

基于LSTM和ARIMA的负荷预测对比分析

张瀚文*, 于长程, 李依婷, 宁世雄

中国矿业大学(北京)机电与信息工程学院, 北京

收稿日期: 2023年5月27日; 录用日期: 2023年6月22日; 发布日期: 2023年6月29日

摘要

针对负荷预测精度有待提升的问题, 基于不同原理进行负荷预测研究。首先进行了数据预处理, 其次形成了基于LSTM的负荷预测流程, 再次形成了基于ARIMA的负荷预测流程, 最后进行算例分析, 验证了基于LSTM的预测方法在预测该地区综合负荷方面具有更高的准确性。进一步地, 分析了两种方法的优缺点, ARIMA模型计算简单, 但对于复杂的非线性负荷变化难以进行准确预测; 而LSTM模型则能够更好地处理复杂的非线性负荷变化, 但对计算资源和训练时间要求较高。

关键词

负荷预测, LSTM模型, ARIMA模型, 预测误差

Comparative Analysis of Load Forecasting Based on LSTM and ARIMA

Hanwen Zhang*, Changcheng Yu, Yiting Li, Shixiong Ning

School of Mechanical Electronic & Information Engineering, China University of Mining & Technology-Beijing, Beijing

Received: May 27th, 2023; accepted: Jun. 22nd, 2023; published: Jun. 29th, 2023

Abstract

To address the issue of improving the accuracy of load forecasting, research on load forecasting based on different principles was conducted. First, data preprocessing was performed. Secondly, a load forecasting process based on LSTM was formed, followed by a load forecasting process based on ARIMA. Finally, a case analysis was conducted, which verified that the LSTM-based prediction method has higher accuracy in predicting the comprehensive load in the region. Furthermore, the advantages and disadvantages of the two methods were analyzed. The ARIMA model has a simple calculation, but it is difficult to accurately predict complex nonlinear load changes. On the other hand, the LSTM model can better handle complex nonlinear load changes, but it requires high

*通讯作者。

computing resources and training time.

Keywords

Load Forecasting, LSTM Model, ARIMA Model, Prediction Error

Copyright © 2023 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

负荷预测是保证电力供需平衡的基础[1]。目前我国的负荷预测方法已经呈现多样化[2]，其中一种常用方法是根据电力负荷历史数据来推测未来数值，充满了不确定性[3] [4]。这种基于历史数据的负荷预测最主要的问题就是解决预测模型的精度问题，如何消除预测模型里的不确定因素，尽可能精确地预测电力负荷的大小。

近些年来，随着科学技术的迅速发展，预测的方法与理论知识不可同日而语，随着人工智能的普及，新的预测方法层出不穷，为电力负荷预测研究提供了有力的工具。具体来说，中国电力企业已经采用了各种预测方法和技术，包括时间序列分析[5]、神经网络[6]、支持向量机[7]、遗传算法[8]等。这些方法和技术能够有效地预测不同地区的负荷需求，提高电力调配的准确性和效率。此外，中国电力企业还积极推广智能电网技术[9] [10]，通过实时监测和分析用电数据，预测未来用电趋势，并根据预测结果进行电力调配，从而提高电力供给的质量和稳定性。总的来说，中国电力企业在区域负荷预测方面已经取得了显著的进展，不断引入新的技术和方法，以适应电力市场和用电需求的不断变化。

在此基础上，本文将重点放在 LSTM 以及 ARIMA 两种方法上，将 ARIMA 作为对照组，LSTM 作为实验组，对比两种预测方法的优缺点。

2. 原理

2.1. 数据预处理

对收集到的数据进行归一化的预处理，可以防止数据输入到 LSTM 中导致神经元饱和，影响预测网络的准确性。其归一化表达式为：

$$Y_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

式中 Y 表示原始数据归一化后的数据； X 表示原始输入数据； X_{\min} 表示原始数据中的最小值； X_{\max} 表示原始数据中的最大值。

2.2. LSTM 神经网络的模型原理

LSTM 是一种基于时间序列的长短期记忆网络。LSTM 神经网络中能够通过更新或移除来实现选择通过信息的节点称为“门”结构，神经网络中包含三种门结构：输入门，输出门，遗忘门。输入门决定哪些信息可以进入细胞，遗忘门决定每一时刻是否有信息被遗忘，而输出门则决定哪些信息是有效的并被输出。其基本的结构原理图见图 1：

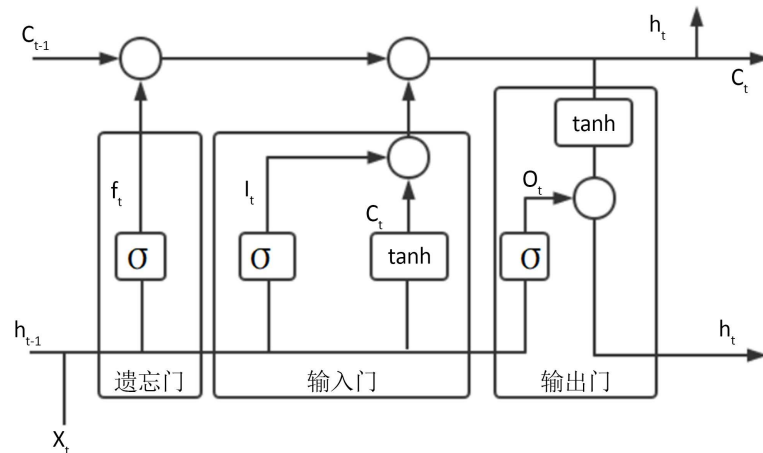


Figure 1. Schematic diagram of LSTM model
图 1. LSTM 模型原理图

前一个记忆单元 C_{t-1} 由遗忘门 f_t 控制保留至当前记忆单元 C_t 的信息量；输入 X_t 由输入门 i_t 决定保留至当前记忆单元 C_t 的信息量；当前记忆单元 C_t 由输出门 O_t 控制输出到 h_t 的信息量。LSTM 的公式如下：

$$\begin{aligned}
 \text{记忆单元: } C_t &= \tanh(W_c [h_{t-1}, x_t] + b_c) \\
 C_t &= f_t C_{t-1} + i_t C_t \\
 \text{遗忘门: } f_t &= \sigma(W_f [h_{t-1}, x_t] + b_f) \\
 \text{输入门: } i_t &= \sigma(W_i [h_{t-1}, x_t] + b_i) \\
 \text{输出门: } O_t &= \sigma(W_o [h_{t-1}, x_t] + b_o) \\
 \text{最终输出: } h_t &= O_t \tanh(C_t)
 \end{aligned} \tag{2}$$

2.3. ARIMA 预测原理

ARIMA (Autoregressive Integrate Moving Average) 是一种广泛应用于时间序列分析中的统计模型，用于预测未来的趋势和波动。ARIMA 模型基于对时间序列数据的分析，包括分析趋势、季节性和随机性等特征，以及进行差分和平滑处理，以便将时间序列转化为平稳的序列。ARIMA 模型在电力预测中应用广泛，可以用于预测电力负荷、电价、电量等指标。

ARIMA 模型由三个参数来描述时间序列，分别是 p 、 d 、 q ，其中：

p 表示自回归(AR)项，表示当前时刻的值与之前 p 个时刻的值之间的关系；

d 表示差分(I)项，表示需要进行几次差分才能使序列平稳；

q 表示移动平均(MA)项，表示当前时刻的值与之前 q 个时刻的噪声项之间的关系。

3. 负荷预测过程

3.1. LSTM 预测流程

预测需要多方面验证 LSTM 模型预测的准确性，故采用多组对照实验。将不同时间尺度实验组的预测结果，以及同一时间尺度下用 LSTM 方法、ARIMA 两种方法进行预测的结果分别进行对照分析。预测步骤如下：

1) 导入某地区某年的区域负荷数据样本数据集，对样本数据集进行预处理。预处理包括归一化处理和划分训练集。根据不同的时间分度，将每组数据按 6:2:2 划分训练集、测试集和验证集；

2) 搭建 LSTM 神经网络模型。设置神经元核心的个数, 迭代次数, 优化器等。输入训练集训练 LSTM 网络, 通过对比分析预测结果, 调整模型参数;

3) 向训练好的 LSTM 神经网络输入测试集, 预测下一个时间段的用电负荷, 以此类推, 实现滚动预测。选取合适的评价指标的预测结果进行分析。

3.2. ARIMA 预测流程

ARIMA 模型的搭建过程通常可以分为以下几个步骤:

- 1) 数据准备: 收集并准备好时间序列数据, 确保数据完整、连续、准确。
- 2) 数据预处理: 对原始数据进行必要的预处理, 如填充缺失值、一阶差分和二阶差分等。
- 3) 确定 ARIMA 模型阶数: 通过分析时间序列数据的未平稳与一阶差分后的自相关函数 ACF 和偏自相关函数 PACF, 并根据 AIC、BIC 准则, 以及对残差序列的分析, 选择合适的 ARIMA 模型阶数。
- 4) 模型训练: 使用选定的 ARIMA 模型阶数, 对数据进行训练。
- 5) 模型检验: 对训练出来的模型进行检验, 包括检查残差序列的正态性和自相关性等, 以确保模型具有良好的性质和可靠性。
- 6) 模型预测: 使用训练好的 ARIMA 模型进行预测, 可以预测未来某一时刻或一段时间内的数据。
- 7) 模型评估: 对模型预测结果进行评估, 比较预测值与实际值之间的误差, 根据 MAE、RMSE 两个评价指标对模型进行优化和改进。

3.3. 评估指标

为了评估模型性能, 本文采用平均绝对误差(MAE), 均方根误差(RMSE)。数学表达式如式(3)~(4)所示。

$$MAE = \frac{\sum_{g=1}^{n_0} |\hat{Y}_g - Y_g|}{n_0} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{g=1}^{n_0} (\hat{Y}_g - Y_g)^2}{n_0}} \quad (4)$$

4. 负荷预测案例分析

4.1. 数据集处理

本文收集的是 A 地区 2016 年一年的用电负荷量, 采样间隔为 15 分钟, 共 35,316 个数据。在归一化处理, 根据不同的时间尺度, 分为三组, 分别选取一年、一个月、一天的数据, 划分前 60% 的数据为训练集, 中间 20% 的数据为测试集, 末尾 20% 的数据为验证集。

4.2. LSTM 模型搭建过程

4.2.1. 训练模型

初始隐态向量设置为 16, LSTM 单元分别设置为 32 和 64, 每次使用 16 个自变量来预测 1 个因变量, 输出维度为设置为 2, 迭代次数设置为 100, dropout 参数为 0.1、patience 参数为 20。设定后将训练集放入 LSTM 中进行训练, 并加入验证集进行验证, 得出符合数据特点模型。

4.2.2. 停止训练的条件

停止训练有两种情况: 训练过程中达到 epoch 最大值或满足符合提前停止的次数上限时, 该训练则

停止，模型训练完成。

4.2.3. 参数设置

参数设置见表 1。

Table 1. LSTM model parameter settings
表 1. LSTM 模型参数设置

参数	层数	1	2
units		32	64
input size		1	1
batch size		16	16
dropout		0.1	0.1
patience		20	20
epoch		100	100

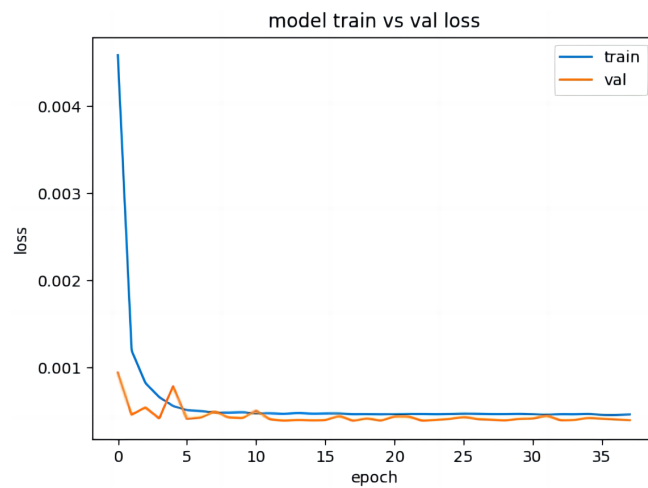


Figure 2. Annual forecast loss graph
图 2. 年预测的损失曲线图

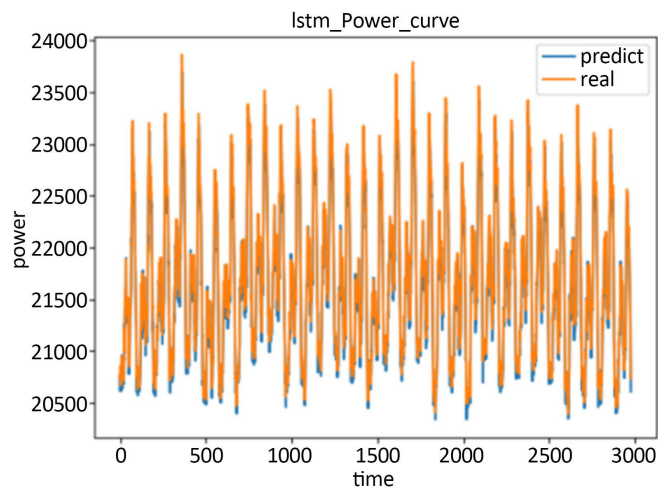


Figure 3. The annual predicted value versus the true value (MW)
图 3. 年预测值和真实值对比图(MW)

4.3. 结果分析

年预测组的训练集和验证集损失曲线见图 2，预测值和真实值对比图见图 3。

月预测组的最终训练损失曲线见图 4，真实值与预测值对比图见图 5。

日预测组的最终训练损失曲线见图 6，绘制真实值与拟合值对比图见图 7。

电力负荷预测是具有高度随机性和变化性的数据，比较上图可得，LSTM 模型在这三组不同时间尺度的预测中，预测值和真实值的曲线接近，预测效果良好，三组训练集和预测集的曲线相关性均较强，验证了 LSTM 预测方法的准确性。所收集负荷数据平均值为 20,635.11 MW，由评估指标 MAE 和 RMSE 可以得出真实值和预测值偏离的相对误差均在 3% 以下，可以用于本案例负荷预测。其中年预测和月预测的准确性更高，可见 LSTM 更适合长期预测。

三组预测模型的测试的准确性见表 2：

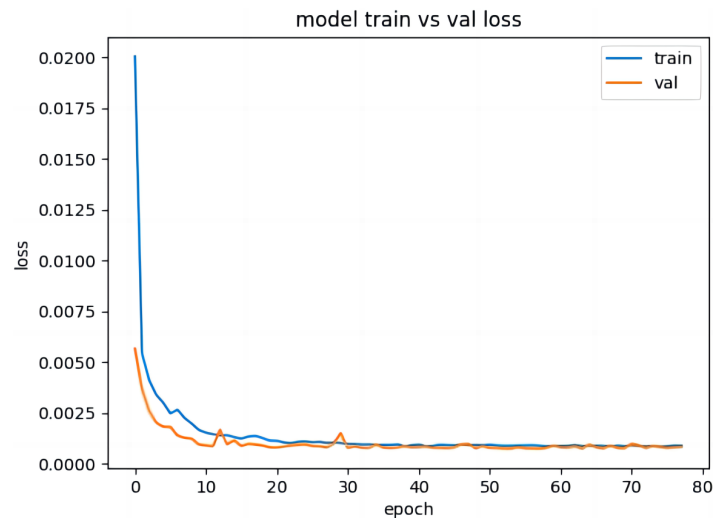


Figure 4. Monthly forecast loss graph

图 4. 月预测的损失曲线图

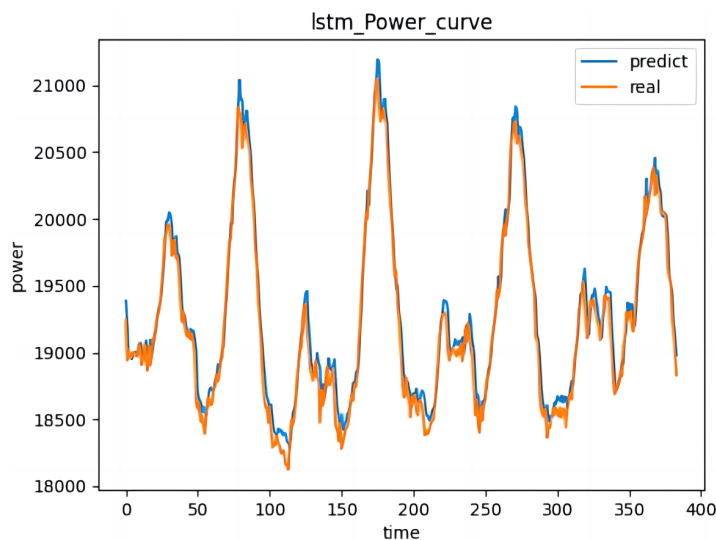


Figure 5. Graph of the predicted and true monthly values (MW)

图 5. 月预测值和真实值对比图(MW)

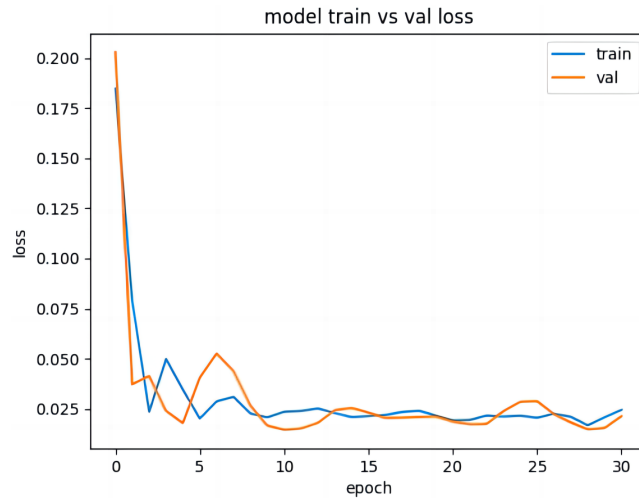


Figure 6. Daily forecast loss graph
图 6. 日预测的损失曲线图

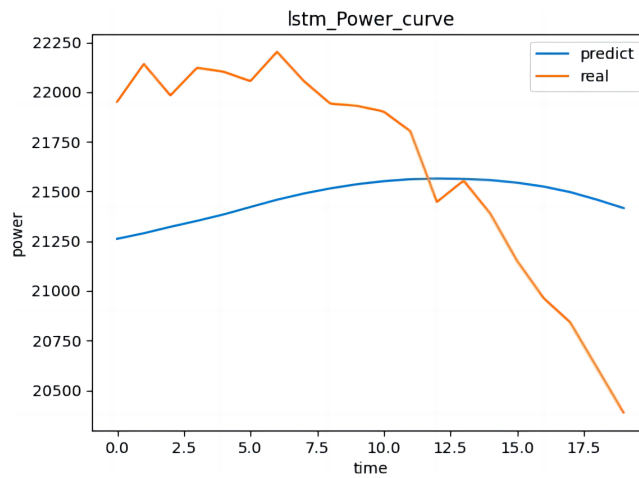


Figure 7. Daily predicted value versus true value (MW)
图 7. 日预测值和真实值对比图(MW)

Table 2. Prediction evaluation index (MW)
表 2. 预测评估指标(MW)

指标	年预测	月预测	日预测
MAE	95.18	89.55	555.79
RMSE	120.68	116.71	618.67

4.4. ARIMA 模型搭建过程

4.4.1. 数据预处理

对原始数据进行填充缺失值处理,采用 $n = 2$ 的 n 最近邻均值法填充,并对填充缺失值后的数据进行一阶差分与二阶差分。

4.4.2. 确定 ARIMA 模型阶数

通过分析时间序列数据的未平稳与一阶差分后的自相关函数 ACF 和偏自相关函数 PACF,并结合 AIC、

BIC 准则，确定 ARIMA 模型的 p 、 q 参数， $p = 4$ 、 $q = 0$ 。对原数据与一阶差分后的数据进行 ADF 检验，得到 $p < 0.05$ 小于显著性水平，我们即可认为序列具有平稳性，则确定 ARIMA 模型的 $d = 1$ 。即 $(p,d,q) = (4,1,0)$ 。划分训练集、测试集和验证集同 LSTM 模型。

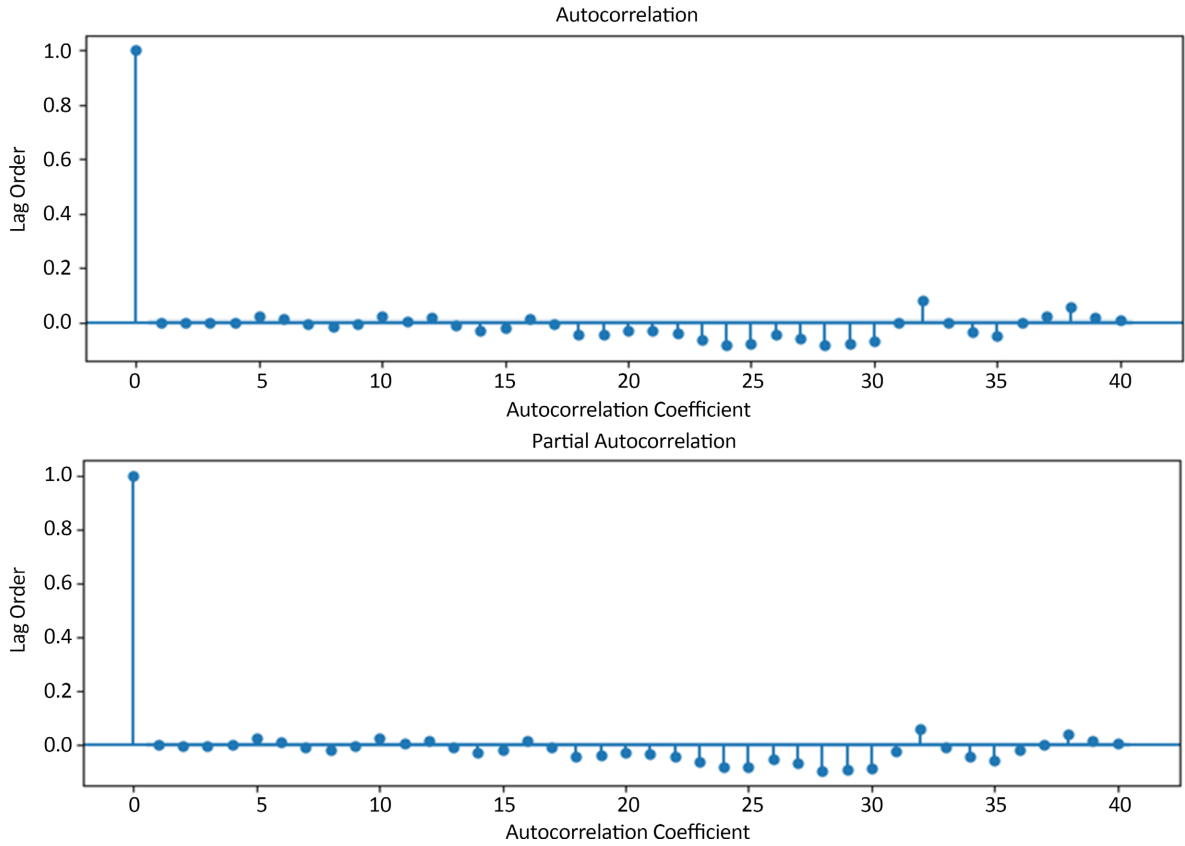


Figure 8. Residual ACF and PACF plots

图 8. 残差 ACF 与 PACF 图

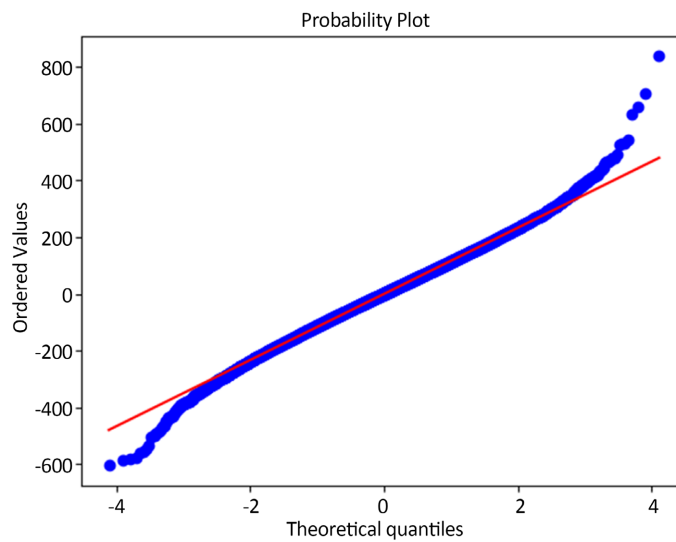


Figure 9. Residual normality test QQ graph

图 9. 残差正态性检验 QQ 图

4.4.3. 模型检验

在 ARIMA 负荷预测中,进行了多种检验,有效地对模型进行评估和选择,提高预测的准确性和可靠性。其中, $pvalue = 4.07974$, DW 检验结果为 2.00187。残差 ACF 与 PACF 图见图 8, 残差正态性检验 QQ 图见图 9, 残差正态性检验直方图见图 10。这些检验从不同角度评估了模型的合适性, 确保了模型的稳定性、可靠性和准确性。

4.4.4. 模型预测

原数据与预测数据对比图见图 11:

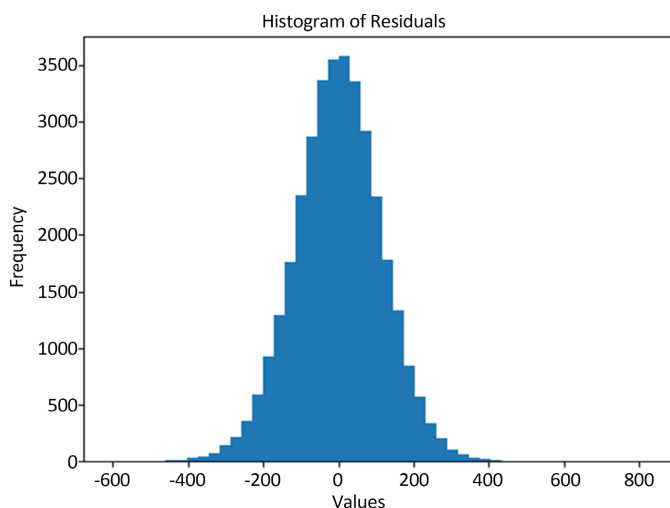


Figure 10. Histogram of residual normality test

图 10. 残差正态性检验直方图

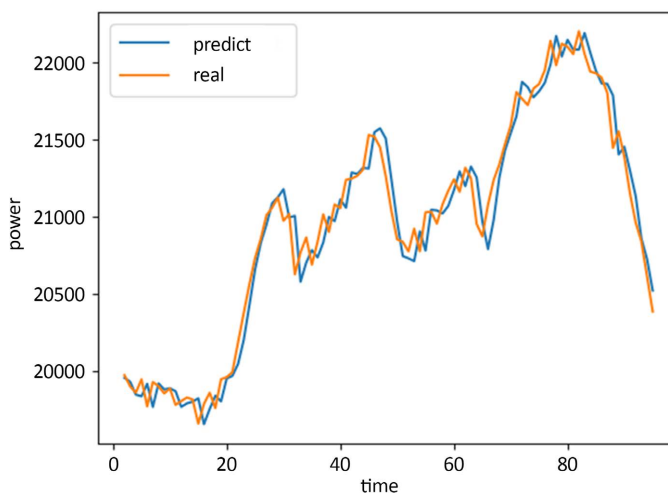


Figure 11. Graph comparing original data with predicted data (MW)

图 11. 原数据与预测数据对比图(MW)

从原数据与预测数据对比图看出 ARIMA 模型拟合效果较好, 但有些值偏高或偏低。

4.4.5. 模型评估

RMSE (Root Mean Squared Error): 均方根误差, 是 MSE 的平方根。RMSE 越小, 说明预测结果越准确。

MAE (Mean Absolute Error): 平均绝对误差, 是预测值与实际值之差的绝对值的平均值。MAE 越小, 说明预测结果越准确。

最终由 ARIMA 模型拟合出的均方根误差 RMSE = 1117.10 MW, 平均绝对误差 MAE = 890.12 MW。LSTM 模型的 MAE 为 95.18 MW, RMSE 为 120.68 MW, 而 ARIMA 模型的 MAE 为 890.12 MW, RMSE 为 1117.10 MW, 说明 LSTM 模型在预测误差方面表现更优秀。因此, 在实际应用中, 可以优先考虑使用 LSTM 模型进行电力负荷预测, 同时也需要根据实际情况进行选择 and 调整。

5. 结语

本文基于 LSTM 和 ARIMA 模型, 形成了相应预测流程, 根据区域负荷历史数据进行预测。区域负荷预测案例分析结果表明, 在均方根误差和平均绝对误差两项指标中, LSTM 通常比 ARIMA 的预测更精确, 但 ARIMA 模型具有更强的可解释性, 只需要进行参数估计和预测就能够清楚地展示趋势和周期性, LSTM 模型的预测结果则需要进一步解释其内部神经网络结构的作用且 LSTM 的计算复杂度更高, 需要大量计算资源和时间。综合来说, 应根据具体应用场景选择合适的模型: 对于简单的平稳时间序列数据, ARIMA 更适合; 对于复杂的非线性时间序列数据, LSTM 更具优势。

基金项目

中央高校基本科研业务费专项资金; 中国矿业大学(北京)大学生创新训练项目(202204020)。

参考文献

- [1] 康重庆, 夏清, 张伯明. 电力系统负荷预测研究综述与发展方向的探讨[J]. 电力系统自动化, 2004(17): 1-11.
- [2] 廖旒焕, 胡智宏, 马莹莹, 卢王允. 电力系统短期负荷预测方法综述[J]. 电力系统保护与控制, 2011, 39(1): 147-152.
- [3] 郑志杰, 李磊, 赵兰明. 考虑数据不确定性的中长期电力负荷预测[J]. 电力系统保护与控制, 2011, 39(7): 123-126+132.
- [4] 邓永生, 焦丰顺, 张瑞锋, 王将平. 配电网规划中电力负荷预测方法研究综述[J]. 电器与能效管理技术, 2019(14): 1-7.
- [5] 王斌. 基于时间序列 ARIMA 模型的电力负荷短期预测分析[J]. 石河子科技, 2019(3): 43-47.
- [6] 陆普凡. 基于改进神经网络的电力负荷短期预测方法研究[D]: [硕士学位论文]. 南京: 南京邮电大学, 2022.
- [7] 叶远胜, 张静. 基于时间序列的 SVM 短时电力负荷预测[J]. 现代信息科技, 2020, 4(24): 17-19.
- [8] 刘鹏飞, 付波. 基于遗传算法优化 BP 神经网络的短期热负荷预测方法研究[J]. 区域供热, 2019(1): 43-48+76.
- [9] 周孝信, 陈树勇, 鲁宗相, 黄彦浩, 马士聪, 赵强. 能源转型中我国新一代电力系统的技术特征[J]. 中国电机工程学报, 2018, 38(7): 1893-1904+2205.
- [10] 董朝阳, 赵俊华, 文福拴, 薛禹胜. 从智能电网到能源互联网: 基本概念与研究框架[J]. 电力系统自动化, 2014, 38(15): 1-11.