

基于机器学习的混合模型在电力负荷预测中的应用

徐建平¹, 胡济恒¹, 孙卫哲²

(1. 国网浙江省电力有限公司金华供电公司, 浙江 金华 321000; 2. 北京众智新能科技有限公司, 北京 100080)

摘要:从电力负荷的变化特点入手,提出基于机器学习的混合模型,该模型采用具有自适应噪声的完整集合经验模式分解技术对电力负荷进行分解,剔除高频噪声影响后,对负荷序列进行重构,采用多目标花授粉优化算法对极限学习机进行优化,分析认为该模型可提高极限学习机预测的准确性和稳定性,结合江苏省月度负荷预测案例验证了模型的有效性。

关键词:电力系统;负荷预测;机器学习;分解技术;多目标优化;花授粉优化算法

中图分类号:TM715 **文献标志码:**B **文章编号:**1001-9898(2021)01-0027-04

Application of Hybrid Model Based on Machine Learning in Power Load Forecasting

XU Jianping, HU Jiheng, SUN Weizhe

(1. State Grid Zhejiang Electric Power Co., Ltd. Company Jinhua Power Supply Company, Jinhua 321000, China;

2. Beijing Zhongzhi New Energy Technology Co., Ltd., Beijing 100080, China)

Abstract: Starting from the changing characteristics of the power load, a hybrid model based on machine learning is proposed, and the model is analyzed using a complete set of empirical mode decomposition technology with adaptive noise to decompose the power load. After removing the influence of high-frequency noise, the load sequence is reconstructed. The process of optimizing the extreme learning machine with a multi-objective flower pollination optimization algorithm, and that the model can improve the accuracy and stability of the extreme learning machine prediction, combined with the monthly load forecasting case analysis in Jiangsu Province to verify the effectiveness of the model.

Key words: power load system; load forecasting; machine learning; decomposition technology; multi-objective optimization; flower pollination optimization algorithm

1 概述

电力负荷不仅受到气象因素例如气候、温度、湿度的影响,也会受到消费者居住模式和生活方式的影响。气象因素的变化莫测加上消费者的随机和非线性行为均为电力负荷的预测工作带来巨大的阻碍。在文献中,电力负荷的预测方法主要分为两类:经典的统计模型,具有预测效率高、模型构建容易的优点,但是也存在处理非线性时间序列能力较差、对数据质量要求高以及需要基于一定假设条件的缺点;机器学习模型,具有预测精度高,善于处理非线性时间序列的优点,但是其在处理非线性时间序列时也具有学习能力有限、容易陷入局部最优和过度拟合、灵活参数过多等问题。

鉴于以上问题,混合预测模型逐渐发展起来,

郭威麟等人为了降低负荷序列的非线性,提出了基于集成经验模态分解法(EEMD),BP神经网络和遗传算法的短期负荷预测法,实验结果表明该模型预测精度更高,稳定性更强^[1]。龙金莲采用粒子群算法优化最小二乘支持向量机对短期电力负荷进行了预测,并得到了满意的实验结果^[2]。李冬辉等人提出了一种将多种群的果蝇优化算法(MFOA)和广义回归神经网络(GRNN)相结合的混合年电力负荷预测模型,实验证明提出的模型预测性能优于其它对比模型^[3]。

本文结合了解析技术与优化算法的优势,提出了基于具有自适应噪声的完整集合经验模式分解(CEEMDAN)、多目标花授粉算法(MOFPA)和ELM的混合模型。首先,利用CEEMDAN将原始序列分解成若干IMF,然后剔除其中的高频噪声分

收稿日期:2020-10-13

作者简介:徐建平(1981—),男,高级工程师,主要从事电网运行管理工作。

量,重构负荷序列。最后以预测精度和预测稳定性为2个优化目标,用MOFPA优化ELM,对江苏省月度电力负荷进行预测。

2 基于机器学习的混合预测模型设计

2.1 具有自适应噪声的完整集合经验模式分解(CEEMDAN)

经验模态分解(EMD)是Huang等人^[4]提出的一种自适应信号时频处理方法,该方法根据数据本身在时间尺度上的特征对信号进行分解,而无需预先设置任何基函数。它将序列分解为有限数量的本征模态函数(IMF)。每个IMF分量在不同时间尺度上代表原始信号的不同特征。

EMD在处理非平稳和非线性信号方面具有很大的优势,但是仍然存在“模态混叠”问题。通过将高斯白噪声添加到信号中,集成经验模式分解(EEMD)算法,在很大程度上消除了EMD算法中的模态混叠问题。但是,EEMD算法无法在信号重构后完全消除高斯白噪声,从而导致重构误差。CEEMDAN作为EEMD的改进版本,更有效地消除了模态混叠,重构误差几乎为零,大大降低了计算成本^[4]。

定义通过EMD获得的第 j 个模态分量的算子 $E_j(\cdot)$,将 $w_i(t)$ 设为具有正态分布 $N(0,1)$ 的白噪声。CEEMDAN算法的步骤如下。

a. 通过EMD分解每一个序列 $s_i(t) = s(t) + \epsilon_0 w_i(t)$ 来提取第一个IMF分量,其中 ϵ_0 是噪声系数, $i = 1, 2, \dots, I$,定义第一个模态为 $\overline{IMF}_1 = \frac{1}{I} \sum_{i=1}^I IMF_{i1}$ 。

b. 计算第一个残差 $r_1(t) = s(t) - \overline{IMF}_1$ 。

c. 分解残差 $r_1(t) + \epsilon_1 E_1(w_i(t))$ 来获得第2个模态 $\overline{IMF}_2 = \frac{1}{I} \sum_{i=1}^I E_1(r_1(t) + \epsilon_1 E_1(w_i(t)))$ 。

d. 重复分解其它IMF直到获得残差。最终的残差可以表示为:

$$R_M(t) = s(t) - \sum_{j=1}^M \overline{IMF}_j \quad (1)$$

式中: M 表示IMF的总数量,IMF在不同的时间尺度上共同构成原始信号的特征。

2.2 多目标花授粉优化算法(MOFPA)

花授粉算法是一种元启发式算法,其灵感来自花朵的自然授粉。花授粉算法使用一组规则来简

化自然花授粉的复杂过程。在花授粉算法中,假定花粉 i 的第 t 次迭代的解为 x_i^t ,当前全局最优解为 x^* ,异花授粉的迭代过程可以表示如下:

$$x_i^{t+1} = x_i^t + \gamma L(\lambda) (x^* - x_i^t) \quad (2)$$

式中: γ 为一个常数,表示步长尺度因子; $L(\lambda)$ 为Levy飞行的步长。自花授粉过程可以表述如下:

$$x_i^{t+1} = x_i^t + \epsilon (x_j^t - x_i^t) \quad (3)$$

式中: ϵ 为一个随机数,且 $\epsilon: U(0,1)$, x_j^t 和 x_k^t 是与 x_i^t 不同的2朵花。

与解决单目标问题不同,多目标优化问题无法通过算术关系运算符比较不同解决方案的优劣。为此,通常定义包括帕累托占优、帕累托最优性、帕累托最优集和帕累托最优前沿来解决多目标优化问题^[5]。为了提高预测精度和稳定性,本文定义了2个目标函数如下:

$$\min \begin{cases} obf_1(y) = MAPE = \frac{1}{N} \sum_{i=1}^N abs((\hat{y}_i - y_i)/y_i) \\ obf_2(y) = std(\hat{y}_i - y_i), i = 1, 2, \dots, n \end{cases} \quad (4)$$

式中: N 为观测值的数量; $\hat{y}_i$ 和 y_i 分别为预测值和观测值。基于上述目标函数,进行迭代寻优,得到的最优解为帕累托最优解。

2.3 ELM

极限学习机由Huang等人开发,是一种紧凑高效的单层前馈神经网络学习算法,其学习效率和泛化能力均高于传统的前馈网络^[6]。它具有结构简单、预测准确性高、预测能力和实时学习能力强、计算速度快和训练样本需求少的特点。假定 X 是一个单标签样本, Y 表示分类,其中 $X \in R^d$ 且 $Y \in R^C$ 。极限学习机的输入层,隐藏层和输出层分别有 d, h 和 C 个节点。该网络的独特之处在于,输入层的权重和偏差会随机初始化,并且不会进一步更新。模型过程可表示为:

$$Y = Z\beta$$

式中: $Y = \{y_1, y_2, \dots, y_C\}$, $Z = \{z_1, z_2, \dots, z_h\}$ 表示隐藏层输出; β 表示隐藏层和输出层之间维数为 $h \times C$ 的权重矩阵。进而,权重矩阵 β 可以表示如下:

$$\beta = Z^+ Y$$

式中: Z^+ 表示 Z 的摩尔-彭罗斯逆。一旦获得了权重矩阵 β ,极限学习机方法就已经完成了训练阶段,并准备进入测试阶段。

2.4 混合预测模型的建立

机器学习模型通过对观测值进行训练学习,来拟合序列的复杂特征,进而提供较为精确的预测结果。因此,建立机器学习模型时往往需要基于一定量的历史观测值,具体观测值的数量通常依据多次实验经验确定。现有研究证明混合预测模型在电力负荷预测方面具有良好的预测性能,基于文献[7]的研究,本文提出了基于机器学习的混合预测模型,其在数据处理方面能更好地消除随机噪声的影响,且基于 ELM 的机器学习预测模型计算效率显著提高。提出的基于机器学习的混合预测模型框架主要可以分为 2 个阶段。

第 1 阶段:由于电力负荷数据固有的非线性和不稳定性特征增加了电力负荷预测的难度,为了减弱原始负荷数据中噪声的负面影响并进一步提高最终的预测精度,提出的混合预测框架首先对电力负荷数据进行数据预处理。利用 CEEMDAN 技术将原始序列分解为多个子序列,以捕获和挖掘负荷时间序列的主要特征。通过剔除高频成分来剔除噪声影响,得到重构的负荷序列,为提高模型整体的预测精度做出了重要贡献。

第 2 阶段:模型的预测准确性和稳定性对于电力系统的平稳运行和高效调度是必不可少的,因此本文将预测精度和预测稳定性被定义为多目标花授粉算法的 2 个优化目标,具体如式(4)。鉴于 ELM 出色的模拟与预测能力,本文利用优化的 ELM 通过滚动预测形式对重构负荷序列的未来变化趋势进行预测,预测过程中通过多目标优化确保最终预测结果的准确性和稳定性。

3 实例分析

3.1 数据选取

选取 2011 年 1 月至 2018 年 12 月江苏省月度工业用电量作为实验数据集来验证提出模型的良好性能。根据多次实验,将训练集与测试集的比例设为 7:3。江苏省工业用电量曲线见图 1,从图 1 可以看出,序列整体呈上升趋势,非线性特征和不稳定性特征显著。

3.2 预测结果分析

本文应用 4 个常见的误差指标,包括平均绝对百分比误差(MAPE),平均绝对误差(MAE),

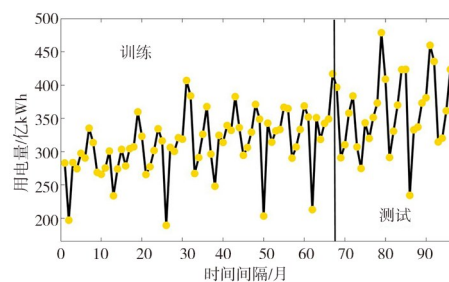


图 1 江苏省工业用电量曲线

均方根误差(RMSE)和预测误差标准差(S_{id}),来评估提出模型的预测准确性和预测稳定性。具体地,MAPE,MAE 和 RMSE 越小,表明预测准确度越高; S_{id} 越小表明预测稳定性越高。

从传统的统计模型、机器学习模型和混合模型中选取了 5 个对比模型,来证明提出的混合预测模型优于传统模型,不同预测模型负荷预测结果评价见表 1,且提出的混合预测模型的每一个组成部分均为最终预测精度的提高做出了重大贡献。

表 1 不同预测模型负荷预测结果评价

模型	MAPE /%	MAE /亿 kWh	RMSE /亿 kWh	S_{id} /亿 kWh
BP	18.956 3	72.440 0	72.440 0	92.947 0
ELM	18.304 2	67.289 3	83.778 1	64.112 0
ARIMA	24.414 3	82.152 7	107.495 3	108.906 7
CEEMDAN- ELM	14.086 2	47.515 6	63.593 3	64.195 9
MOFPA- ELM	14.382 8	52.711 5	68.439 5	53.785 5
CEEMDAN- MOFPA- ELM	8.938 0	31.520 1	39.453 1	38.554 7

表 1 列出了不同预测模型的预测精度与预测稳定性对比情况。从表 1 可以看出,BP,ELM 和 ARIMA 的 MAPE 值分别为 18.956 3%,18.304 2%和 24.414 3%,BP 和 ELM 表现相近,但是优于 ARIMA 模型,说明机器学习模型在处理非线性时间序列方面具有一定优势,ARIMA 模型不擅长预测非线性趋势明显且波动剧烈的时间序列。CEEMDAN-ELM 的 MAPE 为 14.086 2%显著低于单模型,且与 ELM 模型相比,降低了 ELM 模型 MAPE 的 34.57%,说明 CEEMDAN 在提取原始时间序列特征,剔除噪声因素负面影响方面具有显著成效,且非常有利于提高模型的预测精度。MOFPA-ELM 的 MAPE,MAE 和 RMSE 分别为 14.382 8%,52.711 5 亿 kWh 和 68.439 5 亿 kWh,显著优于 ELM 模型,说明 MOFPA 算法对于提高 ELM 的预测性能具有重要的帮助作用。

CEEMDAN-MOFPA-ELM 模型具有最小的 $MAPE$, MAE 和 $RMSE$ 值,说明在众多对比模型中,提出的模型具有最高的预测精度。

从 S_{td} 的值可看出模型预测的稳定性,具体地,MOFPA-ELM 的 S_{td} 为 53.785 5 亿 kWh,而 ELM 模型的 S_{td} 为 64.112 0 亿 kWh,说明多目标优化算法在提高模型预测精度的同时提高了模型预测的稳定性。CEEMDAN-MOFPA-ELM 模型的 S_{td} 为 38.554 7 亿 kWh,显著低于其他对比模型的方差,说明提出的模型的预测稳定性最高。

图 2 描绘了不同对比模型和提出模型的预测值情况。从图 2 可以看出,提出的 CEEMDAN-MOFPA-ELM 模型的预测值与真实值更加接近,拟合效果更好。其它对比模型的预测值与真实值偏离较远,说明模型的训练与学习能力有限,无法提供准确的预测结果。

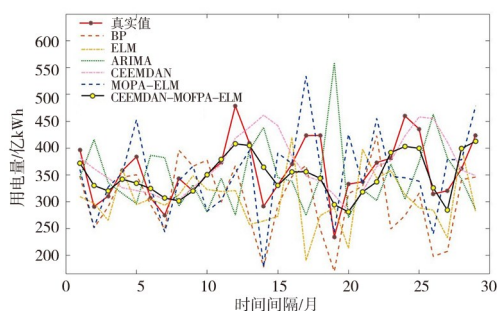


图 2 不同预测模型电力负荷预测结果折线

4 结论

本文提出了一种基于机器学习的电力负荷混合预测模型,它结合了 CEEMDAN 分解技术和

MOFPA 优化算法,改进了传统的极限学习机的缺陷,剔除了数据中的随机噪声对预测结果的负面影响。从实验结果可以看出,混合预测模型的预测效果显著优于传统的统计模型 ARIMA 以及机器学习模型 BP 和 ELM,同时,提出的模型在所有混合对比模型中表现最优。因此,提出的混合模型显著提高了预测精度和预测稳定性,可以为电网调度提供有力的信息支撑。

参考文献:

- [1] 郭威麟,蒋晓艳,罗意,等.基于 EEMD-GABP 的某地区短期负荷预测研究[J].江苏电机工程,2019,38(6):93-98.
- [2] 龙金莲.基于改进粒子群优化 LSSVM 的短期电力负荷预测[J].机电信息,2016(36):25-25.
- [3] 李冬辉,尹海燕,郑博文.基于 MFOA-GRNN 模型的年电力负荷预测[J].电网技术,2018,42(2):585-590.
- [4] ZHANG W Y, QU Z X, ZHANG K Q, et al. A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting[J]. Energy Conversion and Management, 2017, 136(3): 439-451.
- [5] YANG W D, WANG J Z, LU H Y, et al. Hybrid wind energy forecasting and analysis system based on divide and conquer scheme: A case study in China[J]. Cleaner Production, 2019, 222 (6): 942-959.
- [6] FU W L, WANG K, LI C S, et al. Multi-step short-term wind speed forecasting approach based on multi-scale dominant ingredient chaotic analysis, improved hybrid GWO-SCA optimization and ELM[J]. Energy Conversion and Management, 2019, 187(5): 356-377.
- [7] 吕勤,杨靖萍,李峰.基于分解集成策略和多目标优化算法的混合预测模型在电力负荷预测中的应用[J].河北电力技术,2020,39(6):24-27,39.

本文责任编辑:王洪娟

下 期 要 目

城市配电网故障定位技术的分析

660 MW 亚临界机组清洁疏水系统节能改造

发变组过励磁保护误动原因分析及处理措施

励磁变压器励磁涌流引起的调相机跳闸分析

500 kV 断路器失灵保护误动作跳闸故障分析