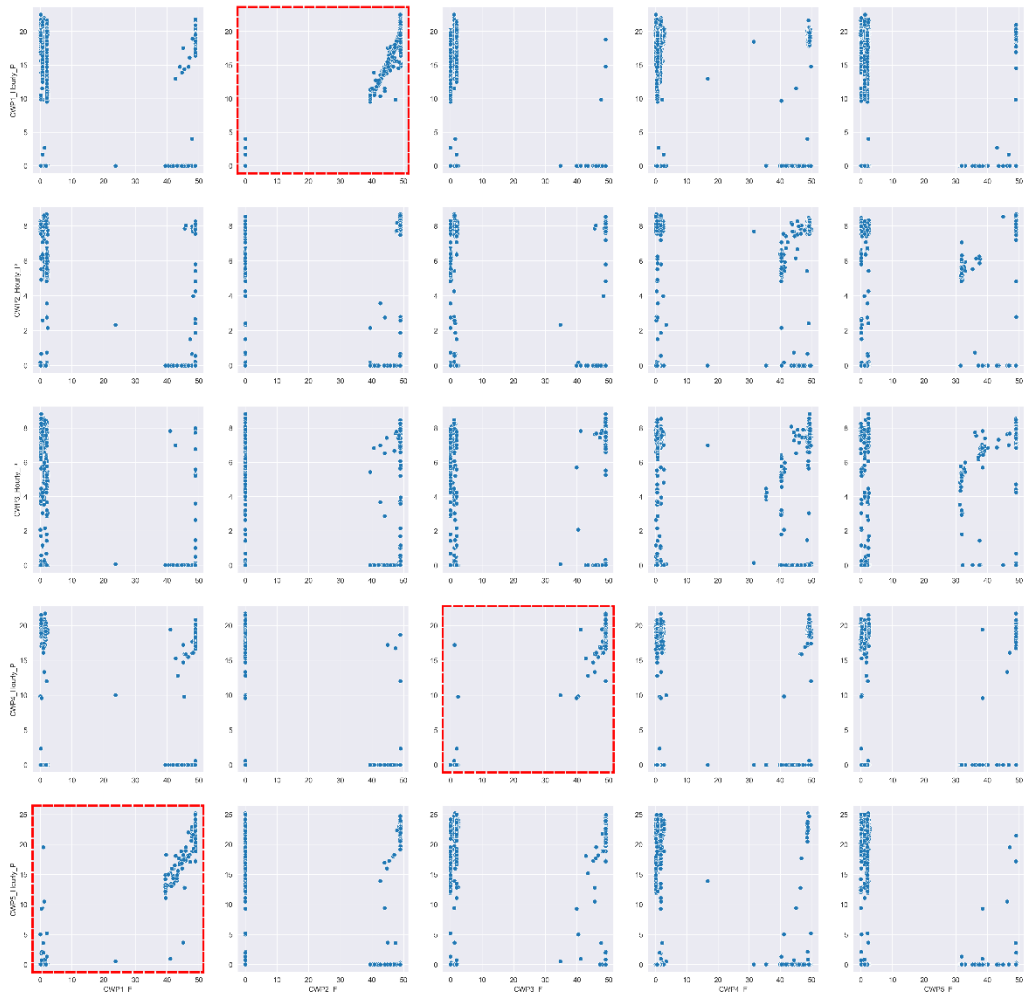
在此之后，我们需要处理单台水泵的能耗和频率之间关系，根据流体力学中相似性原理清洗数据。冷冻/冷却水泵的标签不对应的问题见图 5.39（上）和图 5.40（上）。以图 5.39 为例， 5×5 的（上）子图中横坐标分别为冷冻水泵 1~5 的频率，纵坐标为冷水泵 1~5 的逐时功耗，我们可以发现较为水泵 3 的频率和水泵 1 的功耗符合相似定律，同理水泵 2 的频率和水泵 4 的功耗，水泵 1 的频率和水泵 5 的功耗。 5×5 的（下）子图是将标签对应后的关系图。水泵 2 和水泵 3 的频率和功耗没有特别明显的对应关系，但从功率上可以看出这两台冷冻水泵均为传输能力较小的水泵，其中一个还是备用泵，这两台小水泵可能为交替开启的状态。上述分析同样适用于图 5.40。



第 5 章 基于实际建筑案例的建筑节能潜力分析验证



图 5.39 处理前（上）和处理后（下）冷冻水泵频率和功耗的标签对应关系图



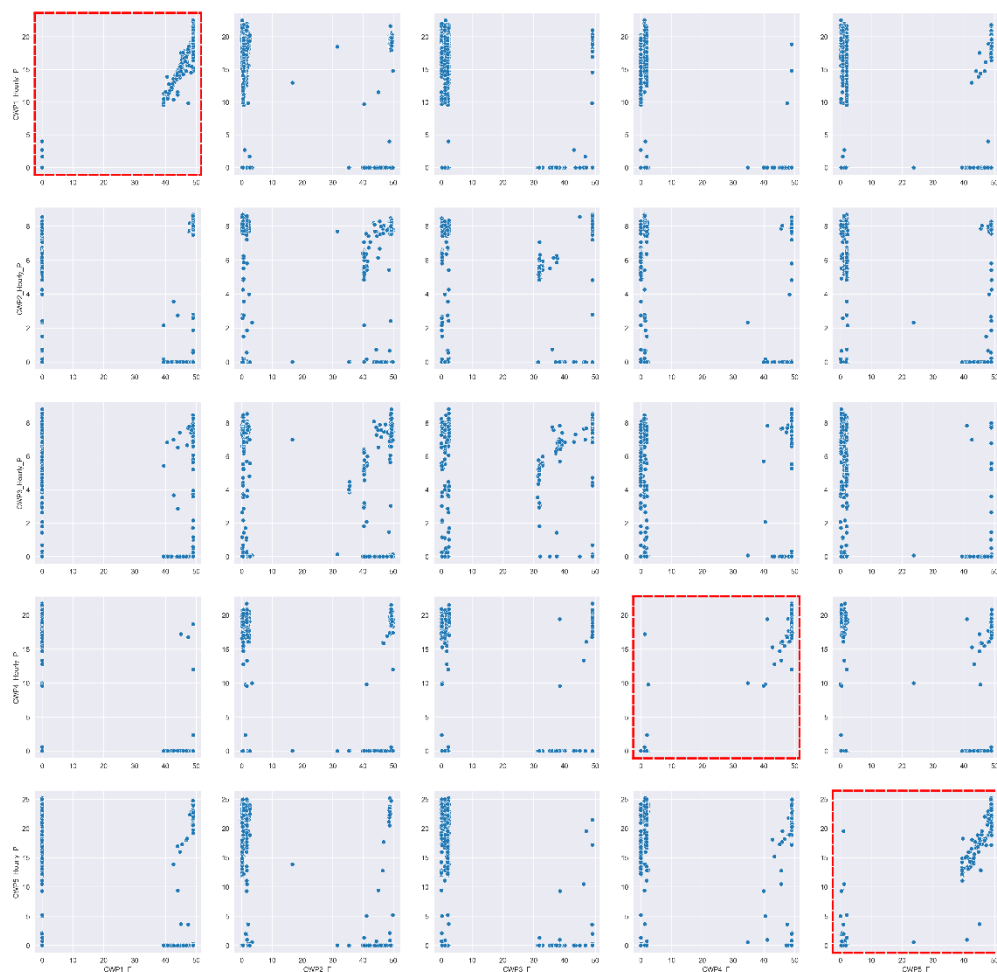


图 5.40 处理前（上）和处理后（下）冷却水泵频率和功耗的标签对应关系图

5.2.3.3 冷却塔系列数据处理

冷却塔系列数据仅包含单个风机的频率和功率数据，并且风机和冷却塔对应关系无从得知，因为原始标签信息模糊且每个风机是可以独立启停控制的。从图 5.41 中不难看出，冷却塔系列数据问题较为严重，自身数据丰富性不足，数据又存在大面积缺失，很难诊断冷却塔性能是否出现劣化。但幸运的是，从图 5.42 的风机频率的相关系数热力图中可以发现冷却塔的一些运行规律，所有的冷却塔应该是同开同关同时调频的，但这需要和现场工程师进一步核实。

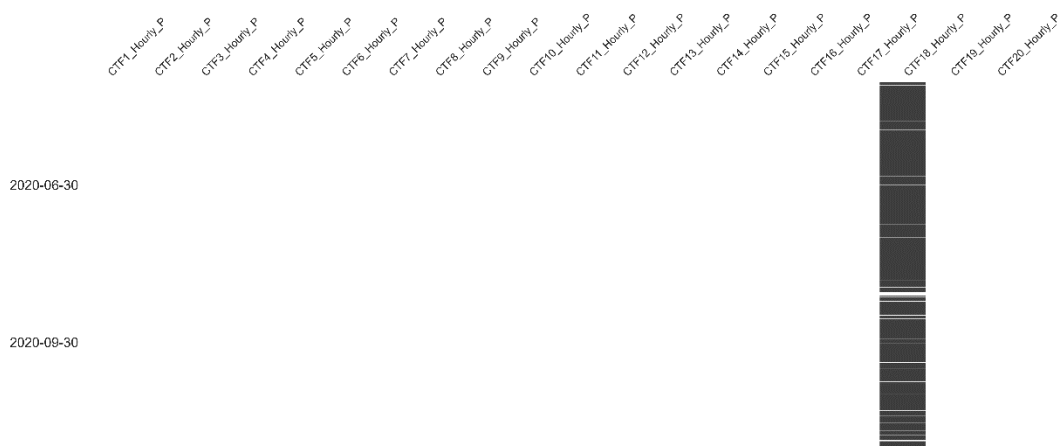


图 5.41 冷却塔风机能耗的缺失问题

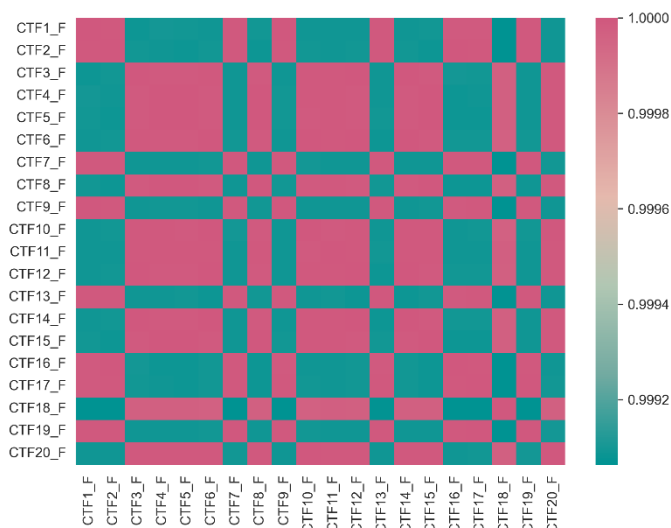


图 5.42 冷却塔风机频率的 Pearson 相关系数矩阵

5.2.3.4 系统层级数据处理

根据清洗后的冷冻/冷却水泵频率，利用相似准则可以近似计算出额定情况下冷冻侧和冷却侧干管环路的流量，但实际管路特性可能与理想情况下的并不相同。但是作为参考，结合专家知识，我们可以初步判断流量是否合理，管道是否存在阻塞的情况。

图 5.43 展示了冷冻水和冷却水流量之间关系，左上角子图中橙色散点表示用相似准则计算的冷冻水流量，蓝色散点表示实测冷冻水流量，主要差别出现在 11 月份，实测冷冻水干管流量突然升高明显不合理，所以将 11 月份的数据全部剔除。此外在左下角子图中，橙色散点为实测冷却水流量，蓝色散点为实测冷冻水量，两者相差数倍，但右下角子图表明冷冻水/冷却水干管温差均处于 5℃ 的水平，那么流量差距属于不合理范围，采用能量守恒定理将冷却水干管流

量进行修正，修正后的结果见图 5.44。从图 5.45 中也可以看出：修正干管流量前，数据之间是不满足能量守恒定理的（冷却侧和冷冻侧在能量交换上无法耦合）。

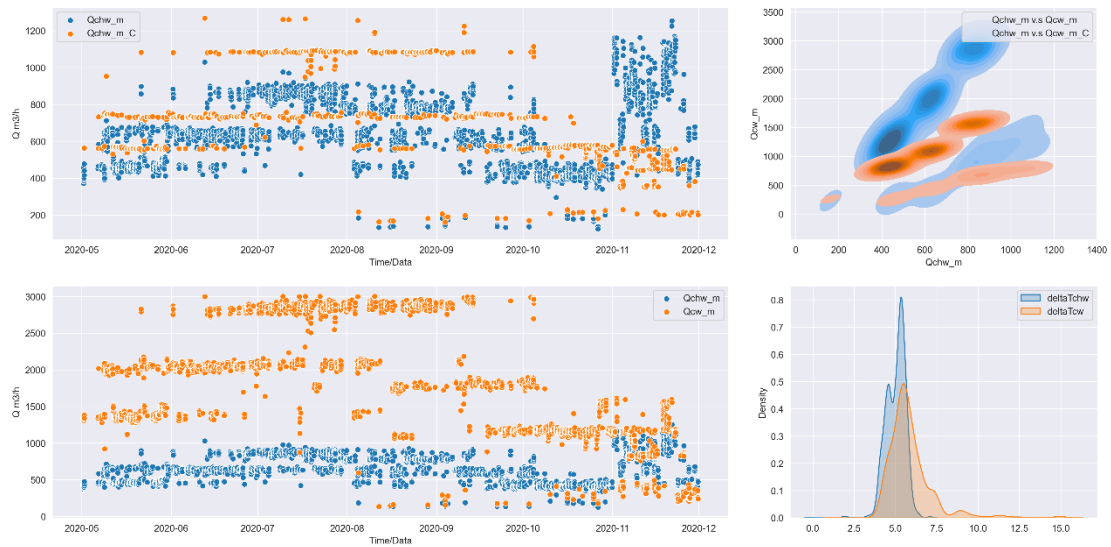


图 5.43 修正前冷冻/冷却水环路干管流量之间的关系

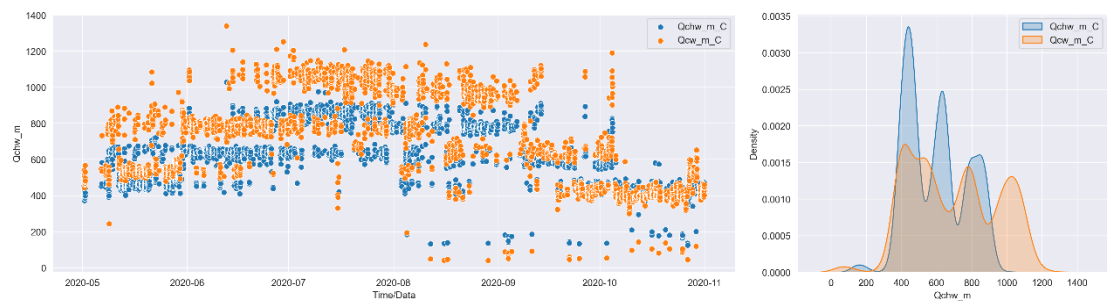


图 5.44 修正后冷冻/冷却水环路干管流量之间的关系

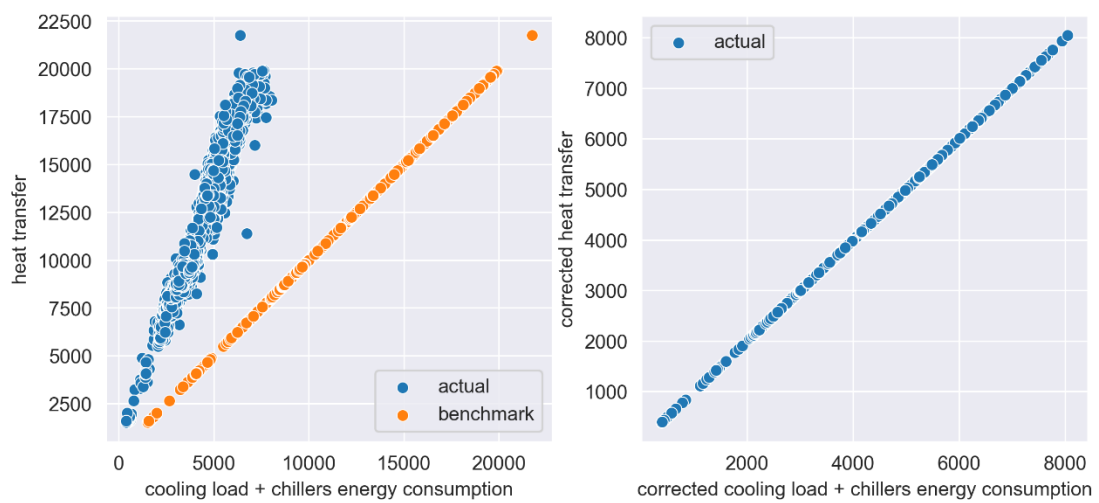


图 5.45 修正前（左）后（右）能量守恒定理的满足情况

5.2.4 深诊断方法验证——能效问题诊断

5.2.4.1 冷机相关问题

根据修正后的数据（图 5.46），我们可以直接判断冷机启停是否异常，我们可以发现在冷机开启的方式中存在一定的节能潜力：

- 1) 原先“开启一台大冷水机组一台小冷水机”的情况可以改为“仅开启一台大冷水机”；
- 2) 原先“开启两台大冷水机组”的大部分情况可以更改为“开启一台大冷水机组一台小冷水机组”；
- 3) 原先“开启两台大冷水机组一台小冷水机组”的大部分情况可以改为“开启两台大冷水机组”，这样可以进一步提高 PLR 从而实现这一部分节能潜力。

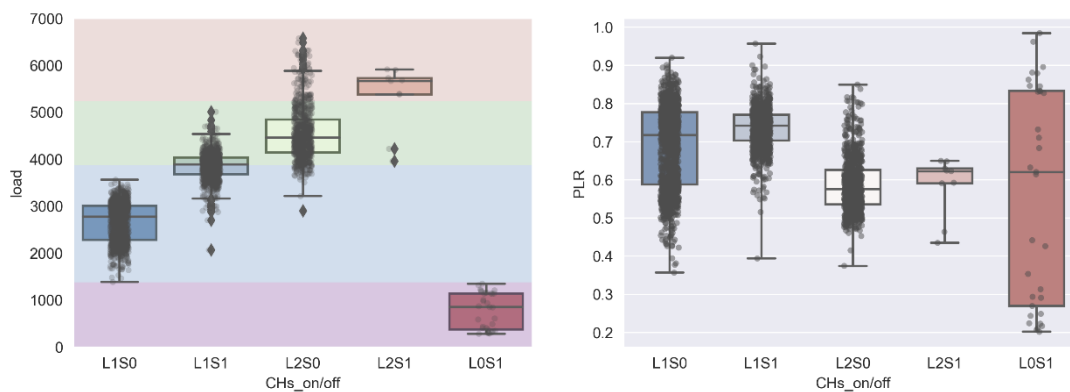


图 5.46 冷机开启情况和负荷之间的关系（左）冷机开启情况和 PLR 之间的关系（右）

图 5.47 表示的是四台冷水机组在开启状态的进出口温度核密度图，红色虚线表示额定工况下的温度，从中不难看出冷冻水进出水温度都偏高，冷却水回水温度保持较好。

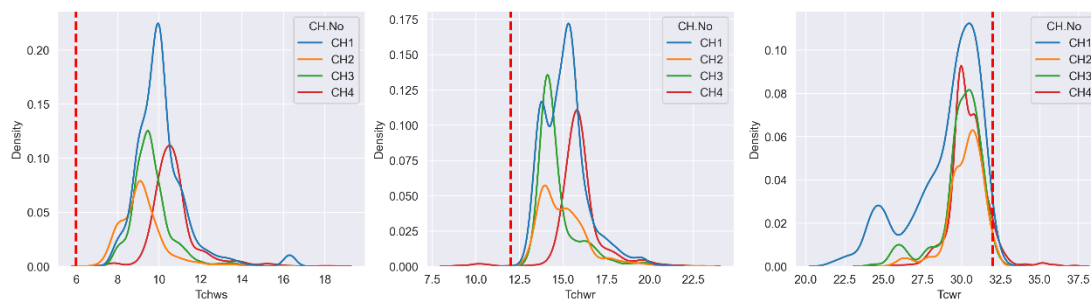


图 5.47 冷冻水供水温度（左）、冷冻水回水温度（中）和冷却水回水温度（右）核密度图

5.2.4.2 冷冻/冷却水泵相关问题

在图 5.49 中，左侧子图系列是冷冻水泵和冷机启停之间关系，两张子图为一组，左上角第一张子图时冷机启停与各个冷冻水泵功率之间的关系，左上角第二张子图时冷机启停与各个冷冻水泵频率之间的关系，其余以此类推。

不论从频率和功率上分析，水泵和冷机的联动关系是一致的，例如冷冻水泵 1 对应冷水机组 1，冷冻水泵 4 对应冷水机组 2，冷冻水泵 5 对应冷水机组 3，均为“大水泵对应大冷机”，小水泵 2、3 则是交替配合小冷机 4。右侧有关冷却水泵对应冷机的子图可以反应同样的规律：冷却水泵 1 对应冷水机组 1，冷却水泵 4 对应冷水机组 2，冷却水泵 5 对应冷水机组 3，冷却水泵 2、3 交替对应冷水机组 4，同样也是满足“大冷机对应大水泵、小冷机对应小水泵”的原则。另外，图 5.48 应证了运行阶段“大流量小温差”的状况不太明显。

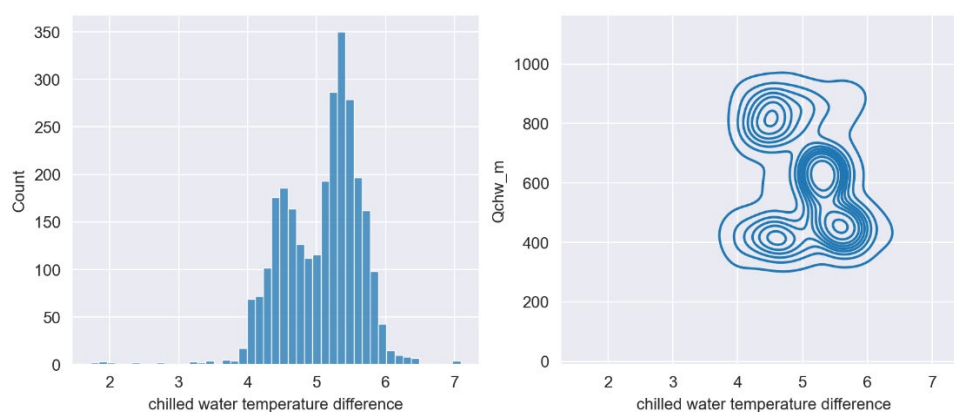


图 5.48 系统运行时冷冻水干管供回水温差直方图（左）和冷冻水干管供回水温差和冷冻水干管流量的关系（右）

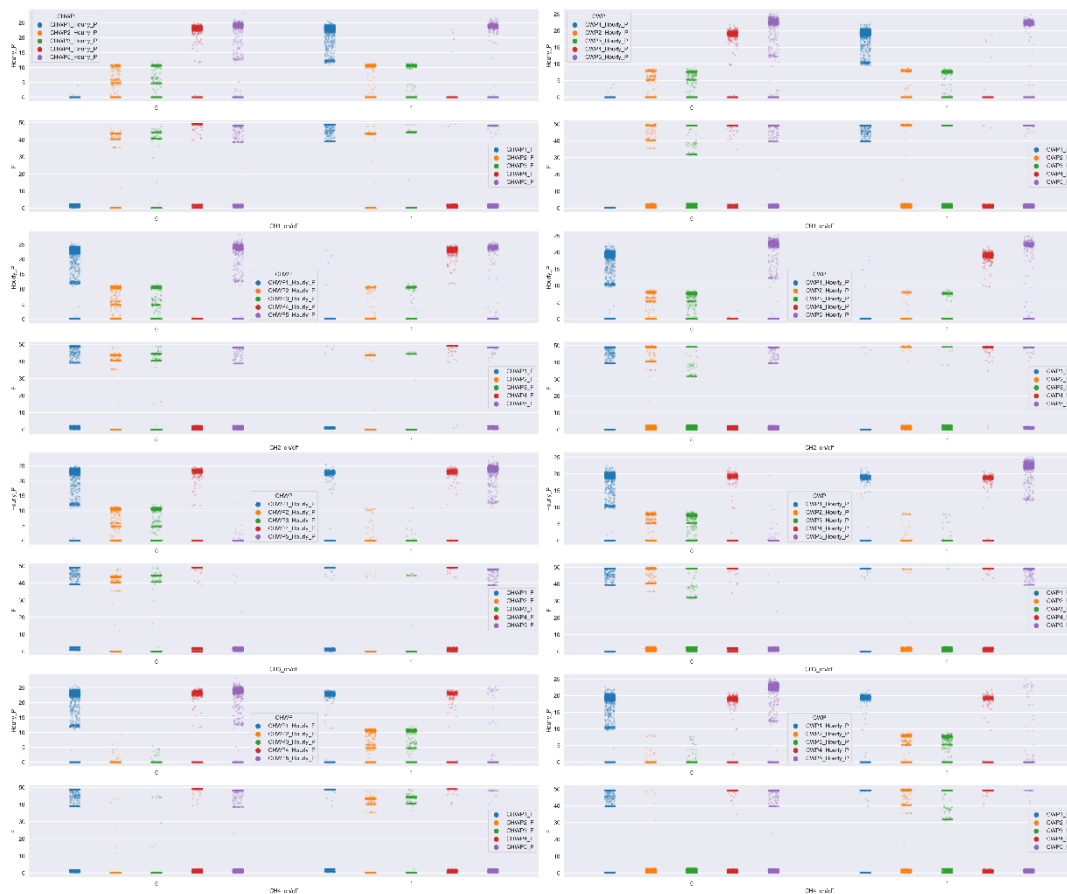


图 5.49 冷机启停与冷冻（左）/冷却（右）水泵频率、功率之间的关系

在图 5.50 和图 5.51 中，黄色表示理想工况情况而蓝色表示实际情况，不论是冷冻侧环路还是冷却侧环路均发现，实际情况和理想工况均存在一定差距。

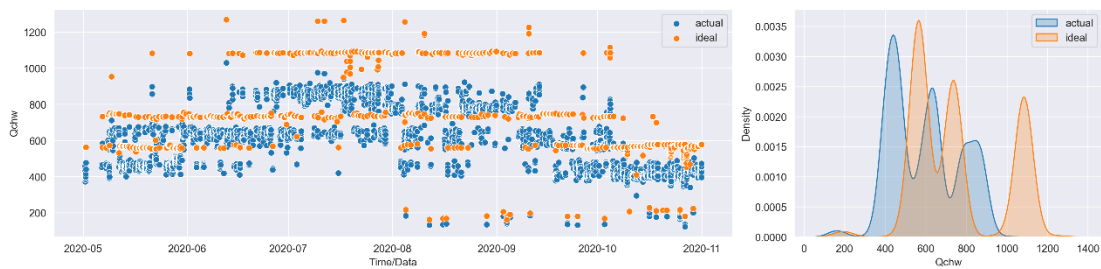


图 5.50 系统运行时冷冻水主干实测流量和理想工况下的流量关系



图 5.51 系统运行时冷冻水干管实测流量和理想工况下的流量关系

5.2.4.3 冷却塔相关问题

在数据预处理框架中已经明确说明：冷却塔数据信息的贫乏阻碍对冷却塔性能的分析，但是我们可以看到冷却塔所有风机时同开同关的，充分利用了冷却塔换热面积^[59]。图 5.52 表明了冷却水供回水温差一直保持在 5℃左右，冷却塔运行情况良好，若想要进步在冷却侧进行节能，可以将定温差控制改为逼近度控制。

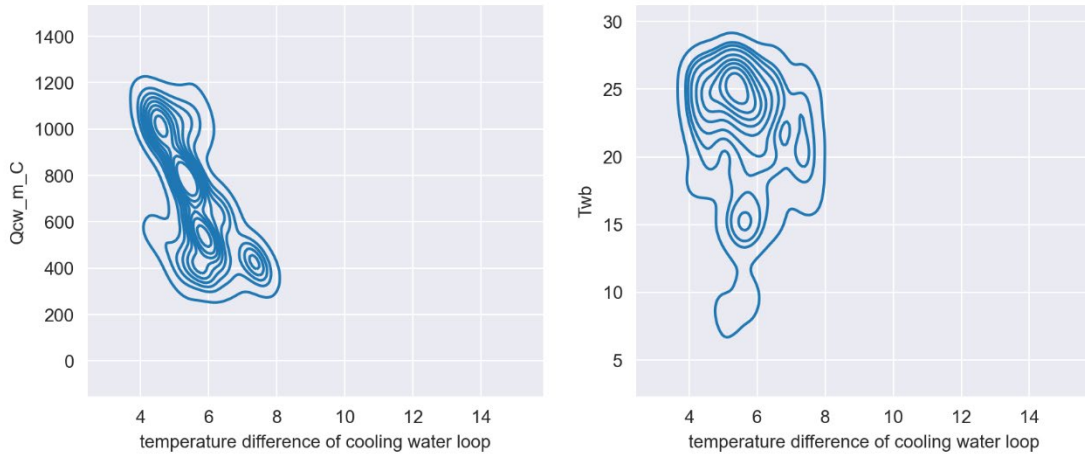


图 5.52 系统运行时冷却水干管供回水温差和冷却水干管流量的关系（左）和冷却水干管供回水温差和室外湿球温度的关系（右）

5.2.5 案例节能潜力分析总结

将上述可视化分析的结果进行汇总形成表 5.21。

表 5.21 案例 2 节能潜力分析总结

问题类别	具体问题	节能潜力分析依据	节能潜力分析结果及建议
冷机系列	开启台数异常	图 5.44	冷机开启台数存在节能潜力空间：1) “开启两台大冷机和一台小冷机”的情况可以改成“开启两台大冷机”，2) “开启两台大冷机”的情况可以改为“开启一台大冷机和一台小冷机”，3) “开启一台大冷机和一台小冷机”的情况可以改为“开启一台大冷机”
	冷冻水供水温度异常	图 5.45	冷冻水供水温度过高
	冷冻水回水温度异常	图 5.45	冷冻水回水温度过高

	冷却水回水温度异常	图 5.45	无
冷冻水泵系列	开启台数异常	图 5.46	无
	大流量小温差	图 5.47	无
	冷冻侧环路阻塞	图 5.48	不确定，理想工况和实际工况确实存在差距，但是缺少压力数据进一步判断
冷却水泵系列	开启台数异常	图 5.46	无
	冷却侧环路阻塞	图 5.49	不确定，理想工况和实际工况确实存在差距，但是缺少压力数据进一步判断
冷却塔系列	冷却塔换热性能变差	图 5.50	不确定，但冷却侧环路温差能达到名义温差，需要数据进一步判断，控制逻辑可以改为逼近度控制。
节能潜力分析建议			<ol style="list-style-type: none"> 1. 冷机开启台数在建筑负荷较大的时候需要优化，具体建议见上述表格 2. 进一步检查实测水量低于名义水量的原因，水泵自身劣化还是环路出现阻塞 3. 可以将控温差的冷却塔控制方式改为逼近度控制方式

5.3 本章小结

本章通过两个案例验证了本文提出的公共建筑节能潜力分析方法的实用性和准确性。

第一个建筑案例有较为全面的建筑信息但缺少分项数据的建筑案例，通过该案例验证了建筑内热负荷分析、冷站设备选型方案优化以及冷站运维能效诊断——浅诊断的准确性，其中分析总结了黑箱模型超参数调优的规律：虽然采用同样的黑箱模型，若模型的任务发生变化或训练数据出现增减，模型对不同超参数的敏感性会发生变化，自动优化黑箱模型超参数是提高分析精度的关键步骤。

第二个建筑案例无法获知全面的建筑信息，但是有较为丰富的分项数据描绘建筑的运维画像，与第一个案例形成互补，通过第二个案例验证本文提出的深诊断中分项计量数据的数据清洗流程和框架，清洗后的数据可以直接用于分析各种常见的能效问题。

第 6 章 总结与展望

6.1 本文总结

本文提出一种基于数据挖掘的公共建筑节能潜力分析方法，涵盖从建筑内热负荷、设备选型优化、冷站运维的三个方面，充分利用不断发展并完善的建筑大数据加速传统的既有公共建筑能源审计的速度。

- 1) 关于建筑内热负荷的分析：在影响建筑能耗的关键变量的研究成果上，采用 Grasshopper 和 Eppy 生成建筑-能耗数据库，以此为基础训练并优化黑箱模型，后接全局搜索算法推测关键变量，根据关键变量的推断结果分析既有建筑的内热负荷构成。
- 2) 关于设备选型方案的节能潜力分析：在已知影响建筑负荷的关键变量真值的前提下，判断设计阶段的冷机选型方案与建筑负荷的匹配度，用平均 PLR 衡量匹配程度，后采用全局搜索算法优化冷水机组名义制冷量。
- 3) 关于冷站运维的节能潜力分析：本文提出浅诊断-深诊断相结合的节能潜力分析架构。浅诊断利用冷机总能耗数据，通过建模和变量推断的方式分析冷机自身性能劣化的问题。深诊断在对分项计量数据清洗后，分析设备以及系统的能效问题。
- 4) 在黑箱模型的超参数调优方面，输出为冷站总能耗的 XGBoost 模型对 `colsample`、`learning_rate` 这两个超参数最敏感；另外模型的超参数会因为任务、数据变化等原因，在重要性排序上会发生变化。

6.2 主要贡献

- 1) 改进了基于 Gaussian Process 的关键变量推测算法，本文采用全局随机搜索算法推测关键变量，优势在于：在不牺牲推测精度的前提条件下，可同时推测多个关键变量，且时间效率得到提升。
- 2) 本文引入一种针对数据驱动模型的基于树模型的超参数调优方法，以求尽可能减少黑箱模型的误差。数据驱动模型对超参数的高敏感特性决定了在建筑大数据积累的情况下，数据驱动模型的超参数需要根据情况的变化自动调整。本文以泛化能力得到广泛认可的 LightGBM、XGBoost 模型为例，当模型存在诸多超参数需要调节时，若训练数据或预测目标发生变化，即便采用同一模型，模型对不同超参数的敏感性也会随之发生变化。

- 3) 本文对分项计量数据进行缺失性分析，在“特性中寻找共性”，提出一套相对通用的，基于数据交叉验证的针对建筑冷站分项计量数据的数据清洗框架。根据本文提出的数据清洗框架清洗后的数据可直接用作对设备、系统的能效诊断中。

6.3 局限性与展望

- 1) 建筑围护结构的保温性能也会影响建筑负荷，并且随着时间推移围护结构的保温性能可能会不断变化，但本文在建筑负荷分析中并未考虑围护结构保温性能下降对建筑冷站能耗的影响，可向该方向进一步研究。
- 2) 本文的多变量推测方法提高了时间效率，但部分变量的推测精度有待进一步提高。
- 3) 本文搭建的节能潜力分析多停留在定性层面，对定量分析涉及不深，因为缺乏同时拥有改造前后建筑冷站数据并悉知建筑改造细节的建筑案例；若后期有较为理想的建筑案例可进一步验证本文方法的有效性，也可进一步深入探究定量分析。
- 4) 本文在深诊断部分提出的针对分项计量数据的数据清洗框架中，并未纳入对压力数据的清洗方法，因为本研究缺少实际压力数据的支持。
- 5) 本文的验证案例一为政府办公建筑，该案例的人流相对稳定，减少了人员不确定性对方法准确验证的影响，本文提出的方法在商场建筑、商业综合体建筑中有效性和准确性需要在对应的建筑案例中进一步验证。
- 6) 本文的验证案例二的分项计量数据中，直接与冷机相关的测点点位相较于其他设备，如冷却塔的测点点位数量多且类型丰富，受限于数据量，并不是所有常见的空调系统、设备能效问题均能作出了诊断。

参考文献

- [1] Cristina Baglivo, Paolo Maria Congedo, Graziano Murrone, Dalila Lezzi. Long-term predictive energy analysis of a high-performance building in a mediterranean climate under climate change[J], *Energy*, Volume 238, Part A, 2022, 121641.
- [2] Xiaoma Li, Yuyu Zhou, Sha Yu, Gensuo Jia, Huidong Li, Wenliang Li. Urban heat island impacts on building energy consumption: A review of approaches and findings[J], *Energy*, Volume 174, 2019, Pages 407-419.
- [3] 中国建筑节能协会,能耗统计专业委员会.中国建筑能耗研究报告(2020)[R].2020.11.厦门.
- [4] Zhou, N., Khanna, N., Feng, W. *et al.* Scenarios of energy efficiency and CO₂ emissions reduction potential in the buildings sector in China to year 2050[J]. *Nat Energy* 3, 978–984 (2018).
- [5] 潘毅群,梁育民,朱明亚.碳中和目标背景下的建筑碳排放计算模型研究综述[J].暖通空调,2021,51(07):37-48.
- [6] 住房和城乡建设部节能与科技司,公共建筑能源审计导则[S],建科[2007]49号.北京,2016.
- [7] 上海市住房和城乡建设管理委,上海市发展改革委员会.2020年上海市国家机关办公建筑和大型公共建筑能耗监测及分析报告[R].2021.7.
- [8] 胡振中,袁爽.建筑能耗与环境监测系统标准化数据提取技术[J].清华大学学报(自然科学版),2020,60(04):357-364.
- [9] Data collection device and associated system for monitoring and storing performance and maintenance data related to a component of an electrical system[P]. United State. US9037428B2, 2015.
- [10] 潘毅群.实用建筑能耗模拟手册[M].第一版,北京,中国建筑工业出版社,2013:50-72.
- [11] Akshay Ajagekar, Fengqi You. Quantum computing based hybrid deep learning for fault diagnosis in electrical power systems[J], *Applied Energy*, Volume 303, 2021, 117628.
- [12] Shunian Qiu, Zhenhai Li, Zhengwei Li & Xinfang Zhang. Model-free optimal chiller loading method based on Q-learning[J], *Science and Technology for the Built Environment*, 2020, 26:8, 1100-1116.
- [13] Shunian Qiu, Zhenhai Li, Zhengwei Li, Jiajie Li, Shengping Long, Xiaoping Li. Model-free control method based on reinforcement learning for building cooling water systems: Validation by measured data-based simulation[J], *Energy and Buildings*, Volume 218, 2020, 110055.
- [14] Zhang, X., Li, Z., Li, Z. *et al.* Differential pressure reset strategy based on reinforcement learning for chilled water systems[J]. *Build. Simul.* 15, 2020, 233–248.
- [15] Cheng Fan, Yongjun Sun, Yang Zhao, Mengjie Song, Jiayuan Wang. Deep learning-based feature engineering methods for improved building energy prediction[J], *Applied Energy*, Volume 240, 2019, Pages 35-45.
- [16] Fan, C., Yan, D., Xiao, F. *et al.* Advanced data analytics for enhancing building performances: From data-driven to big data-driven approaches[J]. *Build. Simul.* 14, 2020, 3–24.
- [17] 田柏秋,潘毅群,黄治钟.离心式冷水机组经验回归模型的分析与评价[J].建筑节能,2014,42(02):15-20+52.

- [18] 高力强,朱丽,周海珠,周立宁,孙勇,刘其.基于运行数据的高校绿色建筑节能潜力分析[J].西安建筑科技大学学报(自然科学版),2017,49(03):422-426.
- [19] 李欣欣. 西安市某商场能耗分析与对策研究[D].西安建筑科技大学,2013.
- [20] 付乐林,王刚,夏积玉,徐科飞.青岛市某办公建筑能耗模拟与节能潜力分析[J].青岛理工大学学报,2018,39(05):86-90.
- [21] Zhihong Pang, Yan Chen, Jian Zhang, Zheng O'Neill, Hwakong Cheng, Bing Dong. Nationwide HVAC energy-saving potential quantification for office buildings with occupant-centric controls in various climates[J]. Applied Energy, Volume 279, 2020, 115727.
- [22] Yixing Chen, Chuhao Yang, Xiao Pan, Da Yan. Design and operation optimization of multi-chiller plants based on energy performance simulation[J], Energy and Buildings, Volume 222, 2020, 110100.
- [23] Tatiana Alves, Luiz Machado, Roberta Gonçalves de Souza, Pieter de Wilde. Assessing the energy saving potential of an existing high-rise office building stock[J], Energy and Buildings, Volume 173, 2018, Pages 547-561.
- [24] Tatiana Alves, Luiz Machado, Roberta Gonçalves de Souza, Pieter de Wilde. A methodology for estimating office building energy use baselines by means of land use legislation and reference buildings[J], Energy and Buildings, Volume 143, 2017, Pages 100-113.
- [25] Chaham Alalouch, Saleh Al-Saadi, Husam AlWaer, Kamel Al-Khaled. Energy saving potential for residential buildings in hot climates: The case of Oman[J], Sustainable Cities and Society, Volume 46, 2019, 101442.
- [26] Yue Pan, Limao Zhang. Data-driven estimation of building energy consumption with multi-source heterogeneous data[J], Applied Energy, Volume 268, 2020, 114965.
- [27] Marjan Ilbeigi, Mohammad Ghomeishi, Ali Dehghanbanadaki. Prediction and optimization of energy consumption in an office building using artificial neural network and a genetic algorithm[J], Sustainable Cities and Society, Volume 61, 2020, 102325.
- [28] Xinyue Li, Shuqin Chen, Hongliang Li, Yunxiao Lou, Jiahe Li. Multi-dimensional analysis of air-conditioning energy use for energy-saving management in university teaching buildings[J], Building and Environment, Volume 185, 2020, 107246.
- [29] Yibo Chen, Fengyi Zhang, Umberto Berardi. Day-ahead prediction of hourly subentry energy consumption in the building sector using pattern recognition algorithms[J], Energy, Volume 211, 2020, 118530.
- [30] Siriwarin Petcharat, Supachart Chungpaibulpatana, Pattana Rakkwamsuk. Assessment of potential energy saving using cluster analysis: A case study of lighting systems in buildings[J], Energy and Buildings, Volume 52, 2012, Pages 145-152.
- [31] Mingya Zhu. The construction of minimum variable set for energy prediction models of office building [D]. Tongji University,2018.
- [32] Huajin Sha. Building heterogeneous data integration method and the establishment of the hybrid energy model[D]. Tongji University,2020.
- [33] WIKIPEDIA. Grasshopper[EB/OL]. (2021.04.29). [2021.04.29]. <https://en.wikipedia.org/wiki/Grasshopper>
- [34] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System[J]. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery

- and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 2016, 785–794.
- [35] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: a highly efficient gradient boosting decision tree[J]. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 2017, 3149–3157.
- [36] Github. LightGBM-parameters tuning[EB/OL]. (2021)[2021].
<https://lightgbm.readthedocs.io/en/latest/Parameters-Tuning.html>
- [37] Edwin K.P Chong, Stanislaw H. Zak. An Introduction to Optimization[M]. Fourth Edition. 北京,电子工业出版社, 2015.10.
- [38] ASHRAE Standards Committee. ASHRAE Guideline 14, Measurement of Energy and Demand Savings[S]. ASHRAE, Atlanta (2002).
- [39] Paolo Tormene, Toni Giorgino, Silvana Quaglini, Mario Stefanelli. Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation[J], Artificial Intelligence in Medicine, Volume 45, Issue 1, 2009, Pages 11-34.
- [40] Ruikai He, Peng Xu, Zhibo Chen, Wei Luo, Zhineng Su, Jiong Mao. A non-intrusive approach for fault detection and diagnosis of water distribution systems based on image sensors, audio sensors and an inspection robot[J], Energy and Buildings, Volume 243, 2021, 110967.
- [41] Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. Practical Bayesian optimization of machine learning algorithms[J]. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'12). Curran Associates Inc., Red Hook, NY, USA, 2012, 2951–2959.
- [42] James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for hyperparameter optimization[J]. In Proceedings of the 24th International Conference on Neural Information Processing Systems (NIPS'11). Curran Associates Inc., Red Hook, NY, USA, 2011, 2546–2554.
- [43] Daniel Golovin, Benjamin Solnik, Subhodeep Moitra, Greg Kochanski, John Karro, and D. Sculley. Google Vizier: A Service for Black-Box Optimization[J]. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17). Association for Computing Machinery, New York, NY, USA, 2017, 1487–1495.
- [44] Moritz P, Nishihara R, Wang S, Tumanov A, Liaw R, Liang E, Elibol M, Yang Z, Paul W, Jordan MI, Stoica I. Ray: a distributed framework for emerging AI applications[J]. 13th Symposium on Operating Systems Design and Implementation, 2018, 561-577.
- [45] Nikolaus Hansen. The CMA Evolution Strategy: A Tutorial.
- [46] K. Jamieson and A. Talwalkar, Non-stochastic best arm identification and hyperparameter optimization[J], Proc. 19th Int. Conf. Artif. Intell. Statist., 2016, pp. 240-248.
- [47] Liam Li, Kevin Jamieson, Afshin Rostamizadeh, Ekaterina Gonina, Moritz Hardt, Benjamin Recht, and Ameet Talwalkar. Massively Parallel Hyperparameter Tuning[J]. NeurIPS Workshop on Machine Learning Systems. 2018.
- [48] U.S Department of Energy. Engineering Reference[EB/OL], EnergyPlus Version 9.5.0 Documentation, 2021.03.30.

- [49] Simon Henriët, Umut Şimşekli, Benoit Fuentes, Gaël Richard, A generative model for non-intrusive load monitoring in commercial buildings[J], Energy and Buildings, Volume 177, 2018, Pages 268-278.
- [50] Elnaz Azizi, Roya Ahmadihangar, Argo Rosin, Joao Martins, Rui Amaral Lopes, M.TH. Beheshti, Sadegh Bolouki. Residential energy flexibility characterization using non-intrusive load monitoring[J], Sustainable Cities and Society, Volume 75, 2021, 103321.
- [51] Alan Meier, Dan Cautley. Practical limits to the use of non-intrusive load monitoring in commercial buildings[J], Energy and Buildings, Volume 251, 2021, 111308.
- [52] 王延凯.python对语音文件加入高斯白噪声[EB/OL],(2020.07.20)[2020.07.20].
https://blog.csdn.net/weixin_38468077/article/details/107456886
- [53] 孟欣. 建筑能源监测系统中缺失数据的重构与修补方法研究[D].大连理工大学,2021.
- [54] Yangyang Fu, Zhengwei Li, Fan Feng & Peng Xu. Data-quality detection and recovery for building energy management and control systems: Case study on submetering[J], Science and Technology for the Built Environment, 22:6, 2016, 798-809.
- [55] Tong Xiao, Peng Xu, Ruikai He, Huajing Sha, Status quo and opportunities for building energy prediction in limited data Context—Overview from a competition, Applied Energy[J], Volume 305, 2022, 117829.
- [56] Ying Ji. FDD Method of HVAC System in Summer in Large Public Buildings with Submetering Electricity Data[D]. Tongji University, 2017.
- [57] Abraham Lee. Pyswarm-Particle swarm optimization (PSO) with constraint support[EB/OL], (2014.10.21)[2014.10.21]. <https://pythonhosted.org/pyswarm/>
- [58] Ryan Solgi. GENETICALGORITHM[EB/OL],(2020.12.30)[2020.12.30].
<https://github.com/rmsolgi/geneticalgorithm>
- [59] J.E. Braun, G.T. Diderrich. NEAR-OPTIMAL CONTROL OF COOLING TOWERS FOR CHILLED-WATER SYSTEMS[J]. ASHRAE TRANSACTIONS, 1990, Volume 96, Part 2, 806-813.
- [60] 深圳市住房和建设局,深圳市建设科技促进中心,深圳市建筑科学研究院股份有限公司. 深圳市大型公共建筑能耗监测情况报告(2020年度)[R].2021.2.
- [61] ISO 16346-2013, Energy performance of buildings – Assessment of overall energy performance[S]. Geneva: International Organization for Standardization: 2013
- [62] ISO 13790-2005, Thermal performance of buildings – Calculation of energy use for space heating and cooling[S]. Geneva: International Organization for Standardization: 2005.
- [63] Chung W. Review of building energy-use performance benchmarking methodologies[J]. Applied Energy, 2011, 88(5): 1470-1479
- [64] Sicheng Zhan, Zhaoru Liu, Adrian Chong, Da Yan, Building categorization revisited: A clustering-based approach to using smart meter data for building energy benchmarking[J], Applied Energy, Volume 269, 2020, 114920.
- [65] 住房和城乡建设部,建筑节能与可再生能源利用通用规范[S]. 2021.09.08. 北京.
- [66] J Z, D L. Current Situation of Energy Consumption and Energy Saving Analysis of Large Public Building[J], Procedia Engineering, Volume 121, 2015, Pages 1208-1214.

致谢

在同济的两年半既漫长又短暂，纵有千言万语时，但思绪万千无从提笔。

首先感谢各位老师对我的帮助和指导，在此特别感谢我的导师许鹏教授，对我的指导，是我成长路上的灯塔，给予我自由和包容能让我在科研的海洋探索，寻我所思，求我所想。

感谢同门陪我走过这段时光，相伴学习，嬉笑打闹，恍如昨日；感恩我在这段时光里遇到的每个人，教我成长，教我感恩。

若重逢若回首，长亭古道，黄云白日，祝好祝好。

2022年3月

个人简历、在读期间发表的学术成果

个人简历:

何睿凯, 男, 1997年3月生。

2019年6月毕业于中国人民解放军陆军工程大学建筑环境与能源应用工程专业, 获学士学位。

2019年9月进入同济大学机械与能源工程学院供热、供燃气、通风及空调工程专业, 攻读硕士学位。

已发表论文:

- [1] **Ruikai He**, Peng Xu, Zhibo Chen, Wei Luo, Zhineng Su, Jiong Mao. A non-intrusive approach for fault detection and diagnosis of water distribution systems based on image sensors, audio sensors and an inspection robot, *Energy and Buildings*, Volume 243, 2021, 110967.
- [2] Tong Xiao, Peng Xu, **Ruikai He**, Huajing Sha, Status quo and opportunities for building energy prediction in limited data Context—Overview from a competition, *Applied Energy*, Volume 305, 2022, 117829.
- [3] Mingyue Guo, Peng Xu, Tong Xiao, **Ruikai He**, Mingkun Dai, Shelly L. Miller. Review and comparison of HVAC operation guidelines in different countries during the COVID-19 pandemic, *Building and Environment*, Volume 187, 2021, 107368.
- [4] 郭明月, 许鹏, 肖桐, **何睿凯**, 戴明坤. 应对新型冠状病毒国内外暖通相关指南对比. 暖通空调, 2020, 50(11):13-20.

发明专利:

- [1] **何睿凯**、许鹏、顾洁帆. 一种建筑自动划分防烟分区的方法 (申请公布号: CN202110332196.4, 专利申请号: CN202110332196.4)
- [2] 许鹏、**何睿凯**、陈喆、陈永保、肖桐、陈智博. 一种空调系统中水系统路径寻优算法 (申请公布号: CN112241564A, 专利申请号: CN202011126801.4)

在校期间获得奖励:

2021.11 研究生国家奖学金

2021.11 同济大学优秀学生

同济大学学位论文原创性声明

本人郑重声明：所提交的学位论文《基于数据挖掘的公共建筑节能潜力分析方法》，是本人在导师指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人创作的、已公开发表或者没有公开发表的作品的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。本学位论文原创性声明的法律责任由本人承担。

学位论文作者签名：

日期： 年 月 日

同济大学学位论文版权使用授权书

本人完全了解同济大学关于收集、保存、使用学位论文的规定，同意如下各项内容：按照学校要求提交学位论文的印刷本和电子版本；学校有权保留学位论文的印刷本和电子版，并采用影印、缩印、扫描、数字化或其它手段保存论文；学校有权提供目录检索以及提供本学位论文全文或者部分的阅览服务；学校有权按有关规定向国家有关部门或者机构送交论文的复印件和电子版；允许论文被查阅和借阅。学校有权将本学位论文的全部或部分内 容授权编入有关数据库出版传播，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于（在以下方框内打“√”）：

保密，在_____年解密后适用本授权书。

不保密。

学位论文作者签名：

指导教师签名：

日期： 年 月 日

日期： 年 月 日