

[负荷预测]

基于MIC-LightGBM的综合能源系统多元负荷预测

文/张河宜¹ 周 毅¹ 张 军¹ 陈俊杰¹ 罗俊杰²

1. 国网神农架供电公司; 2. 国网江陵县供电公司

提出了基于最大互信息系数 (Maximal Information Coefficient, MIC) 结合轻梯度提升机 (Light Gradient Boosting Machine, LightGBM) 的综合能源系统多元负荷预测模型。首先利用MIC进行相关因素分析, 筛选特征降低预测模型冗余度, 最后输入至LightGBM模型获得多元负荷预测结果。经算例验证, 基于基学习器集成的LightGBM具有最低预测误差和较快的预测速度。

综合能源系统 (Integrated Energy System, IES) 是新型的能源系统结构, 能够将系统内部的多异质类型能源通过能量交换装置紧密结合, 为用户提供供电、供热和供冷等服务, 在提高能源效率、减少碳排放和增加可再生能源的渗透等方面起到重要作用。但IES具有多能互补、系统内部多元负荷相互耦合的特点, 增加了系统内部能源的灵活性, 同时也使得系统更具复杂性。为了确保IES优化调度的可行性和可靠性, 对IES多元负荷预测准确性和有效性提出更高要求。

近些年来, 以机器学习为主的人工智能方法被广泛应用在单一能源系统和IES负荷预测领域。由于单一的机器学习算法自身机理单一, 使得预测泛化性能弱, 在IES多元负荷预测场景下可能存在预测结果精度不足、预测耗时过长等问题。针对上述问题, 充分发挥机器学习优势, 本文提出基于集成学习的IES多元负荷预测方法。集成学习可以博采众长, 结合不同预测方法的优势, 在利用MIC进行特征筛选的基础上, 采用基于集成学习的LightGBM算法, 以决策树为基学习器进行集成, 构建MIC-LightGBM的IES多元负荷预测模型, 经算例验证分析, 所提预测模型可以更少的预测耗时

获得较为准确的多元负荷预测结果。

集成学习

集成学习由多个基学习器通过一定结合策略形成。根据基学习器的生成方式, 集成学习分为两类: 基学习器依次生成且存在强依赖关系的为串行集成; 基学习器之间不存在强依赖关系的为并行集成。串行集成的基学习器一般是同质的, 典型算法是以决策树为基础的Boosting系列算法, 并行集成典型的是Bagging系列算法。Bagging算法属于集成学习的并行集成范式, 将多个基学习器结合形成强学习器。Bagging算法通过自主采样对原始训练集任意有放回地抽取形成若干训练数据集, 基于这些不同训练数据集训练若干基学习器, 再以某种策略结合这些基学习器获得输出。

LightGBM预测模型

为提高大规模数据处理的计算耗时和精度, Microsoft的研究人员于2017年推出了LightGBM算法, LightGBM高效率并行的模型训练优势, 使得其具有更快速的训练速度、更低的内存占用、更准确的预测结果等优秀的数据处理能力。该算法一

经提出，就在Kaggle等国际数据处理大赛上广泛应用，并取得了优秀的成绩。LightGBM算法是在XGBoost算法基础上优化训练样本输入，改进叶结点分裂策略，以提升预测精度。

$$f_T(X) = \sum_i^T a_i f_i(X), f_i \in F \quad (1)$$

式中， F 代表所有决策树的集合空间。LightGBM算法的基学习器为决策树，各基学习器串行生成，并按照式（1）的形式将 T 棵决策树结合起来，形成新的决策树目标是使包括回归误差和复杂度在内的损失函数最小。MIC-LightGBM综合能源系统多元负荷预测流程如图1所示。

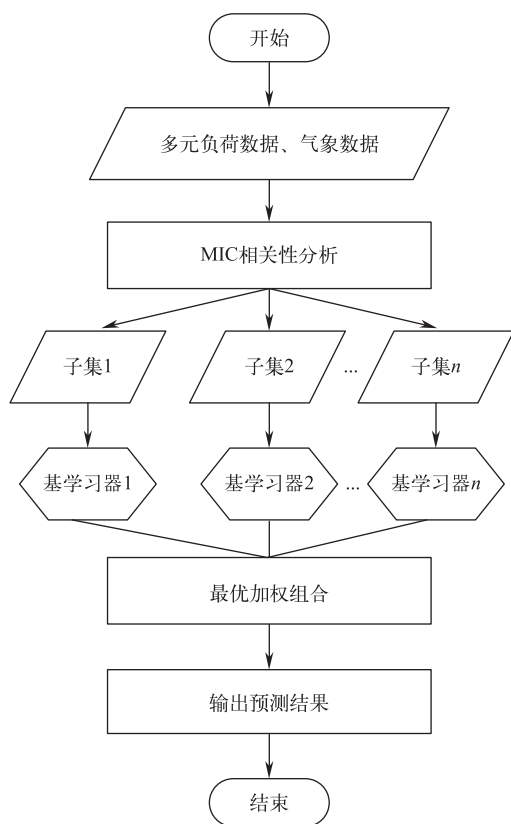


图1 MIC-LightGBM综合能源系统多元负荷预测流程

最大互信息系数

MIC是一种基于互信息系数的相关性分析方法，MIC相比于Pearson和Spearman更具普适性，能够同时有效地衡量两个变量之间的线性或非线性的关联强度，在保持强相关性变量的属性信息的同时，弥补弱相关性变量属性间非线性关系。MIC计算公式如下。

互信息 (Mutual Inforamtion)

$$I[X;Y] = \sum_{X,Y} p(X,Y) \log_2 \frac{p(X,Y)}{p(X)p(Y)} \quad (2)$$

最大互信息系数

$$M[X;Y] = \max_{|X||Y| \leq B} \frac{I[X;Y]}{\log_2(\min(|X|, |Y|))} \quad (3)$$

式中， $|X||Y|$ 为在 X 、 Y 方向上划分大的网格数； B 为网格限制变量， $B=n^\theta$ ($\theta=0.55$ 或 0.6)。

算例分析

1. 算例介绍

为说明本文提出的基于MIC-LightGBM的综合能源系统短期负荷预测模型的有效性和优越性，采用区域校园综合能源系统算例进行验证，该数据集时间尺度为2020年4月9日—6月9日，数据粒度为1 h，每日采集24个数据点，数据集内容包含校园中日常、教学活动所需的电、冷、热负荷。且气象数据集时间尺度以及粒度与负荷数据集相适配，包含温度、压强、相对湿度和风速等。

2. 评价指标

为体现预测结果的精确性，选择均方根误差 (RMSE) 和平均绝对百分比误差 (MAPE) 作为评估指标。

RMSE表示预测值与真实值的均方根误差，以标准差的形式衡量预测结果中极端异常值对整体预测性能的影响。

$$E_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (4)$$

MAPE表示预测值与真实值的平均绝对百分比误差，以百分比的形式衡量预测模型在不同负荷量纲下的预测效果。

$$E_{\text{RMSE}} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (5)$$

式中， $\hat{y}_i$ 、 y_i 为第 i 个时刻的预测值和真实值； n 为数据总体个数。

3. 多元负荷特征选取

IES区别于单一能源系统的最大特点在于：IES包含了多种不同的能源形式之间的互补互济，所以有必要对其进行定量的相关性分析，更好地利用多元负荷的相关性因素进行模型训练，获得准确、可

靠且符合实际工程背景的多元预测结果。依据MIC计算结果绘制热力图，如图2所示。将系数不小于0.5视为具有较强相关性的特征。根据MIC系数选取温度和电冷热负荷历史数据构建特征矩阵。

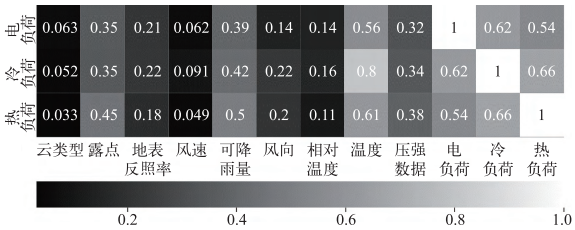


图2 MIC相关性系数热力图

4. 对比实验

为了验证MIC-LightGBM在预测时间和预测精度两个方面的优越性，选择基于统计学习理论和采用超平面分割的SVR及基于深度学习可利用时序特征的LSTM进行对比说明。

5. 参数设置

大量研究说明机器学习算法的有效性和可靠性对自身的参数设置具有高度的敏感性。故本节着重说明LightGBM算法中关键参数及关键参数如何确定。

LightGBM算法中涉及的核心参数：学习率影响模型训练过程中收敛速度；最大深度的大小控制训练模型过程中特征利用能力；最大叶子结点数表

示随着结点数目增大，可以提高精度，但也会发生过拟合现象；正则化参数防止预测模型过拟合。具体参数的选取根据验证集的预测效果，采用试错法进行确认。具体参数为：学习率为0.05、最大深度为5、最大叶子结点数为300、正则化参数为0.5。

6. 结果分析

结合MIC特征筛选的SVR、LSTM和目标模型LightGBM的电、冷、热负荷预测任务的误差结果见表1。由表1分析可知，本文所提出的基于MIC-LightGBM在电负荷、冷负荷和热负荷的预测任务上均具有最优的预测精度。以MAPE为例，MIC-LightGBM较MIC-LSTM电、冷、热预测误差分别下降了35%、40%和31%；较MIC-SVR分别下降了48%、51%和47%，验证了LightGBM基于集成思想，利用式（1）将基学习器组合集成，可以有效降低预测误差。对比各预测耗时可知，基于统计学习理论的SVR的预测耗时最短，基于深度学习的LSTM预测耗时最长，本文提出的LightGBM耗时介于两者之间，但相较于这两个预测模型，LightGBM具有最佳的预测准确性，说明了LightGBM并行计算在不失预测精度的情况下，还能有效降低预测耗时。

表1 各预测模型的RMSE、MAPE 误差指标对比

模型	RMSE			MAPE			预测耗时 /s
	电 /kW	冷 /Ton	热 /mmBTU	电 /kW	冷 /Ton	热 /mmBTU	
MIC-SVR	465.38	527.22	0.24	1.64	5.77	3.19	168
MIC-LSTM	329.42	464.28	0.19	1.32	4.71	2.48	402
MIC-LightGBM	202.54	230.02	0.13	0.85	2.81	1.69	246

结束语

本文提出了一种基于MIC-LightGBM的综合能源系统多元负荷预测方法，利用MIC分析多元负荷相关性因素，筛选出与预测任务具有较强相关性的特征并构建输入矩阵，可以有效降低模型冗余度，减少训练时长；LightGBM高效的并行计算能力，在预测精度和预测耗时上明显优于其他对比模型，验证了LightGBM应用于综合能源系统多元负荷预测问题上的优越性。EA

(收稿日期：2023-04-04)

参考文献

[1] 朱继忠, 董瀚江, 李盛林, 等. 数据驱动的综合能源系统负荷预测综述 [J]. 中国电机工程学报. 2021: 1-20.

[2] 李文武, 张鹏宇, 石强, 等. 基于聚合混合模态分解和时序卷积神经网络的综合能源系统负荷修正预测 [J]. 电网技术, 2022, 46(9): 3345-3357.

[3] 屈小云, 吴鸣, 李奇, 等. 多能互补综合能源系统综合评价研究进展综述 [J]. 中国电力, 2021: 1-13.