

基于多种模型的电力系统负荷预测

王 格, 陈 鑫

重庆理工大学理学院, 重庆

收稿日期: 2024年1月28日; 录用日期: 2024年2月22日; 发布日期: 2024年2月29日

摘 要

对于时间序列的拟合预测模型有很多, 但随着实际问题的不断变化, 传统的时间序列预测模型已经渐渐不满足要求, 现如今更加需要的是以问题为导向的, 具有高精度, 高泛化能力的预测模型。本文采用了ARIMA、机器学习算法对某地区未来十天间隔十五分钟用电负荷进行预测, 并通过预测值和观测值的MAE、RMSE、MAPE对精度进行了比较, 研究发现精度最好的为GBDT模型, 所以最终选取GBDT模型来进行后续的预测工作。

关键词

ARIMA, 机器学习, 电力负荷预测

Power System Load Forecasting Based on Multiple Models

Ge Wang, Xin Chen

College of Science, Chongqing University of Technology, Chongqing

Received: Jan. 28th, 2024; accepted: Feb. 22nd, 2024; published: Feb. 29th, 2024

Abstract

There are many fitting prediction models for time series, but with the constant change of practical problems, the traditional time series prediction model has gradually failed to meet the requirements, and now more problem-oriented prediction models with high precision and high generalization ability are needed. In this paper, ARIMA and machine learning algorithms are used to predict the electricity load in a region at an interval of 15 minutes in the next ten days, and the accuracy is compared by MAE, RMSE and MAPE of the predicted value and observed value. It is found that the GBDT model has the best accuracy, so GBDT model is finally selected for the subsequent prediction work.

Keywords

ARIMA, Machine Learning, Power System Load Forecasting

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

电力系统负荷(电力需求量, 即有功功率)预测是指充分考虑历史的系统负荷、经济状况、气象条件和社会事件等因素的影响, 对未来一段时间的系统负荷做出预测。负荷预测是电力系统规划与调度的一项重要内容。短期(两周以内)预测是电网内部机组启停、调度和运营计划制定的基础; 中期(未来数月)预测可为保障企业生产和社会生活用电, 合理安排电网的运营与检修决策提供支持; 长期(未来数年)预测可为电网改造、扩建等计划的制定提供参考, 以提高电力系统的经济效益和社会效益。

对此, 很多学者和工程师对电力系统负荷预测做了大量的研究。常见的预测模型有基于统计模型, 如时间序列预测、回归预测等, 基于机器学习和深度学习, 如支持向量积、随机森林等。Huang [1]等人提出一种带分层规则的阈值的自回归模型对电力系统小时负荷需求预测, 实验证明该方法提高了预测精度和可靠性。Lee [2]等人采用提升方案和自回归综合移动平均(ARIMA)模型进行短期电力负荷预测, Chen [3]等人首次研究非对称效应在负荷时间序列的应用, 对比各种模型的预测精度, 建立了基于多种非对称自回归条件异方差模型; Lloyd [4]等人提出一种基于梯度提升机和高斯过程的短期电力负荷预测方法, 该方法通过机器学习回归方法, 极大地提高了预测精度。Yu [5]等人提出了一种结合动态时间规整的递归神经网络, 用于预测日高峰负荷。复杂多变的气象条件和社会事件等不确定因素都会对电力系统负荷造成一定的影响, 使得传统负荷预测模型的应用存在一定的局限性。同时, 随着电力系统负荷结构的多元化, 也使得模型应用的效果有所降低, 因此电力系统负荷预测问题亟待进一步研究。本文采用了机器学习算法、深度学习算法以及新的时间序列预测模型在某地区未来十天间隔十五分钟用电负荷的预测精度进行了比较, 选取最优的模型来进行后续的预测工作。

2. 数据预处理

数据说明及预处理

针对不同问题选用不同时间段的样本数据进行数据集的划分。对于区域未来十天十五分钟负荷数据预测选近来 6 个月作为训练集, 近来十天作为测试集; 对于取悦未来三十天十五分钟符合数据预测选用近来 18 个月样本数据作为训练集, 近来三个月作为测试集。对于不同行业未来用电负荷预测, 根据时序图分析波动性与周期性, 灵活选取训练集与测试集。同时将天气数据利用 encoder 编码与行业按日期数据合并, 以作为行业数据预测的协变量。

3. 基于 ARIMA 模型的电力负荷的中短期预测

3.1. 模型数据的分析

对于时间序列数据[6], 我们首先观察序列的时序图、自相关图以及偏自相关图, 见图 1、图 2、图 3。

15分钟电力负荷图

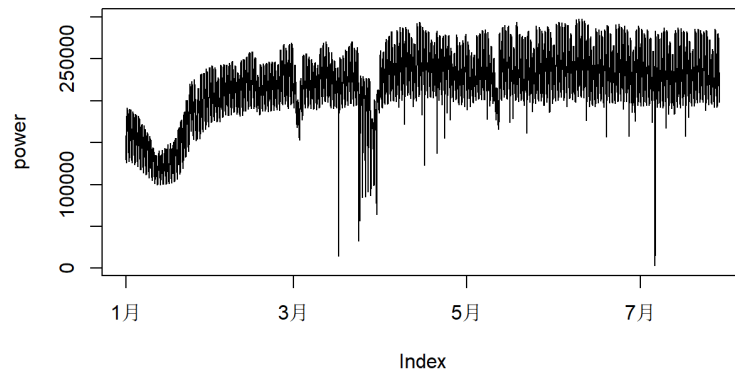


Figure 1. Sequence timing diagram

图 1. 序列时序图

Series power

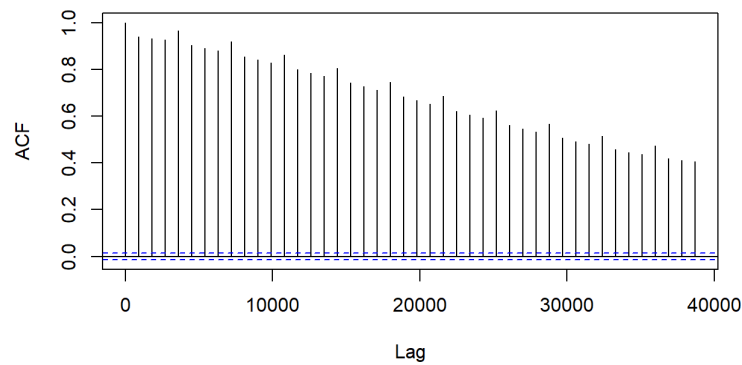


Figure 2. Sequence autocorrelation diagram

图 2. 序列自相关图

Series power

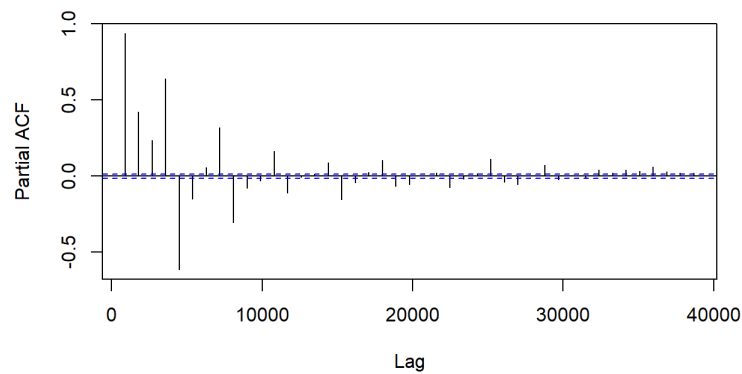


Figure 3. Sequence partial autocorrelation diagram

图 3. 序列偏自相关图

由于 2021 年 5 月份之前的数据有很多的异常值, 经讨论认为, 认为这些异常值可能是由于节假日或者极端天气的影响, 但由于开始我们只是预测未来 10 天的数据, 并且未来 10 天并没有节假日, 因此我们决定只有 2021 年 5 月份以后的数据来拟合模型。

3.2. 问题一的模型建立与评价

观察该 2021 年 5 月份以后数据的时序图、自相关图、偏自相关图, 见图 4、图 5、图 6:

2021年5月之后数据的15分钟电力负荷图

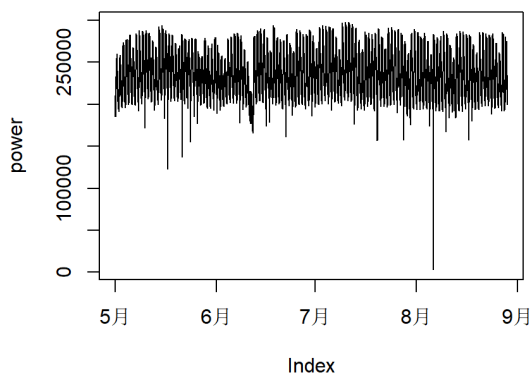


Figure 4. Time sequence diagram of some data

图 4. 部分数据时序图

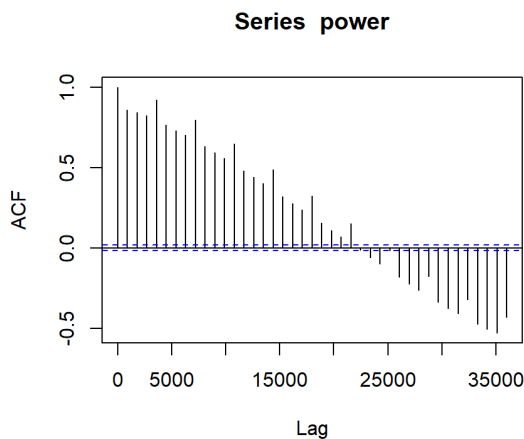


Figure 5. ACF diagram of partial data

图 5. 部分数据 ACF 图

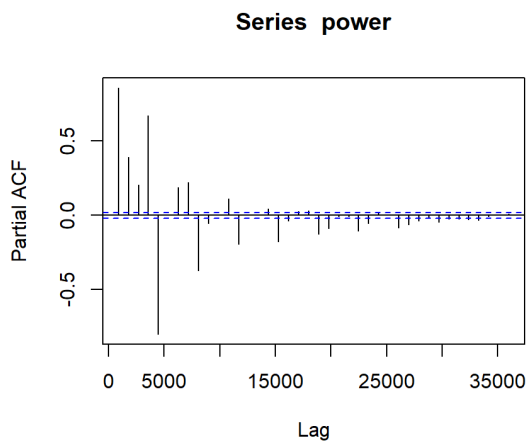


Figure 6. PACF diagram of partial data

图 6. 部分数据 PACF 图

从原始的时序图发现, 序列的周期性很强, 有明显的季节效应, 趋势变化并不是很明显。因此本文第一步先常数使用 ARIMA 模型。

3.2.1. ARIMA 模型

使用 R 语言的包 forecast 中的 auto.arima 函数拟合数据, 得到 ARIMA(3,0,1)模型。为了评价模型的准确率, 本文把 2021 年 8 月 15 日以后的数据作为测试集, 之前的作为训练集重新拟合模型, 得到模型 ARIMA(5,0,0), 模型的在测试集上的效果见表 1 (注 MAE、RMSE、MAPE 越小, 说明模型对该数据来说越好):

Table 1. Indicators of the ARIMA model test set

表 1. ARIMA 模型测试集的指标

MAE	RMSE	MAPE
22562.25	25787.213	9.973476

3.2.2. ARIMA 加法季节模型

由于数据有明显的季节性因素, 因此构建含有季节性的 ARIMA 模型。使用 R 语言的 ARIMA 函数拟合加法季节模型的 ARIMA(4,0,5), 其在测试集上的效果见表 2:

Table 2. Indicators of ARIMA additive seasonal model

表 2. ARIMA 加法季节模型的指标

MAE	RMSE	MAPE
1821.363	4878.357	1.147699

3.2.3. STL 分解模型

使用 R 语言的 STL 函数, 其中参数设置为 s.window = 96, t.window = 25, t.jump = 1。得到序列分解图如图 7:

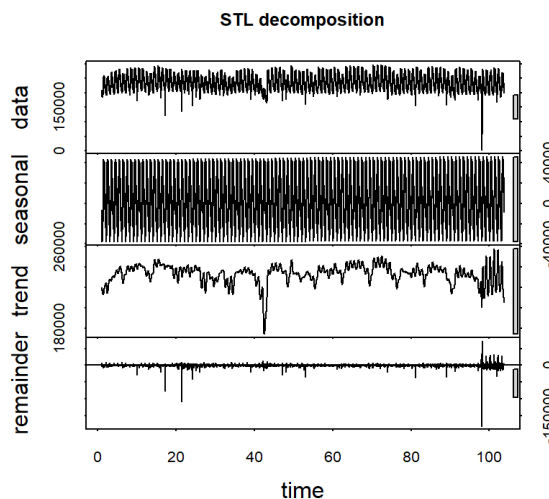


Figure 7. STL decomposition diagram

图 7. STL 分解图

模型预测的准确度[7]如表 3 所示:

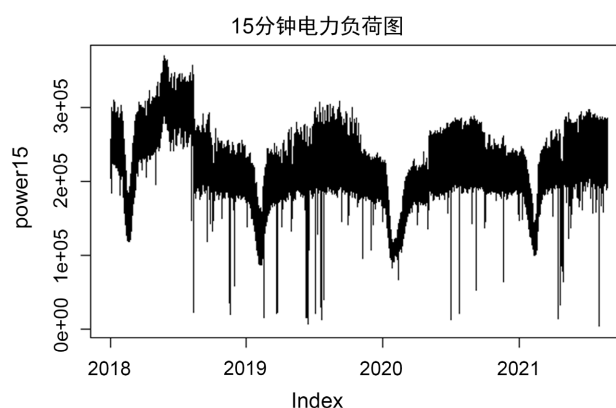
Table 3. Indicators of STL decomposition model**表 3.** STL 分解模型的指标

MAE	RMSE	MAPE
24283.51	31297.69	1.01619

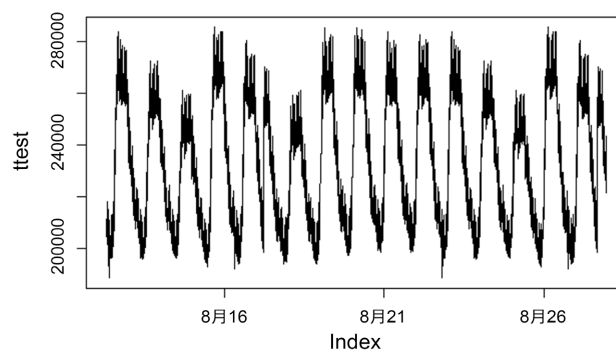
4. 基于机器学习算法的电力负荷的中短期预测

4.1. 模型数据的分析

经过对附件数据信息的预处理,对数据信息进行分析,首先有数据为 2018 年 1 月 1 号 0 点 0 分开始到 2021 年 8 月 31 号 23 点 45 分为止的每隔十五分钟的负荷数据,描述该地区在这三年多内用电量随着时间的变化,画出其整体电力负荷随时间变化如图 8 所示(横坐标为时间,纵坐标为电力负荷即总有功功率):

**Figure 8.** Power time series diagram**图 8.** 电力时序图

从图 8 中可以发现,除去 2018 年中旬的一次断崖式下降和其他年份的小量突变点之外,负荷随着时间的变化由很强的季节性,这也与实际十分相符,因为包括轻重工业在内的大部分工业在除每一天的工作时间、每一周的周一到周五以及法定工作日之外的休息时间其电力负荷都是非常小的,而这些工业用电是占据电力负荷的大头,因此进一步分析该时间序列的季节性主要包括哪些方面,我们分别画出来了该数据集最后十四天如图 9 所示(横坐标为时间,纵坐标为电力负荷划分的测试集),最后四个月的图像,如图 10 所示(横坐标为时间,纵坐标为电力负荷划分的训练集):

**Figure 9.** Time series for the last 14 days**图 9.** 最后 14 天时序图

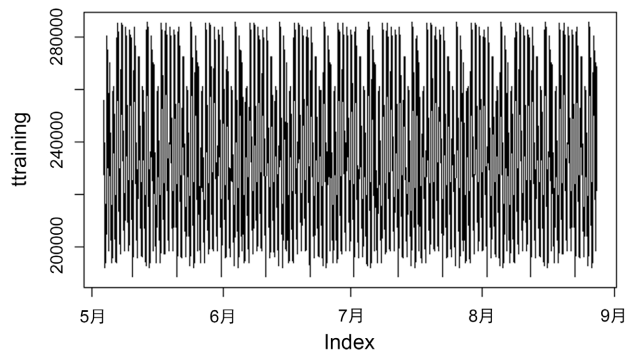


Figure 10. Time series diagram from May to August 2021

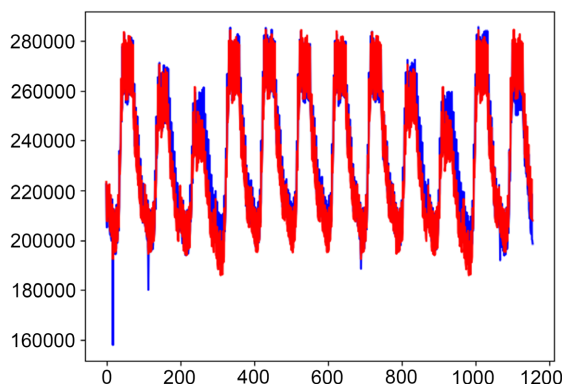
图 10. 2021 年 5~8 月时序图

由此可以看出, 该时间序列的季节性包括, 每天, 每周, 每月以及相应大型节假日, 因此, 为使得机器学习的算法能够识别这些信息, 我们需要生成天变量, 周变量, 月变量以及相应的大型节假日信息变量, 因为机器学习的算法是无法直接从我们的时间序列中去获取这些信息的, 需要把它们全不生成变量告诉它这些是负荷的影响因子, 是这些季节性变量使得负荷数据响应产生部分变化, 除此之外, 也应该把时间序列的其他影响因素也以变量的形式“告诉”机器学习算法, 例如时间序列中的趋势, 随机性等等, 把这些都生成变量, 使得我们利用机器学习的方法对数据的拟合能更加准确。这也是能在时间序列的预测使用机器学习的方法的一种方法。

4.2. 模型建立与评价

4.2.1. 随机森林模型

通过对时间信息的变量化处理, 我们得到了与该时间序列相关的日变量(以其十五分钟取样为周期的循环变量)、周变量(以周为周期的循环变量)、月变量(以月为周期的循环变量)后, 我们根据前面对整体时间序列的分析, 决定把训练集取 2019 年之后的数据集, 由于是短期十天的预测, 因此我们最终训练集只取了 2020 年 1 月之后的数据, 再以最后十天的数据作为验证集, 经过对随机森林模型里的一些调优参数的调整, 建立随机森林模型[8] [9]对数据进行拟合, 其最终的拟合效果如图 11 所示:



(注: 横坐标为最后十天电力负荷记录次数, 纵坐标为电力负荷总有功功率, 红色线为预测值, 蓝色线为真实值。)

Figure 11. Comparison of random forest test set predictions

图 11. 随机森林测试集预测对比图

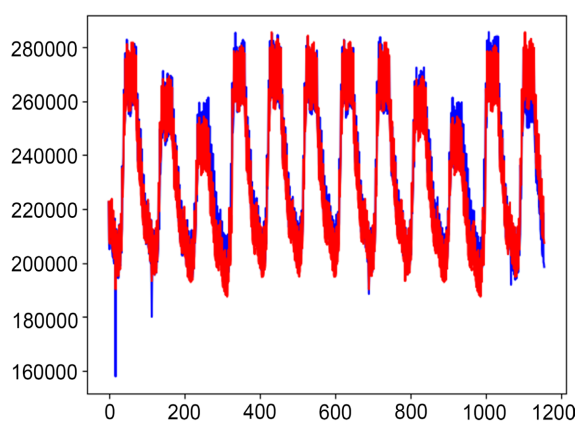
模型预测的准确度如表 4 所示:

Table 4. Model evaluation index**表 4.** 模型评价指标

MAE	RMSE	MAPE
3647.64	4936.00	0.0159

4.2.2. GBDT 模型

通过对时间信息的变量化处理, 我们得到了与该时间序列相关的日变量(以其十五分钟取样为周期的循环变量)、周变量(以周为周期的循环变量)、月变量(以月为周期的循环变量)后, 根据前面对整体时间序列的分析, 决定把训练集取 2019 年之后的数据集, 由于是短期十天的预测, 因此最终训练集只取了 2020 年 1 月之后的数据, 再以最后十天的数据作为验证集, 经过对 GBDT 模型[10] [11]里的一些调优参数的调整, 建立 GBDT 模型对数据进行拟合, 其最终的拟合效果如图 12 所示:



(注: 横坐标为最后十天电力负荷记录次数, 纵坐标为电力负荷总有功率, 红色线为预测值, 蓝色线为真实值。)

Figure 12. Comparison of GBDT test set predictions**图 12.** GBDT 测试集预测对比图

模型预测的准确度如表 5 所示:

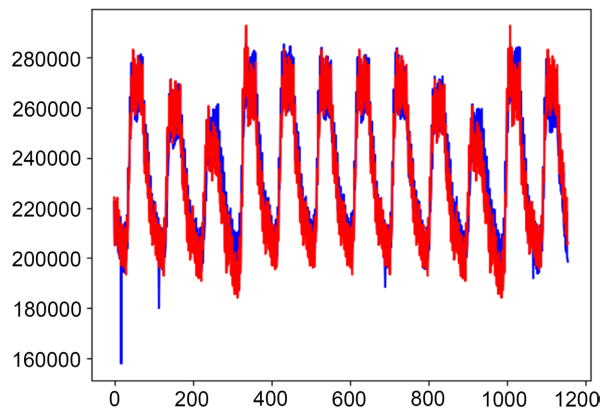
Table 5. Model evaluation index**表 5.** 模型评价指标

MAE	RMSE	MAPE
3351.45	4481.62	0.0147

4.2.3. XGBoost 模型

通过对时间信息的变量化处理, 我们得到了与该时间序列相关的日变量(以其十五分钟取样为周期的循环变量)、周变量(以周为周期的循环变量)、月变量(以月为周期的循环变量)后, 我们根据前面对整体时间序列的分析, 决定把训练集取 2019 年之后的数据集, 由于是短期十天的预测, 因此我们最终训练集只取了 2020 年 1 月之后的数据, 再以最后十天的数据作为验证集, 经过对 XGBoost 模型[12]里的一些调优参数的调整, 建立 XGBoost 模型对数据进行拟合, 其最终的拟合效果如图 13 所示。

模型预测的准确度如表 6 所示:



(注: 横坐标为最后十天电力负荷记录次数, 纵坐标为电力负荷总有功功率, 红色线为预测值, 蓝色线为真实值。)

Figure 13. Comparison of XGBoost test set predictions

图 13. XGBoost 测试集预测对比图

Table 6. Model evaluation index

表 6. 模型评价指标

MAE	RMSE	MAPE
4212.00	5415.24	0.0185

5. 结论与反思

本文采用了 ARIMA、机器学习算法对某地区未来十天间隔十五分钟用电负荷进行预测, 并通过预测值和观测值的 MAE、RMSE、MAPE 对精度进行了比较, 研究发现精度最好的为 GBDT 模型, 所以最终选取 GBDT 模型来进行后续的预测工作。复杂多变的气象条件和社会事件等不确定因素都会对电力系统负荷造成一定的影响, 使得传统负荷预测模型的应用存在一定的局限性。同时, 随着电力系统负荷结构的多元化, 模型应用的效果也有所降低, 因此电力系统负荷预测问题亟待进一步研究。

参考文献

- [1] Huang, S.R. (1997) Short-Term Load Forecasting Using Threshold Autoregressive Model. *IEE Proceedings - Generation, Transmission and Distribution*, **144**, 477-481. <https://doi.org/10.1049/ip-gtd:19971144>
- [2] Lee, C.M. and Ko, C.N. (2011) Short-Term Load Forecasting Using Lifting Scheme and ARIMA Models. *Expert Systems with Applications*, **38**, 5902-5911. <https://doi.org/10.1016/j.eswa.2010.11.033>
- [3] Chen, Y., Xu, P., Chu, Y., et al. (2017) Short-Term Electrical Load Forecasting Using the Support Vector Regression (SVR) Model to Calculate the Demand Response Baseline for Office Building. *Applied Energy*, **195**, 659-670. <https://doi.org/10.1016/j.energy.2017.06.076>
- [4] Lloyd, J.R. (2014) GEFCom2012 Hierarchical Load Forecasting Gradient Boosting Machines and Gaussian Processes. *International Journal of Forecasting*, **30**, 369-374. <https://doi.org/10.1016/j.ijforecast.2013.07.002>
- [5] Yu, Z., Niu, Z., Tang, W., et al. (2019) Deep Learning for Daily Peak Load Forecasting: A Novel Gated Recurrent Network Combining Dynamic Time Warping. *IEEE Access*, **7**, 17184-17194. <https://doi.org/10.1109/ACCESS.2019.2895604>
- [6] 王燕. 时间序列分析——基于 R [M]. 北京: 中国人民大学出版社, 2015.
- [7] Cleveland, R.B., Cleveland, W.S. and Terpenning, I. (1990) STL: A Seasonal-Trend Decomposition Procedure Based On Loess. *Journal of Official Statistics*, **6**.
- [8] 王宇涛, 钱妍竹. Python 大数据分析 with 机器学习 [M]. 北京: 机械工业出版社, 2020.

- [9] 丁伟熊. Python 机器学习核心算法编程实例[M]. 北京: 中国工信出版集团, 2019.
- [10] 张良均, 王路, 谭立云, 苏剑林. Python 数据分析与挖掘实战[M]. 北京: 机械工业出版社, 2015.
- [11] 游皓麟. Python 预测之美[M]. 北京: 电子工业出版社, 2020.
- [12] 朱江行, 邹晓松, 熊炜, 袁旭峰, 艾小清, 彭月. 基于 Prophet 与 XGBoost 混合模型的短期负荷预测[J]. 现代电力, 2021, 38(3): 325-331.