

# 机理计算与机器学习结合的空调负荷预测

李安琳, 龚延风

(南京工业大学, 江苏南京 210009)

**[摘要]** 文章提出了一种机理计算与神经网络学习相结合的在线负荷预测方法。利用 Python 建立了统计学预测模型、支持向量机预测模型、随机森林预测模型。并通过将溧阳某一建筑实际空调负荷数据逐步输入模型之中模拟建筑的实时运行, 对 1 年内和 1 年后的泛化能力进行分析。模拟结果表明, 随着实际负荷数据的逐步增加, 通过在线学习逐步完善性能, 可以快速建立建筑的空调负荷预测模型。机器学习预测空调负荷的方式扩大了负荷预测的使用范围, 提高了冷冻站智能化运行的水平。

**[关键词]** 机器学习; 支持向量机; 随机森林; 在线负荷预测

**[中图分类号]** TU831 **[文献标志码]** A **[文章编号]** 1005-6270(2020)04-0123-04

## Air Conditioning Load Forecasting Based on Mechanism Calculation and Machine Learning

LI An-lin GONG Yan-feng

(Nanjing Technology University, Nanjing Jiangsu 210009 China)

**Abstract:** In this paper, an on-line load forecasting method based on mechanism calculation and neural network learning is proposed. The statistical prediction model, support vector machine prediction model and random forest prediction model are established by python. And through gradually inputting the actual air conditioning load data of a building in Liyang into the model to simulate the real-time operation of the building, analyze the generalization ability within one year and one year later. The simulation results show that with the increase of actual load data and the improvement of performance through online learning, the air conditioning load forecasting model of buildings can be established quickly. The way of machine learning to predict air conditioning load expands the scope of load forecasting and improves the level of intelligent operation of refrigeration station.

**Key words:** machine learning; support vector machine; random forests; online load prediction

### 0 引言

空调负荷预测是冷水机组运行的基础条件, 是冷冻站控制策略制定的必要依据。目前, 冷水机组的选型都是按照最大制冷量选取的, 即按照最大建筑负荷进行选取。但是一般情况下, 冷水机组满负荷的运行时间不到总运行时间的 3%<sup>[1]</sup>, 而且据国外学者实测得到, 空调机组 80% 以上的时间是在 60% 以下的部分负荷情况下运行<sup>[2]</sup>, 所以冷水机组能耗主要是其在部分负荷运行工况下的能耗。如果冷水机组长时间保持高负荷运行, 而不随着室外气象参数扰动和室内状态变化及时进行调整, 会造成巨大的能源浪费。所以, 准确的空调负荷是冷水机组合理运行的必要条件。但是, 冷水机组的台数一般都是通过实测冷冻水供水温差、流量等方法进行控制<sup>[3]</sup>, 导致控制过程往往不能及时

对空调负荷的变化做出准确的响应。除此此外, 有时由于工作人员的人为操作的误差, 也会导致冷水机组开启台数的过多或过少, 一方面影响了室内的热湿环境, 造成室内人员的舒适度下降; 另一方面也造成了能源的大量浪费。鉴于此, 通过对影响空调负荷的主要因素进行分析, 从而对负荷做出准确合理的预测显得至关重要。

### 1 负荷估算模型

#### 1.1 负荷估算方法

基于空调负荷机理的估算模型如式(1):

$$Q=Q_1+Q_2+Q_3 \quad (1)$$

式中,  $Q_1$ : 通过围护结构得热量形成的冷负荷, 可采用冷负

基金项目: 江苏省大学生创新创业训练计划项目 (项目编号 201910291119Y)

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

[收稿日期] 2020-07-03

[作者简介] 李安琳, 南京工业大学, 本科生。

荷系数法求得,分为通过墙体、屋顶和窗户瞬变传热形成的冷负荷和通过窗户日射得热形成的冷负荷,  $W$ ;

$Q_2$ : 通过室内热湿源散热散湿形成的冷负荷, 包括设备、照明和人体的得热,  $W$ ;

$Q_3$ : 新风冷负荷,  $W$ 。

$Q_1$  的计算难点在于通过墙体、屋顶和窗户的传热冷负荷计算。目前常用的负荷计算方法主要分为 2 类。①基于设计条件的动态负荷计算方法,如谐波反应法、反应系数法、Z 传递函数法等。②针对季节或年度负荷的计算方法,如度日法、当量满负荷运行时间法、负荷频率法等,不能用来进行逐时负荷预测。空调负荷的估算面对实际不断变化的室外气象条件,本文提出基于动态负荷修正模型。

通过窗户日射得热形成的冷负荷,以及通过室内热湿源散热散湿形成的冷负荷和新风冷负荷的计算按照实际测得的数据进行计算即可,计算模型不需修正。

## 2 用于预测的机器学习模型

机器学习方法构建模型从数据的角度在处理复杂非线性大量环境数据集之间的关系时有着很大的优越性,它们不需要解决许多统计模型所需的假设:样本的正态分布性和齐方差性、多重共线性、变量之间独立等严格的参数的假设。最常见的机器学习模型包含了人工神经网络(Artificial Neural Networks),支持向量机(Support Vector Machines)等等。

### 2.1 随机森林回归模型

#### 2.1.1 决策树与随机森林概述

关于决策树的早期著作之一是 Breiman 等人的开创性著作《分类与回归树(Classification and Regression Trees, CART)》,其中描述了决策树的基础知识以及它们在分类和回归问题中的应用。为此,最流行的算法之一是 Quinlan 的“C4.5”。虽然决策树被证明是有用的,但它们的应用仍然局限于相对低维的数据。随机森林可以高效应对很多变量,并可以在模型端评估变量的重要性,产出高准确度的回归器。而且随机森林可以有效避免过度拟合的问题,学习计算效率高。

#### 2.1.2 决策树分类器与回归器

在决策分析中,决策树可用于可视化和显式地表示决策和决策制定。虽然它是数据挖掘中常用的一种工具,用于派生一种策略以达到特定的目标,但它也广泛用于机器学习。一棵树的训练过程为:根据一个指标,分裂训练集为几个子集。这个过程不断地在产生的子集里重复递归进行,即递归分割。当一个训练子集的分类都相同时,递归停止<sup>[4]</sup>。类似于支持向量机模型,决策树模型同样需要通过其根据一对剩下全部类别(1 vs all)原理来实现多元分类器,通过连续性目标函数实现回归器。

#### 2.1.3 随机森林

在实现随机森林模型时通过构建多个随机的决策树,并通过投票(加权平均等)方式进行对结果的预测。在随机选取子空间的策略上,亦有通过随机选取互补的数据子空

间训练决策树模型,并整合成综合性能提升的随机森林模型(图 1)。

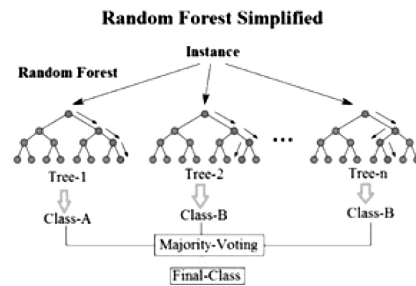


图 1 决策树的原理图

## 2.2 人工神经网络

人工神经网络(Artificial Neural Networks, ANN)是一种复杂的非线性建模技术,能够对复杂函数进行建模。它们可以应用于各种领域的预测、分类或控制问题,是建筑能量预测中应用最广泛的人工智能模型。在过去的 20 年里,研究人员已经应用神经网络来分析各种条件下的各类建筑能耗,如加热/冷却负荷、电力消耗、子级组件运行和使用参数的优化、估计。

人工神经网络由一组模拟神经元组成。每个神经元都是一个节点,通过与生物轴突-突触-树突连接的链接与其他节点相连。每个链接都有一个权重,它决定了一个节点对另一个节点的影响强度。

多层感知器(Multi-Layer Perceptron, MLP)是一种监督学习的算法,是前馈式神经网络的一种,也是结构最为简单的一类神经网络。这个概念常常也和人工神经网络狭义含义相近,在一些研究中,这两者的名字通常会被交替使用(图 2)。

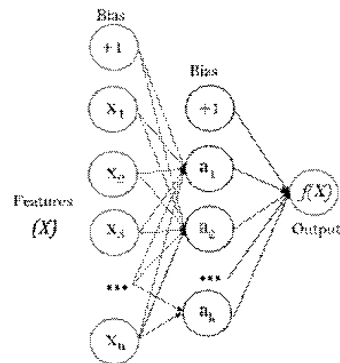


图 2 单隐藏层多层感知器的结构示意图

其中第一列表示  $n$  维数据向量的元素,通过加权求和,并辅以适当的非线性激活函数,获得隐藏层的一个神经元。以图示所示的多层感知器为例,隐藏层中的每一个元素可以表示为:

$$\alpha_i = f_{\text{activate}}(W_0 + \sum_{m=1}^n w_m x_m) \quad (2)$$

其中  $W_m$  表示为每个输入层元素的权重,  $f_{\text{activate}}$  表示非线性的激活函数。激活函数为神经网络引入非线性,常见的激活函数包含 tanh、ReLU、sigmoid 等。通常,多层感知器有 0 到多个隐藏层,而最后的隐藏层则输出为目标。输出目标通

常有多种转化,例如 Softmax 等。

### 2.3 机器学习中的其他问题

#### 2.3.1 交叉验证

交叉验证是将评估统计分析的结果推广到一个独立数据集的模型验证技术的总称。在预测问题中,通常需要两个数据集运行,包括一个确定的训练数据集和一个未知的测试数据集。交叉验证的目的是测试模型预测未用于估计的新数据的能力,以便标记诸如过度拟合或选择偏差之类的问题,并了解模型将如何推广到独立数据集(即来自真实问题的未知数据集)<sup>[5]</sup>。

$k$ -fold 交叉验证是将样本随机分成  $k$  个相等大小的子集中。在  $k$  个子样本中,一个样本被用于验证集,其他  $k-1$  个子样本被用作训练集。模型将被训练  $k$  次,每个子样本都会正好作为验证集一次。通过综合几次的结果,交叉验证可以生成一个最好的模型。这种模型的好处就在于所有的观测值都会被用于作为训练集和测试集。一般常使用 10-fold 交叉验证,但这和能提供的训练样本的规模有关。

在本项目中,对机器学习的模型训练同样需要采取交叉验证的手段去获得最好的回归预测器。

#### 2.3.2 标准化

将数据投入到除了统计模型以外的机器学习模型中,需要对数据进行初始化。进行标准化的主要作用在于防止一些值域较大的变量对模型结果施加了远远大于其他变量的影响。另外标准化还有简化运算的功能。常见的标准化包括正态分布标准化、归一化等<sup>[6]</sup>。

#### 2.3.3 参数训练

超参数是在估计器中不能直接学习的参数。例如在神经元中应当采用哪种激活函数,或者是采用哪一种内核学习器,以及学习率等。scikit-learn 提供了 2 种抽样搜索候选对象的通用方法:(1)对于给定的超参数的范围,Grid-SearchCV 会详细考虑所有的参数组合<sup>[7]</sup>。(2)Randomized-SearchCV 可以从具有指定分布的参数空间中采样给定数目的候选对象。

### 3 案例分析

#### 3.1 数据来源与变量选取

为了模拟小样本条件下的空调负荷预测,采用某一建筑运行数据创建训练所需数据集。建筑地点位于溧阳,建筑高度为 15 m,建筑面积为 4 000 m<sup>2</sup>,建筑类型为酒店,样本日期选取为 2018 年 8 月 1 日—8 月 31 日,时间为每天的 8:00—20:00(周六、周日除外),此为间歇运行模式下的其余时刻负荷的预测。

总体上,建筑空调负荷的影响因素主要包括外扰因素和内扰因素。外扰因素包含室外气象参数,其中室外干球温度和太阳辐射的影响较大;内扰因素包含室内人员、照明和设备的数量、作息和工作形式,由于室内人员情况比较不固定,无法进行定量的说明,需要对其进行模糊化处理。照明和设备相比于人员的影响较小,也相对比较固定,可以不做

考虑。所以,输入层输入参数一般选取  $k-1$  时室外干球温度(即前 1 h)、 $k$  时室外干球温度及含湿度、 $k-1$  时的太阳辐射、 $k$  时的太阳辐射。

#### 3.2 数据预处理

首先,可以观察本实验中的一些数据。针对被解释变量和解释变量进行散点图分析,可以获得图 3 所示结果。

从图中可以看出,上述的变量基本满足正态分布,可以

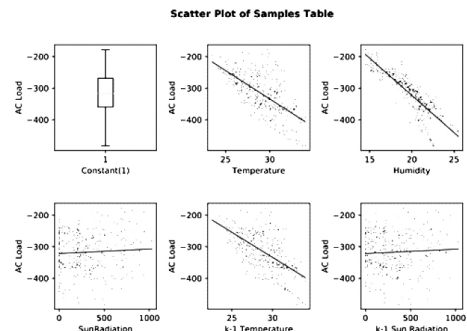


图 3 数据预处理的结果示意图

采取正态分布标准化方法将数据进行标准化处理。

为了建立合理准确的回归模型,将数据分割为训练集和测试集两部分。训练集和测试集都是数据集的子集,前者用于训练模型,后者则用于在模型经训练后测试和评估模型。在训练机器学习的过程中,常常在训练集中提取随机的子集作为验证集,参与到交叉验证的运算中。在本实验中,将模型以 8 月 21 日 0:00 为界限,划分为训练集和测试集。因此,训练集包含 182 条数据,而测试集包含 117 条数据。对于后续的机器学习模型,对解释变量采取了标准化处理,以加速模型的计算速度。

#### 3.3 机器学习模型

使用 Sci-Kit Learn 搭建 3 套机器学习模型。模型训练需要消耗较大的计算资源,其计算复杂度较高,所以通过该软件包提供的并行计算模块来加速模型的运算。通过对参数进行训练和选择,可以得到 3 个预测模型。

##### 3.3.1 支持向量机预测

针对支持向量机回归器,设定了 2 个计算核函数和 3 个正规化权重。计算核的加入使得回归器的评分机制获得了非线性,而正规化参数与损失函数中的正规化部分的强度成反比。经过五折交叉验证和网络搜索超参数,一组最好的超参数如下:机器学习目标函数中的正规化比例为 1,回归器接收范围为  $\epsilon=0.1$  (支持向量机的回归结果在 0.1 的接受范围内不会在目标函数中添加权重),核函数为 rbf (高斯径向基函数核),其参数  $\gamma=0.001$ 。该模型训练后对训练集和测试集的拟合优度分别为 63.7%和 90.2%。

超参数与其对应的五折交叉验证过程中的平均评分的关系如表 1、图 4。

##### 3.3.2 随机森林预测

随机森林回归器的超参数的选择较少,对随机森林里的决策树数量以及深度进行网格五折交叉验证搜索,寻得



表1 机器学习模型的超参数与模型评分表

核函数	正规化系数			
	1	10	100	1000
线性	0.544 806 54	0.506 963 51	0.503 376 67	0.498 985 8
高斯径向基函数核	-121500207	0.115 113 27	0.249 679 48	-0.072 759 6

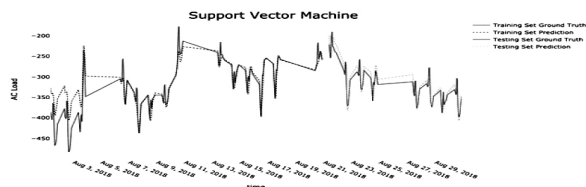


图4 支持向量机预测模型的结果图

的最佳分类器拥有 200 个随机决策树, 其决策树分裂的最大深度为 100 分叉, 决策树叶节点至少需要有一个数据目标值。该框架下, 最佳的随机森林回归器对训练集的拟合优度高达 97.2%, 而对测试集的拟合优度仅 80.7%。

超参数与其对应的五折交叉验证过程中的平均评分的关系如表 2、图 5。

表2 决策树模型的超参数与模型评分表

决策树深度	数量	
	100	200
5	0.024 369 19	0.022 546 33
10	0.021 684 88	0.025 981 54
100	0.043 241 48	0.034 417 84

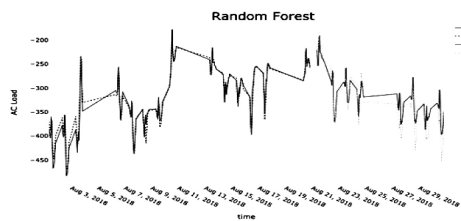


图5 随机森林预测模型的结果图

### 3.3.3 神经网络预测

对于多层感知器, 设定了多种网络结构和求解方法, 具体的参数如图 6 所示。激活函数代表神经元的处理数据信号的转化, 增加多层感知器的非线性和模型的复杂度, 包括罗基斯基回归、双曲正切和线性整流函数。而求解器则采用了多套不同的机器学习框架优化算法, 包含 LBFGS、随机梯度下降(SGD)、随机优化器(ADAM)等。在无法通过直接分析求解的机器学习的训练中, 学习率也是一个重要的参数, 其代表优化函数的过程中每次对参数的变动的尺度。过大的学习率会导致在训练过程中发生振荡而不收敛, 而过小的学习率则会导致陷入局部最优解而非全局最优。

经训练, 本项目中的神经网络使用 LBFGS 优化器, 结构简单的两层隐藏层, 线性整流函数和最大的学习率。该模型的对训练集和测试集的性能较为平衡, 其拟合优度分别为 85.5%与 81.7%。

### 3.3.4 预测方法的比较

从上述的模型来看, 各种回归器对预测空调负荷的能力较为接近, 预测误差总体上在 5%左右, 且均能真实反映

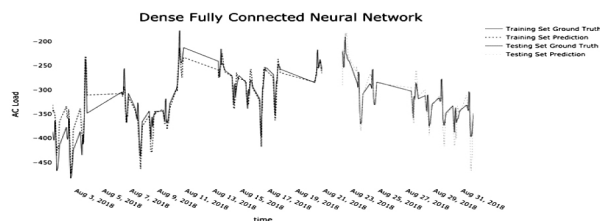


图6 神经网络预测模型的结果示意图

原有的针对空调负荷与解释变量的关联性。从拟合优度的角度来看, 随机森林模型更好地捕捉了训练集的特征, 而多层感知器模型获得了较为均衡的表现。在计算复杂程度上来看, 随机森林模型需要消耗最大的计算资源, 因为其需要建立非常多的随机决策树; 而多层感知器模型也需要较大的运行时间, 因为其拟合求解过程中的复杂度较高。综上所述, 本实验中采取的机器学习模型整体的性能有待提高, 但依然能够取得较为准确的预测结果。

## 4 结论与展望

(1)采用基于机理的经验模型与机器学习相结合的负荷预测方法, 可以实现在小样本条件下建筑空调负荷的实时在线负荷预测。依据预测负荷可对冷水机组进行台数控制, 为冷水机组智能控制打下基础, 便于无人值守机房的实现。

(2)机器学习模型的学习时间冬/夏季分别需要 20 d 左右, 同一年进行中期负荷预测的误差较小, 预测误差总体上在 5%左右, 基本上满足控制精度要求, 模型泛化性能较好。实际工程中, 随着学习样本的不断增长, 本文提出的模型将随时间逐步成长, 预测的精度与泛化能力将逐步提高, 直至完全满足工程要求。

(3)将机理估算模型、机器学习预测模型相结合, 应用于空调负荷的预测, 并根据冷冻站运行模式的不同对神经网络的输入参数进行差异性选取, 同时对室内人员参数进行模糊化处理, 提高了预测的通用性。

## 参考文献

- [1]施丹, 许必熙. 空调系统负荷分段预测与研究[J]. 建筑热能通风空调, 2018, 37(04): 20-24.
- [2]周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 101-106.
- [3]Hecht-Nielsen R. Theory of the backpropagation neural network[J]. Neural Networks for Perception, 1988, 1(1): 445-445.
- [4]J. R. Quinlan, in Induction of Decision Trees. Machine Learning 1, Kluwer Academic Publishers, 1986, pp. 81-106.
- [5]W. Sarle, "Neural Network FAQ," Periodic posting to the Usenet news group comp.ai.neural-nets, 1997.
- [6]F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau and Brucher, "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 1825-2830, 2011.
- [7]S. -K. L. Documentary, "sklearn.model\_selection.GridSearchCV," [Online]. Available: [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html#sklearn-model-selection-gridsearchcv](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html#sklearn-model-selection-gridsearchcv).