

基于Prophet算法和Blending集成学习的 实时负荷中期预测

邰嘉嘉¹, 李代猛², 杜云飞², 沈欣炜², 张璇², 乔百豪³, 何春庚¹, 蓝晓东¹, 罗澍忻¹

(1. 广东电网有限责任公司电网规划研究中心, 广东 广州 510220;

2. 清华大学深圳国际研究生院, 广东 深圳 518055; 3. 中原工学院 电子信息学院, 河南 郑州 451191)

摘要:目前的中期负荷预测一般未考虑负荷实时状态,而负荷数据的非线性、季节性、随机性、时序性特征将影响实时负荷的中期预测。构建一个实时负荷中期预测的框架,采用Prophet算法提取负荷数据的季节性部分,采用Blending集成学习对负荷数据的非季节部分进行滚动预测,将季节性部分和非季节性部分合成中期负荷实时数据。爱尔兰电力系统的算例结果验证了模型的有效性和稳定性。

关键词:负荷预测;Prophet算法;Blending集成学习;季节性

中图分类号:TM714

文献标志码:A

DOI:10.16081/j.epae.202308025

0 引言

负荷预测是电力系统的一项重要研究课题。根据预测时长,负荷预测可以分为短期负荷预测、中期负荷预测和长期负荷预测^[1-2]。其中,中期负荷预测的预测时长通常超过一周,对电力系统的规划、运营和维护具有重要影响^[3]。目前,中长期负荷预测主要关注负荷总量的预测,其中所使用的负荷数据更新频率为天或月甚至更长时间,而没有考虑负荷的实时状态^[4-6]。

负荷数据是具有明显的非线性、季节性、随机性和时序性的序列,这增加了负荷预测的难度^[7]。机器学习是一种有效解决数据中非线性问题的方法,然而机器学习方法种类繁多,各有优缺点。集成学习是一种用于提高模型性能的有效机器学习技术,通过组合多个基本学习器提高模型的预测能力^[8-9]。长短期记忆(long short-term memory, LSTM)网络因能够兼顾时序性和非线性而逐渐成为负荷预测中最常见的机器学习模型^[10],例如:文献[8]利用适应性矩估计优化算法和多层LSTM网络,提出一种考虑温度模糊化的短期负荷预测方法;文献[11]利用双向LSTM神经网络和注意力机制,提出一种具有气象数据修正的短期负荷预测方法。然而,LSTM网络主要用于短期负荷预测,没有充分考虑负荷数据的季节性。其他机器学习方法,如支持向量回归^[12]、

Stacking集成学习^[13]等,也面临上述问题。一些研究考虑了负荷季节性的影响,例如:文献[14]结合极限梯度提升(extreme gradient boosting, XGBoost)算法和Prophet算法对短期日负荷进行研究,考虑负荷的季节效应,获得了具有较高准确度的负荷预测结果;文献[15]利用奇异谱分析方法将负荷分解为主要周期成分,对分解出的序列进行预测后再将其叠加,实现了中期负荷预测。尽管文献[14-15]考虑了负荷季节性,但没有预测负荷的实时状态。

基于上述背景,本文主要致力于构建一个实时负荷中期预测框架,以同时考虑负荷数据的非线性、季节性、随机性和时序性特征。首先,采用Prophet算法^[16]将负荷数据分解为季节性和非季节性两部分,其中季节性部分仅与时间相关,而非季节性部分受多种因素的影响;然后,针对非季节性部分,采用Blending集成学习进行预测,其中预测模型采用自回归模型进行拟合,并集成已有研究中的多种负荷预测算法,以实现准确预测中期负荷的目标;最后,采用滚动预测的方式获得未来一段时间内负荷非季节性部分的实时数据,并将其与季节性部分结合,生成中期负荷的实时数据。

1 Prophet算法和Blending集成学习

1.1 Prophet算法

Prophet算法本质上是一个时间序列分解模型,将时间序列分解为趋势项、季节项、节假日项和误差项,模型可以描述为^[17]:

$$y(t)=g(t)+s(t)+h(t)+\epsilon(t) \quad (1)$$

式中: t 表示时刻; $y(t)$ 为待分解的时间序列; $g(t)$ 为趋势项,用于表征时间序列的非周期趋势; $s(t)$ 为季节项,用于表征具有固定周期性的影响; $h(t)$ 为节假日项,用于表征时间序列在特定日期受到的影响;

收稿日期:2023-04-04;修回日期:2023-06-29

在线出版日期:2023-09-27

基金项目:中国南方电网重点科技项目(支撑多能互补园区规划的能效数据挖掘技术研究项目)(GDKJXM20202019)

Project supported by the Key Science and Technology Project of China Southern Power Grid (Research Project of Energy Efficiency Data Mining Technology for Multi-energy Complementary Park Planning)(GDKJXM20202019)

$\epsilon(t)$ 为误差项,服从高斯分布。 $g(t)$ 、 $s(t)$ 和 $h(t)$ 的具体介绍如附录A所示。

1.2 Blending集成学习

集成学习通过构建并结合多个学习器来完成学习任务,也被称为多分类器系统、基于委员会的学习等^[17]。模型集成方法大致分为投票法、平均法、排序法、Stacking堆叠法和Blending堆叠法。后2种方法是相对高级的模型集成方法。Stacking堆叠法的建模思路是基于原始数据,训练出多个基学习器,再将基学习器的预测结果组合成新的训练集,以训练一个新的学习器(即元学习器),但Stacking集成学习的基学习器和元学习器在训练时采用相同的训练集,容易造成信息泄露。而Blending集成学习会对原始数据划出留出集,用于元学习器的训练,可以有效保护信息隐私,因此,本文选择Blending集成学习。

2 基于Prophet算法和Blending集成学习的中期负荷预测框架

本文将负荷分为季节性部分和非季节性部分。季节性部分是特定周期下的循环波动,仅与时间相关。非季节性部分是受一系列外部因素影响而产生的波动。因此,负荷预测模型可以表示为:

$$y_{\text{pred}}(t) = d_{\text{pred}}(t) + s(t) \quad (2)$$

式中: $y_{\text{pred}}(t)$ 、 $d_{\text{pred}}(t)$ 分别为总负荷的预测序列和负荷非季节项的预测序列。

2.1 数据集的创建与划分

假设已有数据的时间跨度为0— T ,读取该时段的历史负荷曲线(包括冷、热、电、气负荷曲线)和外部数据曲线(包括但不限于天气曲线),并根据数据特征采用合理的数据预处理方法,进而构造整个模型所需的数据集。按照合适比例将数据集按时间顺序划分为 $[0, T_1]$ 、 $[T_1, T_2]$ 和 $[T_2, T]$ 3个时段的数据集,分别命名为 data_1 、 data_2 和 data_3 , T_1 和 T_2 为间断点的时间戳。

2.2 负荷季节性效应的提取

利用Prophet算法在数据集 data_1 中提取负荷的季节性效应,模型为:

$$y(t) = g(t) + s(t) + \epsilon(t) \quad (3)$$

季节项 $s(t)$ 包含 K 个分量 $s_{p_1}(t) \sim s_{p_K}(t)$,分别表示周期步长为 $P_1 \sim P_K$ 的季节效应,即:

$$s(t) = s_{p_1}(t) + s_{p_2}(t) + \dots + s_{p_K}(t) \quad (4)$$

令一天的时间步长为 N_{day} ,年度步长为:

$$P_r = 365.25 N_{\text{day}} \quad r = 1, 2, \dots, K \quad (5)$$

周度步长为:

$$P_r = 7 N_{\text{day}} \quad r = 1, 2, \dots, K \quad (6)$$

日度步长为:

$$P_r = N_{\text{day}} \quad r = 1, 2, \dots, K \quad (7)$$

上述3个时间步长的季节项分量是该框架主要考虑的季节项分量。年度季节性的考虑主要是为了去除年度气象因素的季节性影响;人类生产生活在工作日和非工作日存在显著不同,因此,负荷存在周度季节性;由于人类生产生活和外部气象条件在一天内的不同时刻也存在一定的季节性,因此,考虑了日度季节性。需要说明的是,由于实时数据通常是分钟级的数据,引入节假日效应会在模型中引入过多的指示变量,从而使模型变得复杂,容易导致过拟合问题,因此,不考虑节假日效应。

2.3 负荷非季节部分预测模型

模型需要生成负荷的非季节项 $d(t)$,因此还需要构建一个去除季节效应的数据集,用于训练负荷非季节项预测模型,即:

$$d(t) = y(t) - s(t) \quad (8)$$

通过上述方式去除数据集 data_2 的季节效应,得到数据集 data_{20} ,用于负荷非季节性部分预测模型的训练和测试。值得注意的是,季节效应分解和负荷非季节项预测采用的是不同的数据集,这是由于季节效应分解后的数据集受到了污染,会对负荷非季节项预测模型的训练造成影响,进而影响负荷非季节项数据生成的有效性。

模型采用Blending集成学习预测 $t+1$ 时刻负荷的非季节性部分,步骤如下。

1)数据集的划分。将数据集 data_{20} 按照合适的比例划分为基学习器的训练集 data_{21} 和测试集 data_{22} 。将数据集 data_{22} 也按照一定的比例划分为元学习器的训练集和测试集。

2)基学习器的训练与测试。基学习器的标签项为 $d(t)$,特征项包括 $d(t)$ 的前几期值和 t 时刻的外部数据 $\mathbf{x}(t)$ 。学习器的算法包括线性回归(linear regression, LR)、 k 近邻(k -nearest neighbor, kNN)算法、CatBoost算法、岭回归(Ridge)算法、Lasso回归算法、多层感知机(multilayer perceptron, MLP)算法、决策树(decision tree, DT)算法、极限树(extra tree, ET)算法、XGBoost算法、轻量级梯度提升机(light gradient boosting machine, LightGBM)算法、随机森林(random forest, RF)算法、AdaBoost算法、梯度提升决策树(gradient boosting decision tree, GDBT)算法、引导聚集(Bagging)算法等。因此,基学习器的学习模型为:

$$d(t) = f_z(d(t-g), d(t-g+1), \dots, d(t-1), \mathbf{x}(t)) \quad (9)$$

式中: $f_z(\cdot)$ 为第 z 个学习器的模型,模型具体形式根据对应的算法确定; g 为 $d(t)$ 的滞后期数,其值根据需求确定。

在基学习器训练结束后,将 data_{22} 作为测试集,

根据不同基学习器的预测效果调整出合适的模型参数。

3)元学习器的训练与测试。利用所有训练完成的基学习器对数据集 data_{22} 进行预测,第 z 个学习器预测结果为 $d_z(t)$ ($z=1, 2, \dots, Z, Z$ 为基学习器总数)。元学习器的标签项为 $d(t)$,元学习器的特征项为 $d_z(t)$,构造新的数据集 data_{23} ,元学习器采用的算法根据需求确定。将数据集 data_{23} 按照一定比例划分为训练集和测试集。因此,元学习器的学习模型为:

$$d(t)=f_{\text{meta}}(d_1(t), d_2(t), \dots, d_z(t)) \quad (10)$$

式中: $f_{\text{meta}}(\cdot)$ 为元学习器模型,模型具体形式根据对应的算法确定。模型训练结束后在测试集上进行测试并辅助模型的调参。

对基学习器和元学习器进行整合,可以得到负荷非季节项的预测模型,如式(11)所示。

$$d_{\text{pred}}(t)=f(d(t-g), d(t-g+1), \dots, d(t-1), x(t-1)) \quad (11)$$

式中: $f(\cdot)$ 为集成学习模型。

2.4 负荷曲线合成

负荷曲线合成共分为2步:第一步是利用训练好的基学习器和元学习器滚动预测生成未来一段时间的负荷非季节项序列;第二步是根据负荷所在时间点,通过添加季节效应合成未来一段时间的负荷曲线。

假设以 T 时刻为时间起点,可将 T 时刻的负荷非季节项值和 $T+1$ 时刻的外部数据作为输入,生成 $T+1$ 时刻的负荷非季节项预测值,再将 $T+1$ 时刻的负荷非季节项预测值和 $T+2$ 时刻的外部数据作为输入,生成 $T+2$ 时刻的负荷非季节项预测值,如此循环,可滚动生成 $T+1-T+N$ 时段的负荷非季节项序列,其中 N 为负荷预测的时长。

通过上述过程合成了一条未来的负荷非季节项曲线 $d_{\text{pred}}(t)$,再根据式(2)即可合成一条未来的总负荷曲线 $y_{\text{pred}}(t)$ 。

2.5 测试与验证

在测试和验证阶段使用的数据集为 data_3 ,采用6个评价指标,分别为决定系数 R^2 、中位绝对误差(median absolute error, MDAE)、均方根误差(root mean square error, RMSE)、平均绝对误差(mean absolute error, MAE)、平均绝对百分比误差(mean absolute percentage error, MAPE)及对称平均绝对百分比误差(symmetric mean absolute percentage error, SMAPE),计算公式分别见附录B式(B1)~(B6)。

3 算例分析

本文基于 EirGrid 集团提供的爱尔兰电力系统负荷数据集进行实时负荷的中期预测,以验证所提

框架的性能。考虑到真实数据较少,仅采用爱尔兰能源系统数据集中2019年1月1日00:00至2020年4月30日23:45的数据(共计46652个时间点)。该数据集主要包括爱尔兰与北爱尔兰的总发电量、电力需求量、风电发电量、风电可用性、非同步发电瞬时渗透率等指标数据,数据为没有进行质量控制的15 min数据采集与监视控制系统读数。数据更新周期为15 min。

3.1 数据来源

以爱尔兰的发电量实时中期预测为例,将电力需求量、风电发电量、风电可用性和非同步发电瞬时渗透率等指标作为外部输入。按本文框架的划分方式:数据集 data_1 为2019年1月1日00:00至2019年6月30日23:45的数据(共17372个时间点),用于季节效应的提取;数据集 data_2 为2019年7月1日00:00至2020年1月20日23:45的数据(共19680个时间点),用于构建非季节性部分的预测模型;数据集 data_3 为2020年1月21日00:00至2020年4月20日23:45的数据(共9600个时间点),用于测试整个生成模型的效果。

3.2 季节性效应的提取

由于数据集 data_1 的数据量不足以支持年度效应的提取,因此,仅考虑周度效应和日度效应。本文利用 Prophet 算法提取数据集 data_1 中发电量负荷曲线的季节效应,分解出周度和日度负荷曲线,结果如图1所示。为了清晰地展示结果,图1仅显示前8000个时间点的数据,时间起点为2019年1月1日

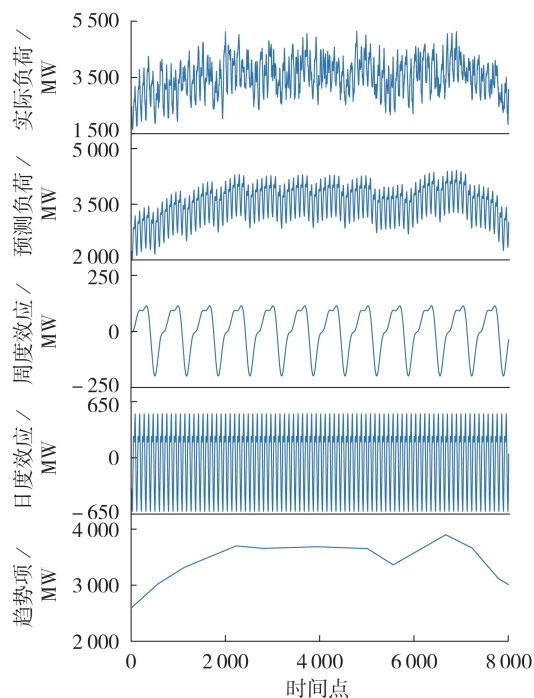


图1 发电量曲线分解

Fig.1 Decomposition of power generation curve

00:00。图中从下往上分别为负荷的趋势项、日度季节性和周度季节性分解结果,3个分解结果合成的预测负荷,实际负荷。

由图 1 可知,发电量负荷呈现明显的周度和日度效应。日度效应比周度效应呈现更高的波峰,这表明负荷的日度效应比周度效应更明显。在日度效应中,白天的效应强于夜间,并且在午间存在 1 个局部波谷,这可能是由于午休期间的社会用电需求会短暂降低,而在 15:00 左右,全社会的用电需求达到波峰。对于周度效应而言,工作日的效应强于周末,波峰出现在周三左右,波谷出现在周六左右,这与实际场景相符合。总体而言,模型的分解效果非常理想。

3.3 非季节性部分的预测

在算例预测过程中,基学习器采用的算法包括 kNN、CatBoost、Ridge、Lasso、MLP、DT、ET、XGBoost、LightGBM、RF、AdaBoost、GDBT 和 Bagging,元学习器采用 LR 算法,自回归采用滞后五阶,按照 7:3 划分基学习器和元学习器的样本,按照 7:3 划分元学习器的训练集和测试集。所有学习器在训练集上的表现如表 1 所示。

表 1 不同学习器的性能

Table 1 Performance of different learners

学习器	$R^2 / \%$	学习器	$R^2 / \%$
kNN	98.09	XGBoost	98.95
CatBoost	99.04	LightGBM	99.00
Ridge	99.12	RF	98.91
Lasso	99.05	AdaBoost	98.48
MLP	98.97	GDBT	98.81
DT	97.97	Bagging	98.83
ET	97.94	LR	96.31

由表 1 可知,每个学习器均有不错的预测性能, R^2 均高于 95%。值得注意的是,基学习器在单点预测方面的准确度明显优于整个集成学习模型,但是在滚动预测方面,基学习器的精度和稳定性均低于整个集成学习模型。以滚动预测 100 d(9 600 个时间点)为例,对表 1 基学习器中表现更优的 Ridge、Lasso、CatBoost 以及元学习器 LR 进行单独训练并进行滚动预测,将其预测性能与 Blending 集成学习模型的预测性能进行对比,结果如图 2 所示。

由图 2 可知:Blending 集成学习在 6 个性能指标上均表现最优,这表明 Blending 集成学习能够同时提高模型的预测精度和稳定性;单点预测精度并不能直接反映滚动预测精度。由于本文所提模型需要预测中期负荷曲线,因此,更精确的滚动预测结果是模型的主要目标。

3.4 负荷曲线合成

根据训练好的 Blending 集成学习模型和外部输入,本文进行 100 d(9 600 个时间点)的滚动预测,得

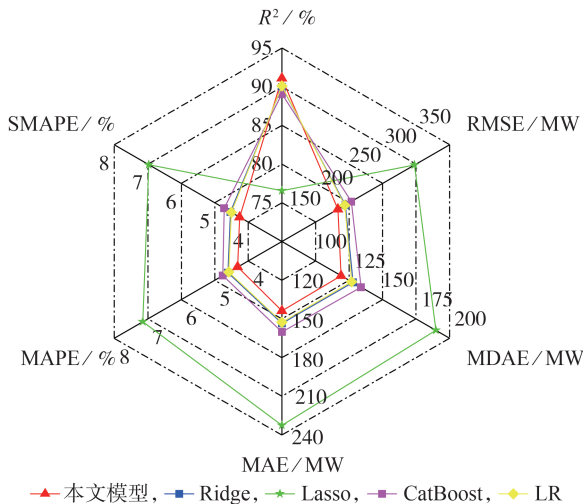


图 2 不同模型的滚动预测性能

Fig.2 Rolling forecasting performance of different models

到负荷的非季节性部分,再根据时间属性和季节效应生成最终的负荷曲线,如图 3 所示。由图可知,预测曲线和实际曲线的走势非常相似,除个别点存在较大的误差外,其他点的误差都在 0 附近且较小,这充分说明了本文模型预测的准确性。

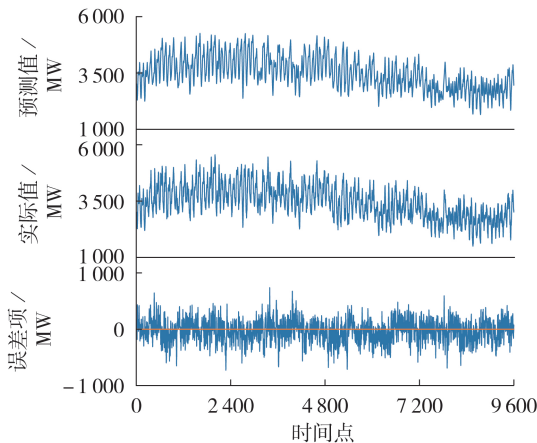


图 3 负荷的预测结果和实际结果

Fig.3 Load forecasting results and actual results

表 2 展示了模型评价指标。由表可知:尽管非季节部分已经达到了非常高的预测精度,但是加入季节项后,模型的绝对指标明显提升,特别是 R^2 指标;MAPE 和 SMAPE 略有减小,这是由于这 2 个指标是相对指标而不是绝对指标,加入季节效应后,指标计算公式的分母增大,从而使指标减小;加入季节项后整体预测精度得到提升,这表明 Prophet 算法和 Blending 集成学习模型的结合能够有效提升模型的预测效果。

3.5 模型预测稳定性

为了评估模型预测稳定性,对不同预测时间尺度下的模型性能进行测试。图 4 展示了不同预测时

表2 模型评价指标

Table 2 Evaluation index of model

负荷	R^2 / %	RMSE / MW	MDAE / MW	MAE / MW	MAPE / %	SMAPE / %
非季节部分	91.01	184.48	120.01	144.45	4.34	4.28
实际负荷	93.66	183.58	119.17	143.74	4.36	4.30

间尺度下的模型性能。由图可知,每个指标均呈现明显的收敛趋势,并收敛到一个非常理想的值,这说明在满足精度要求的前提下,所提模型具有良好的稳定性,适用于中期负荷预测。

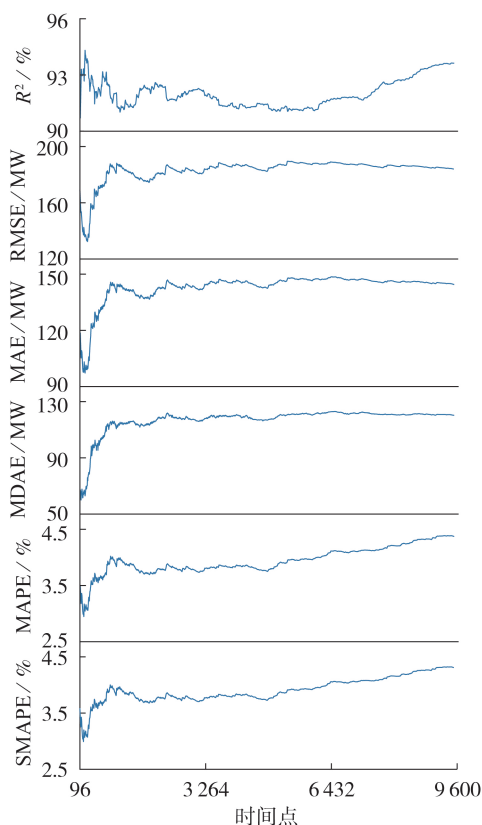


图4 不同预测时间尺度下的模型性能

Fig.4 Model performance under different forecasting time scales

4 结论

本文结合 Prophet 算法和 Blending 集成学习,综合考虑负荷波动的非线性、季节性、随机性和时序性特点,构建一个实时负荷中期预测框架。爱尔兰电力系统实际数据的验证结果表明,该框架能够准确地反映负荷的季节性,并准确预测非季节性负荷,从而实现高质量的实时负荷中期预测。后续笔者将进一步改进 Prophet 算法和 Blending 集成学习的融合方式,以降低对数据量的需求,并进一步提升模型的预测性能。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] MOTEPE S, HASAN A, STOPFORTH R. Improving load forecasting process for a power distribution network using hybrid AI and deep learning algorithms[J]. IEEE Access, 2019, 7: 82584-82598.
- [2] PATEL A, DAS M, GHOSH S K. Short-term load forecasting: an intelligent approach based on recurrent neural network [M]//Hybrid Intelligent Systems. [S.l.]: Springer International Publishing, 2020: 52-62.
- [3] PATIL M S, LOKA R, PARIMI A M. Application of ARIMA and 2D-CNNs using recurrence plots for medium-term load forecasting [C]//2021 IEEE 2nd China International Youth Conference on Electrical Engineering (CIYCEE). Chengdu, China: IEEE, 2022: 1-5.
- [4] 李亦言, 严正, 冯冬涵. 考虑城市化因素的中长期负荷预测模型[J]. 电力自动化设备, 2016, 36(4): 54-61.
LI Yiyang, YAN Zheng, FENG Donghan. Mid / long-term load forecasting model considering urbanization characteristics[J]. Electric Power Automation Equipment, 2016, 36(4): 54-61.
- [5] 王雁凌, 吴梦凯. 经济新常态下基于偏最小二乘回归的中长期负荷预测模型[J]. 电力自动化设备, 2018, 38(3): 133-139.
WANG Yanling, WU Mengkai. Medium and long term load forecasting model based on partial least-square regression under new normal economy[J]. Electric Power Automation Equipment, 2018, 38(3): 133-139.
- [6] 刘俊, 赵宏炎, 刘嘉诚, 等. 基于协整-格兰杰因果检验和季节分解的中期负荷预测[J]. 电力系统自动化, 2019, 43(1): 73-80.
LIU Jun, ZHAO Hongyan, LIU Jiacheng, et al. Medium-term load forecasting based on cointegration-Granger causality test and seasonal decomposition[J]. Automation of Electric Power Systems, 2019, 43(1): 73-80.
- [7] SINGH P, DWIVEDI P. A novel hybrid model based on neural network and multi-objective optimization for effective load forecast[J]. Energy, 2019, 182: 606-622.
- [8] 郑瑞晓, 张姝, 肖先勇, 等. 考虑温度模糊化的多层长短期记忆神经网络短期负荷预测[J]. 电力自动化设备, 2020, 40(10): 181-186.
ZHENG Ruixiao, ZHANG Shu, XIAO Xianyong, et al. Short-term load forecasting of multi-layer long short-term memory neural network considering temperature fuzziness[J]. Electric Power Automation Equipment, 2020, 40(10): 181-186.
- [9] MAHESH T R, KUMAR D, KUMAR V, et al. Blended ensemble learning prediction model for strengthening diagnosis and treatment of chronic diabetes disease[J]. Computational Intelligence and Neuroscience, 2022, 2022: 1-9.
- [10] 陆继翔, 张琪培, 杨志宏, 等. 基于 CNN-LSTM 混合神经网络模型的短期负荷预测方法[J]. 电力系统自动化, 2019, 43(8): 131-137.
LU Jixiang, ZHANG Qipei, YANG Zhihong, et al. Short-term load forecasting method based on CNN-LSTM hybrid neural network model[J]. Automation of Electric Power Systems, 2019, 43(8): 131-137.
- [11] 王继东, 杜冲. 基于 Attention-BiLSTM 神经网络和气象数据修正的短期负荷预测模型[J]. 电力自动化设备, 2022, 42(4): 172-177, 224.
WANG Jidong, DU Chong. Short-term load prediction model based on Attention-BiLSTM neural network and meteorological data correction[J]. Electric Power Automation Equipment, 2022, 42(4): 172-177, 224.
- [12] CHEN Y B, XU P, CHU Y Y, et al. Short-term electrical load forecasting using the support vector regression (SVR) model

- to calculate the demand response baseline for office buildings [J]. *Applied Energy*, 2017, 195: 659-670.
- [13] 崔树银, 汪昕杰. 基于最大信息系数和多目标Stacking集成学习的综合能源系统多元负荷预测[J]. *电力自动化设备*, 2022, 42(5): 32-39, 81.
- CUI Shuyin, WANG Xinjie. Multivariate load forecasting in integrated energy system based on maximal information coefficient and multi-objective Stacking ensemble learning[J]. *Electric Power Automation Equipment*, 2022, 42(5): 32-39, 81.
- [14] 刘涅焯, 杨学良, 陶晓峰, 等. 基于改进Prophet算法的短期日负荷预测方法研究[J]. *电力需求侧管理*, 2022, 24(5): 58-63.
- LIU Niexuan, YANG Xueliang, TAO Xiaofeng, et al. Short-term power load forecasting based on improved Prophet algorithm[J]. *Power Demand Side Management*, 2022, 24(5): 58-63.
- [15] 陈浩文, 刘文霞, 李月乔. 基于奇异谱分析与神经网络的中期负荷预测[J]. *电网技术*, 2020, 44(4): 1333-1347.
- CHEN Haowen, LIU Wenxia, LI Yueqiao. Medium-term load forecast based on singular spectrum analysis and neural network[J]. *Power System Technology*, 2020, 44(4): 1333-1347.
- [16] TAYLOR S J, LETHAM B. Forecasting at scale[J]. *The American Statistician*, 2018, 72(1): 37-45.
- [17] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 171-190.

作者简介:

郇嘉嘉(1983—), 女, 高级工程师, 博士, 主要研究方向为电网规划、综合能源系统规划等(E-mail: winnie5983@126.com);

张璇(1988—), 男, 助理教授, 博士, 通信作者, 主要从事物理信息系统、现代控制理论、凸优化、基于学习的控制等领域的研究工作(E-mail: xuanzhang@sz.tsinghua.edu.cn)。

(编辑 王锦秀)

Mid-term forecasting of real-time load based on Prophet algorithm and Blending integrated learning

HUAN Jiajia¹, LI Daimeng², DU Yunfei², SHEN Xinwei², ZHANG Xuan², QIAO Baihao³,
HE Chungeng¹, LAN Xiaodong¹, LUO Shuxin¹

(1. Power Grid Planning Research Center of Guangdong Power Grid Co., Ltd., Guangzhou 510220, China;

2. Tsinghua Shenzhen International Graduate School, Shenzhen 518055, China;

3. School of Electronic and Information, Zhongyuan University of Technology, Zhengzhou 451191, China)

Abstract: The current medium-term load forecasting generally doesn't consider the real-time state of the load. However, the characteristics of load data such as nonlinearity, seasonality, randomness and temporality will influence the medium-term forecasting of real-time load. A framework for mid-term forecasting of real-time load is constructed. The Prophet algorithm is adopted to extract the seasonal component of the load data. The Blending integrated learning is adopted for the rolling forecasting of non-seasonal component of the load data. The seasonal and non-seasonal components are combined to synthesize the real-time data of mid-term load. The effectiveness and stability of the model are verified by Irish Power System.

Key words: load forecasting; Prophet algorithm; Blending integrated learning; seasonality

附录 A:

A1.趋势项 $g(t)$

趋势项为分段函数，有两种形式：分段 logistic 函数和分段线性函数。相连的两个分段函数所在的点为变化点，假设存在 S 个变化点，第 j 个变化点的时间戳为 s_j ，则趋势项的两种表达形式分别为：

$$g(t) = \frac{C(t)}{1 + \exp\left(-\left(k + \mathbf{a}(t)^T \boldsymbol{\delta}\right)\left(t - \left(m + \mathbf{a}(t)^T \boldsymbol{\gamma}\right)\right)\right)} \quad (\text{A1})$$

$$g(t) = \left(k + \mathbf{a}(t)^T \boldsymbol{\delta}\right) \times t + \left(m + \mathbf{a}(t)^T \boldsymbol{\gamma}\right) \quad (\text{A2})$$

式中： $C(t)$ 表示承载能力，为 logistic 函数的最大值； k 表示基础增长率； m 表示基础偏移量； $\mathbf{a}(t)$ 是指示向量，反映 k 和 m 在不同时间区间的变化情况：

$$\mathbf{a}(t) = [a_1(t), a_2(t), \dots, a_s(t)]^T \quad (\text{A3})$$

$$a_j(t) = \begin{cases} 1, & t \geq s_j \\ 0, & \text{其他} \end{cases} \quad j = 1, 2, \dots, S \quad (\text{A4})$$

$\boldsymbol{\delta}$ 表示增长率在不同时间区间的变化向量， δ_j 表示在时间戳 s_j 上的增长率变化量， $\delta_j \sim \text{Laplace}(0, \tau)$ ， $\text{Laplace}(\cdot)$ 表示拉普拉斯分布， τ 表示变点变化率范围，当 τ 接近于零的时候， δ_j 也是趋向于零的； $\boldsymbol{\gamma}$ 表示偏移量在不同变化点的调整向量，对于非线性模型：

$$\gamma_j = \left(s_j - m - \sum_{l < j} \gamma_l\right) \left(1 - \frac{k + \sum_{l < j} \delta_l}{k + \sum_{l \leq j} \delta_l}\right) \quad (\text{A5})$$

对于线性模型：

$$\gamma_j = -s_j \delta_j \quad (\text{A6})$$

A2.季节项 $s(t)$

Prophet 算法使用傅立叶级数来模拟时间序列的季节性。周期为 P 的季节项 $s_p(t)$ 的傅立叶级数形式为：

$$s_p(t) = \sum_{u=1}^U \left(\alpha_u \cos\left(\frac{2\pi ut}{P}\right) + \beta_u \sin\left(\frac{2\pi ut}{P}\right) \right) \quad (\text{A7})$$

U 表示展开长度，Prophet 算法对于常见的周期步长设定了特定的 U ， α_u 和 β_u 都是函数参数。周期步长为 P 的函数参数可以形成列向量 $\boldsymbol{\beta}_p$ ，其表示为：

$$\boldsymbol{\beta}_p = [\alpha_1, \beta_1, \alpha_2, \beta_2, \dots, \alpha_U, \beta_U]^T \quad (\text{A8})$$

周期步长为 P 的正弦项和余弦项合成向量 $\mathbf{X}_p(t)$ 可以表示为：

$$\mathbf{X}_p(t) = \begin{bmatrix} \cos\left(\frac{2\pi(1)t}{P}\right), \sin\left(\frac{2\pi(1)t}{P}\right), \\ \cos\left(\frac{2\pi(2)t}{P}\right), \sin\left(\frac{2\pi(2)t}{P}\right), \dots, \\ \cos\left(\frac{2\pi(U)t}{P}\right), \sin\left(\frac{2\pi(U)t}{P}\right) \end{bmatrix} \quad (\text{A9})$$

因此，周期步长为 P 的季节效应项为：

$$s_p(t) = \mathbf{X}_p(t) \boldsymbol{\beta}_p \quad (\text{A10})$$

式中： $\boldsymbol{\beta}_p \sim N(0, \sigma_p^2)$ ， σ_p 是周期为 P 的季节效应的范围， σ_p 越大，表示季节效应越显著；反之表示季节效应不明显。

A3.节假日项 $h(t)$

节假日（或者某些大事件）具有一定的时间窗口，因此需要增加一个指示向量 $\mathbf{Z}(t)$ ，表征时间 t 是否属于假期 l 的窗口期 D_l ，并给每个假期分配一个参数 κ_l ，即：

$$\mathbf{Z}(t) = [z_1(t), z_2(t), \dots, z_L(t)] \quad (\text{A11})$$

$$z_i(t) = \begin{cases} 1 & t \in D_l \\ 0 & \text{其他} \end{cases} \quad l = 1, 2, \dots, L \quad (\text{A12})$$

式中： L 表示假期的个数。节假日项模型可以表示为：

$$h(t) = \mathbf{Z}(t) \boldsymbol{\kappa} \quad (\text{A13})$$

$$\boldsymbol{\kappa} = [\kappa_1, \kappa_2, \dots, \kappa_L]^T \quad (\text{A14})$$

式中： $\boldsymbol{\kappa} \sim N(0, v^2)$ ， v 表示节假日效应的范围。

附录 B:

$$R^2 = 1 - \frac{\sum (y(t) - y_{pre}(t))^2}{\sum (y(t) - y_{mean})^2} \quad (\text{B1})$$

$$\text{MDAE} = \text{median}(|y(t) - y_{pre}(t)|) \quad (\text{B2})$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum (y(t) - y_{pre}(t))^2} \quad (\text{B3})$$

$$\text{MAE} = \frac{1}{n} \sum |y(t) - y_{pre}(t)| \quad (\text{B4})$$

$$\text{MAPE} = \frac{100\%}{n} \sum \left| \frac{y(t) - y_{pre}(t)}{y(t)} \right| \quad (\text{B5})$$

$$\text{SMAPE} = \frac{100\%}{n} \sum \left| \frac{y(t) - y_{pre}(t)}{(|y(t)| + |y_{pre}(t)|) / 2} \right| \quad (\text{B6})$$

式中： $y(t)$ 为负荷实际值； $y_{pre}(t)$ 为预测值； y_{mean}

为负荷实际值的均值， n 是负荷预测的时间长度，

$\text{median}(\cdot)$ 表示中位数函数。