

基于最优特征选择和“ABL+XGB”的电力负荷混合预测方法

刘康宁¹, 代业明¹, 曹书宁²

(1. 青岛大学商学院, 青岛 260001; 2. 山东省高密市住房和城乡建设局, 高密 261502)

摘要: 精准的电力负荷预测是保障电网安全与经济调度的关键。然而, 预测过程受时间、天气、经济等多重因素影响, 高维数据和冗余信息使预测过程复杂化, 降低了预测精度。为此, 提出一种基于最优特征选择与“Attention-Bi-LSTM + XGBoost”(“ABL+XGB”)的混合预测方法。该方法综合考虑 27 种影响因素, 通过对比 6 种特征选择方法, 筛选出最优特征子集, 并采用“ABL+XGB”组合模型进行预测。基于新加坡与挪威电力市场实际数据集的实验结果表明, 所提方法具有较高的预测精度, Lasso 回归结合组合预测模型结果最接近真实负荷且误差最小, 验证了有效的特征选择对于提升电力负荷预测准确性的重要作用。

关键词: 电力负荷预测; 特征选择; 混合预测; Lasso 回归

中图分类号: TM 73 **文献标志码:** A

Power load hybrid forecasting method based on optimal feature selection and "ABL+XGB"

LIU Kangning¹, DAI Yeming¹, CAO Shuning²

(1. School of Business, Qingdao University, Qingdao 260001, China; 2. Gaomi Housing and Urban-rural Construction Bureau, Gaomi 261502, China)

Abstract: Accurate power load forecasting is crucial to grid security and economic dispatch. However, the forecasting process is influenced by multiple factors including temporal, meteorological, and economic variables. This leads to high-dimensional data with significant redundancy, which increases complexity and reduces forecasting precision. To address this issue, a hybrid forecasting method that integrates optimal feature selection with a combined model of Attention-Bi-LSTM and XGBoost was

收稿日期: 2025-01-04

基金项目: 国家自然科学基金资助项目 (72371139); 教育部人文社会科学规划项目 (20YJA630009); 山东省自然科学基金资助项目 (ZR2022MG002)

第一作者: 刘康宁 (2002—), 女, 硕士研究生。研究方向: 电力市场电价预测。E-mail: L18863978132@163.com

通信作者: 代业明 (1975—), 男, 教授。研究方向: 智能电网系统优化、人工智能。E-mail: yemingdai@163.com

引文格式: 刘康宁, 代业明, 曹书宁. 基于最优特征选择和“ABL+XGB”的电力负荷混合预测方法[J]. 上海理工大学学报, 2026, 48(2): 183-193.

Citation: LIU Kangning, DAI Yeming, CAO Shuning. Power load hybrid forecasting method based on optimal feature selection and "ABL+XGB"[J]. Journal of University of Shanghai for Science and Technology, 2026, 48(2): 183-193.

proposed, referred to as the "ABL+XGB" model. This approach considered 27 influencing factors, employed 6 feature selection methods to identify the optimal feature subset, and utilized the proposed "ABL-XGB" model for prediction. Experimental results based on datasets from the Singapore and Norway electricity markets demonstrate that the proposed hybrid forecasting model has a high accuracy, and the prediction results of Lasso regression method combined with the prediction model are closer to the real data and the error is smaller than other feature selection methods, which indicates that a good feature selection method is essential for achieving accurate power load prediction.

Keywords: *power load forecasting; feature selection; hybrid forecasting; Lasso regression*

电力负荷预测作为电力系统规划和运营的关键环节,对于确保电力供应的可靠性、优化电力资源的利用等方面具有重要意义。现有的电力负荷预测方法主要分为传统统计学方法和机器学习方法。其中,传统统计学方法^[1-2]主要依赖于预定义的数学假设和线性时间序列模型,常见的方法包括多元线性回归^[3]、卡尔曼滤波^[4]和自回归积分移动平均模型^[5]等。这些方法虽然拟合速度快,但往往依赖于经验模型和统计分析,对于日益增长的负荷需求,难以满足其精准预测和精细化管理要求。而机器学习方法^[6-7]主要包括支持向量机(support vector machine, SVM)^[8]、随机森林(random forest, RF)^[9]、极端梯度提升(XGBoost)和K最近邻(K-nearest neighbors, KNN)^[10]等,已广泛应用于中短期电力负荷预测。这些方法虽然能够学习数据中的非线性关系,但难以有效捕捉电力负荷数据的时间依赖性和长期趋势。为了解决机器学习模型的局限性,深度学习技术越来越多地用于电力负荷预测,这是因为它们能够直接从原始时间序列数据中学习分层特征表示,包括长短期记忆(long short-term memory, LSTM)网络^[11]、门控递归单元(gate recurrent unit, GRU)^[12]、卷积神经网络(convolutional neural networks, CNN)^[13]和注意力机制^[13]等。这些方法以其强大的数据分析处理和捕捉数据特征的能力,为电力系统的规划和运营提供更科学、更精确的决策支持,但仍存在一定的局限性。例如,模型通常依赖于大量的数据输入,并且电力负荷数据受到多种复杂因素的影响,这些单一模型难以有效捕捉复杂的非线性关系。因此,开展改进深度学习方法以精确预测电力负荷的研究,对于电力系统的安全、稳定和经济运行具有重要意义,同时也有助于推动能源转型和促进深度学习技术的创新与应用。

在电力系统运行过程中,随着数据体量的增加,数据高维性问题日益显著,如何从海量数据中提取出与负荷需求相关的有效特征是电力负荷预测的关键环节之一,而特征选择是减少维数灾难的有效方法^[14]。在进行电力负荷预测前,通过特征选择方法剔除数据的集中冗余和不相关特征,保留具有相关属性的子集,可以有效减少计算时间,提高预测性能。当前的特征选择方法可分为过滤型、封装型和嵌入型^[15-16],不同的特征选择方法具有各自的特点。过滤型方法通过计算各个特征与目标的相关度,选择与目标相关性较高的特征进行模型训练。皮尔逊相关系数(Pearson correlation coefficient, PC)^[17]和最大相关-最小冗余(max-relevance and min-redundancy, mRMR)^[18]是两种典型的过滤型方法,被广泛用于可再生能源发电预测^[19]、充电负荷估计^[20]、电动汽车充电管理^[21]、负荷预测^[22]等方面,减少了计算时间,提高了预测精度。然而,这种方法在特征子集选择时通常不考虑后续的预测模型,可能导致预测模型的拟合能力较低^[23]。相比之下,封装型方法直接使用预测误差来评估选定的特征子集,并在特征选择过程中使用每个子集来训练新的预测模型。递归特征消除(recursive feature elimination, RFE)是一种常见的封装型方法。虽然封装型方法能够改善模型的预测性能,但也伴随着计算时间和复杂性的增加^[24]。而嵌入型方法综合了过滤型和封装型方法的优点,将特征选择过程与预测模型训练融为一体,通过预测模型的参数估计来实现特征的自动选择。Lasso回归方法、RF^[25]和XGBoost等均是常见的嵌入型方法^[26],在电力负荷预测^[27]、终端区容量预测^[28]、语言识别^[29]、生物医学^[30]等领域取得了显著成果。这些方法将特征选择任务交给预测模型学习,更省时省力,但

会增加预测模型的计算复杂性和训练负担^[31]。

近年来将数据进行特征选择并与多种预测模型进行混合预测的方法开始出现^[32], 但通常以一种特征选择算法结合单一预测模型进行预测, 少有研究分析不同特征选择算法对预测结果精确度的改进程度, 且在能源领域应用较少。因此, 本文将影响电力负荷的因素增加到 27 个, 并将 PC、mRMR、RFE、Lasso 回归、RF 和 XGBoost 这 6 种常见且效果良好的特征选择方法分别与作者团队前期提出的“Attention-Bi-LSTM + XGBoost”^[32] (“ABL+XGB”)组合预测模型结合进行电力负荷混合预测。同时, 通过新加坡和挪威两个电力市场数据进行验证, 探讨不同特征选择算法对模型预测结果产生的影响。最终通过误差比较选出性能最佳的一种特征选择方法, 从而得到一种新的电力负荷混合预测方法。

1 “ABL+XGB”组合预测模型

Attention-Bi-LSTM 模型充分考虑输入数据中隐藏的整体信息并从两个维度获得更全面结果, 强调关键输入数据的影响; 而 XGBoost 模型具有较低的复杂度并能防止过度拟合。因此, 将 Attention-Bi-LSTM 模型与 XGBoost 模型基于误差倒数法进行组合得到“ABL+XGB”组合预测模型, 克服了单一预测模型的固有缺陷, 模型框架如图 1 所示。图中: x_t 表示第 t 个时间步的输入特征向量; e_t 为 x_t 经嵌入层变换后的特征表示; $\vec{h}_t$ 和 $\overleftarrow{h}_t$ 分别表示前向 LSTM 与后向 LSTM 在第 t 个时间步的隐藏状态, 将 $\vec{h}_t$ 和 $\overleftarrow{h}_t$ 拼接后得到双向隐藏状态 h_t 作为注意力层的输入。

具体预测步骤如下:

a. 采用 Attention-Bi-LSTM 和 XGBoost 模型预测电力需求, 并获得相应误差。

b. 根据误差结果, 通过误差倒数法计算 Attention-Bi-LSTM 和 XGBoost 两种模型的权重, 使误差较小的模型权重更大, 从而最大限度发挥模型的优势, 减少组合预测模型的总体误差。

采用误差倒数法为 Attention-Bi-LSTM 和 XGBoost 两种模型分配权重的公式如下:

$$f_i = \omega_1 f_{1i} + \omega_2 f_{2i}, \quad i = 1, 2, \dots, n \quad (1)$$

$$\omega_1 = \frac{\xi_2}{\xi_1 + \xi_2} \quad (2)$$

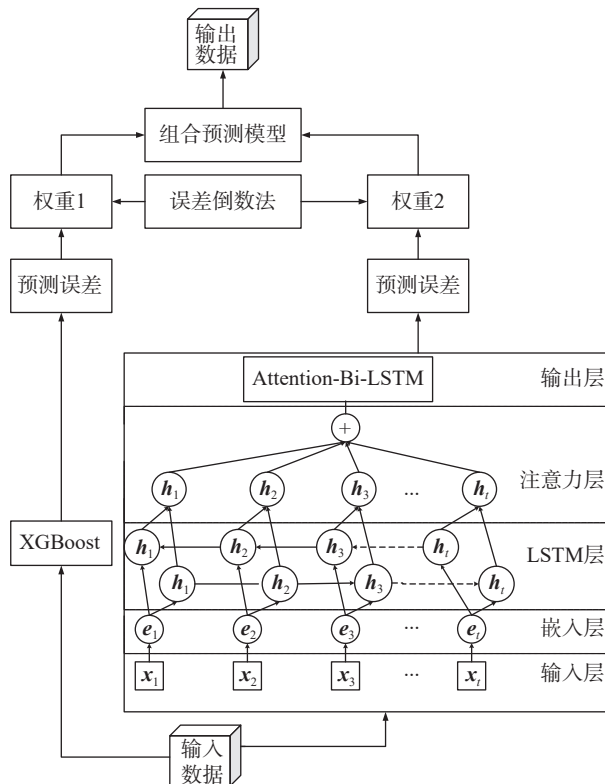


图 1 “ABL+XGB”预测模型框架

Fig.1 “ABL+XGB” prediction model framework

$$\omega_2 = \frac{\xi_1}{\xi_1 + \xi_2} \quad (3)$$

式中: ω_i 为权重值; f_i 为 Attention-Bi-LSTM 和 XGBoost 所得出的预测值; ξ_1 、 ξ_2 分别为预测模型 Attention-Bi-LSTM 和 XGBoost 的误差值。

c. 将 Attention-Bi-LSTM 和 XGBoost 两种模型的不同预测结果进行组合。

2 基于最优特征选择的“ABL+XGB”混合预测方法

为探究 6 种特征选择方法的优劣, 本文研究影响电力负荷的因素, 并将特征选择后的数据代入“ABL+XGB”预测模型中进行误差分析, 对比选出最优的特征选择方法。本文所提出的基于最优特征选择的混合预测方法如图 2 所示, 预测步骤主要分为 5 个阶段。

Step 1 特征输入。选择对电力负荷影响较大的特征共 27 个, 以验证输入特征数量的增加对模型预测结果的影响, 并对特征选择方法进行评估。

Step 2 特征选择。分别使用 PC、mRMR、RFE、Lasso 回归、RF、XGBoost 6 种特征选择方法对 27 个影响因素进行特征选择, 以消除冗余信

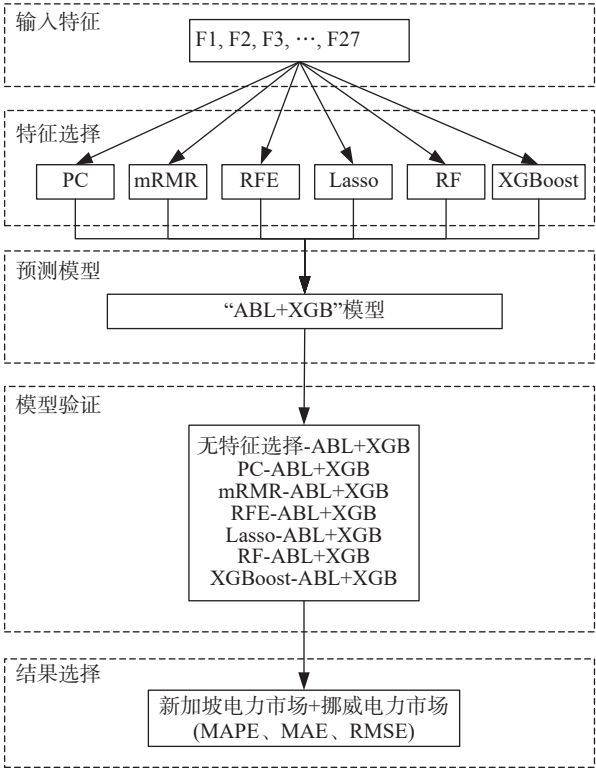


图2 基于最优特征选择的混合预测方法框架图

Fig.2 Frame diagram of hybrid prediction method based on optimal feature selection

息，提取出相关性较大的特征。

Step 3 预测模型。特征选择后的结果结合预测模型得出的误差可以分析出特征选择方法的性能，采用“ABL + XGB”预测模型对上述 6 种特征选择方法和不进行特征选择的模型进行结果比较和误差分析，从而选出性能最佳的特征选择方法。

Step 4 模型验证。分别将经过 PC、mRMR、RFE、Lasso 回归、RF 和 XGBoost 6 种特征选择方法选择后的数据分别代入“ABL + XGB”模型，同时以不进行特征选择的预测结果为对照，进行

模型验证。

Step 5 结果选择。比较基于预测模型的 6 种特征选择方法的预测结果，结合新加坡和挪威两个电力市场数据集，选择出最优的特征选择方法。

3 预测评价

3.1 数据集和实验环境

筛选几个国家电力市场的数据后发现，新加坡和挪威电力市场的数据包含了本文研究需要的所有影响因素，数据完整性程度高，更适合本文提出的模型。这些数据包括 2019 年 1 月 1 日—2020 年 12 月 31 日的新加坡数据和 2020 年 1 月 1 日—2021 年 12 月 31 日的挪威数据。因此，本文使用上述两个电力市场的数据来验证和评估该混合预测方法，并考虑了日类型、时刻、节假日类型、实时电价等因素。电力负荷历史数据的采样周期为 1 h。首先，使用加权灰色关联投影算法 (WGRP)对节假日数据进行优化，选择与当前节假日具有高相似性的历史数据替换异常值，使数据更具一般性；其次，对输入数据进行标准化处理，消除量纲差异；然后，按照时间顺序，以 8 : 2 的比例划分训练集和测试集；最后，将计算结果与实际数据进行比较，并对 3 种误差进行了分析。

本次实验的硬件平台为装有 Intel i5-1035G1 处理器的台式电脑，内存为 8 GB，固态硬盘容量为 477 GB，CPU 显卡为 MX230。使用 Python 3.8 版本实现本文所提出的方法，Attention-Bi-LSTM 模型使用 Keras 深度学习框架，XGBoost 使用 py-xgboost 框架。Attention-Bi-LSTM 模型的参数设置见表 1，XGBoost 的参数设置见表 2。

表 1 Attention-Bi-LSTM 模型的参数

Tab.1 Parameters of the Attention-Bi-LSTM model

参数	参数含义	参数值	设置依据
unit	单元的数量	128	影响模型精度的关键参数,具有最佳数量
time_step	时间步长	24	确定每个输入数据是否与先前连续输入数据的数量相关
num_layers	单元层数	2	默认值为1层,如果设置为2,则第二层接收第一层的计算结果
batch_size	隐藏层宽度	256	一次输入的数据量,通过该参数的设置,可以区分输入数据是否为同一批数据
Epochs	迭代次数	40	取决于计算机的计算能力来确定最佳迭代次数

3.2 预测模型评价

采用“ABL+XGB”组合预测模型对两个数据集进行预测结果和误差分析。首先，使用

Attention-Bi-LSTM 和 XGBoost 模型对同一组数据进行预测，并得出两个模型的预测误差；其次，根据所得误差，分别为上述两个单一的预测模型

表 2 XGBoost 模型的参数
Tab.2 Parameters of the XGBoost model

参数	参数含义	参数值	设置依据
n_estimators	决策树的数量	70	此参数非常强大,可以一次性将模型调整到极限
max_depth	决策树的最大深度	7	常用值范围为10~100,当样本量和特征量较大时,可以适当修改
subsample	随机抽样地点的样本大小	1	控制采样的样本量。默认值为1,这意味着一次提取100%的数据,而0.1表示一次提取10%的数据
eta	迭代决策树	0.1	eta是迭代决策树的步长,也称为学习率,用于确保每个新树具有最佳的预测效果
booster	弱评估器的选择	gbtree	在建树过程中丢弃了一些树,这比梯度提升树具有更好的过拟合功能
alpha	正则项参数	10	当alpha和lambda更大且惩罚更重时,正则项的比例更大,模型的复杂性更低
gamma	模型复杂性的控制	2	防止过度拟合的重要参数

分配权重,并将其与误差倒数法相结合,误差较小的模型被赋予较高的权重;最后,使用 LSTM、Bi-LSTM、Attention-RNN、Attention-LSTM、Attention-Bi-LSTM、XGBoost 和本文预测模型分别进行预测和比较。

3.2.1 预测结果和实际数据比较

图3与图4显示了LSTM、Bi-LSTM、Attention-RNN、Attention-LSTM、Attention-Bi-LSTM、XGBoost 和本文所采用的“ABL+XGB”组合预测模型预测结果及其与实际值的比较,可以从图形趋势的角度观察模型的优越性。由图可知,“ABL+XGB”组合预测模型预测值曲线最接近实际值曲线,并且在预测精度和对比变化的敏感性方面优于其他模型。

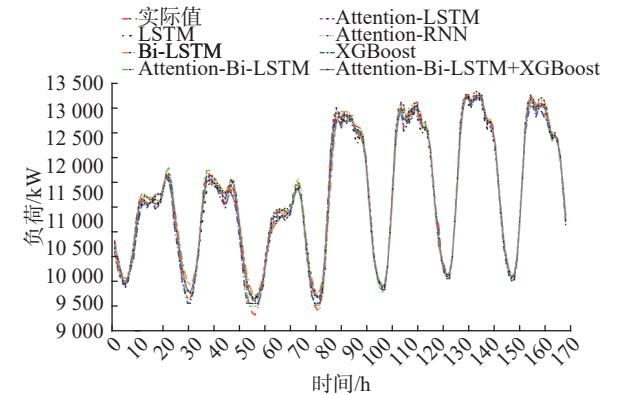


图3 新加坡电力市场预测结果和实际数据的趋势比较图
Fig.3 Trend comparison chart of prediction results and real value in Singapore electricity market

为了更清楚地观察每个模型预测结果的趋势及其与实际值之间的接近程度,本文局部放大了图3和图4。新加坡2020年1月7日24h的局部放大图如图5所示,挪威2021年5月16日24h的局部放大图如图6所示。从图中可以看出,与基准模型和其他模型相比,“ABL+XGB”组合预测模型的预测结果曲线和实际值曲线在趋势和接

近度上都是最接近的,而其余模型的预测值曲线与实际值曲线之间存在很大的偏差,即本文采用的模型拟合效果最好。因此,从总体趋势图和局部放大图的图形趋势来看,“ABL+XGB”组合预测模型比单一预测模型更加接近真实数据且误差更小,预测精度更高,稳定性更好,具备有效性、精准性和实用性的优势。

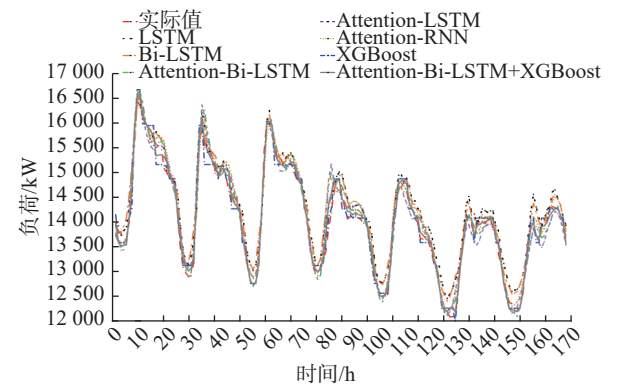


图4 挪威电力市场预测结果和实际数据的趋势比较图
Fig.4 Trend comparison chart of prediction results and real value in Norway electricity market

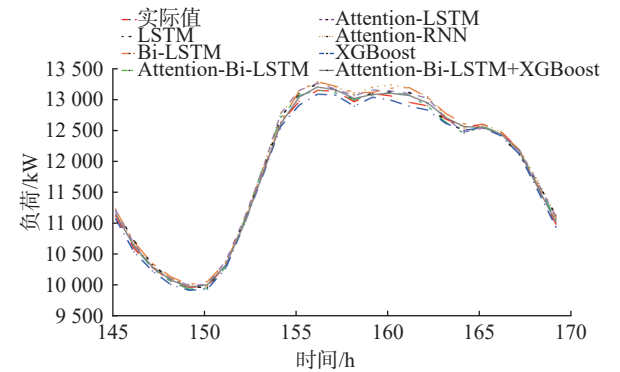


图5 新加坡电力市场2020年1月7日各模型预测结果的局部放大图
Fig.5 Partial enlarged view of models' forecasting results in Singapore electricity market on January 7, 2020

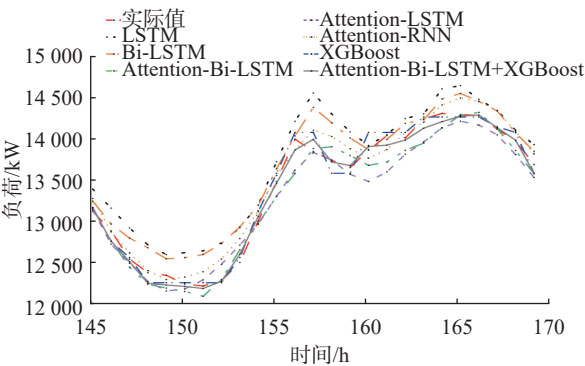


图6 挪威电力市场 2021 年 5 月 16 日各模型预测结果的局部放大图

Fig.6 Partial enlarged view of models' forecasting results in Norway electricity market on May 16, 2021

3.2.2 误差比较分析

本文选取平均相对误差绝对值 (mean absolute percent error, MAPE) 作为对新加坡和挪威两个电力市场的 LSTM、Bi-LSTM、Attention-RNN、Attention-LSTM、Attention-Bi-LSTM、XGBoost 和 “ABL+XGB” 组合预测模型的主要评判指标, 并选取平均绝对误差 (mean absolute error, MAE) 和均方根误差 (root mean squared error, RMSE) 作为辅助评判指标。MAPE、MAE 和 RMSE 的计算公式分别如下所示:

$$X_{MAPE} = \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{y_t} \tag{4}$$

$$X_{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \tag{5}$$

$$X_{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \tag{6}$$

式中: n 为电力负荷数据的数量; y_t 为实际的电力负荷数据; $\hat{y}_t$ 为电力负荷预测数据。

详细的数值结果如表 3 和表 4 所示。通过比较表中的误差值可知: 从整体角度来看, 本文提出的 “ABL+XGB” 组合预测模型的 MAPE、MAE 和 RMSE 最小, 这说明该预测模型的预测精度优于其他 6 种基准模型。

4 特征选择方法评价

本节实验设置同 3.1 节, 同样基于新加坡和挪威电力市场数据集, 分别采用 PC、mRMR、RFE、Lasso 回归、RF 和 XGBoost 这 6 种特征选择方法结合 “ABL+XGB” 预测模型进行对比评价分析,

表 3 新加坡电力市场各预测模型的误差对比
Tab.3 Error comparison of forecasting models in Singapore electricity market

模型	$X_{MAPE} / \%$	X_{MAE} / kW	X_{RMSE} / kW
LSTM	1.011	113.480	147.083
Bi-LSTM	1.155	127.812	168.841
Attention-RNN	0.886	98.16	130.704
Attention-LSTM	0.877	97.691	125.574
Attention-Bi-LSTM	0.711	79.418	105.023
XGBoost	0.526	59.463	62.585
“ABL+XGB” 预测模型	0.387	43.206	54.357

表 4 挪威电力市场各预测模型的误差对比
Tab.4 Error comparison of forecasting models in Norway electricity market

模型	$X_{MAPE} / \%$	X_{MAE} / kW	X_{RMSE} / kW
LSTM	1.011	113.480	147.083
Bi-LSTM	1.155	127.812	168.841
Attention-RNN	0.886	98.16	130.704
Attention-LSTM	0.877	97.691	125.574
Attention-Bi-LSTM	1.091	153.595	210.720
XGBoost	0.814	115.892	148.007
“ABL+XGB” 预测模型	0.682	96.278	125.343

并仍然选取 MAPE、MAE 和 RMSE 作为评判指标。

特征选择的准则因算法而异, 算法由模型参数控制。PC 是选择各特征与电力需求的线性相关系数大于 0.5 的 13 个特征。mRMR 是在最大化特征与目标变量相关性的同时, 最小化特征间的冗余性,

目标函数为 $\max_S \left(\frac{1}{|S|} \sum_{f_i \in S} I(f_i; y) - \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I(f_i; f_j) \right)$ 。

其中: S 为最终选择出的特征子集; $|S|$ 为特征子集 S 中特征的数量; f_i 、 f_j 为子集 S 中的特征; $I(f_i; y)$ 为特征 f_i 与目标变量 y 之间的交互信息。RFE 是通过反复训练模型, 按重要性对特征进行排序, 丢弃最不重要的特征, 直到剩余 7 个特征为止。Lasso 回归是在线性回归模型中加入 L1 正则化项, 将不重要的特征系数压缩为 0, 目标函数为

$\min_{\beta} \left(\sum_{i=1}^n (y_i - \beta X_i)^2 + \lambda \sum_{j=1}^p |\beta_j| \right)$ 。其中: β 为回归系数向量; n 为样本的总数; y_i 为第 i 个样本的真实电力负荷值; X_i 为第 i 个样本的特征向量; p 为特征的总数; β_j 为对应第 j 个特征的回归系数; 正则化参数 λ 设置为 0.02, 其取值基于网格搜索法在区间

[0.01,0.10] 内的初步优化, 并结合 5 折交叉验证确定, 与电力负荷预测场景的 Lasso 应用实践一致^[26]。RF 是选择特征在决策树中的重要性大于等于 0.03 的 7 个特征。XGBoost是选择基于梯度提升树模型的特征重要性大于 0.03 的 7 个特征。

4.1 新加坡电力市场

针对新加坡电力市场数据, 采用 PC、mRMR、RFE、Lasso 回归、RF、XGBoost 6种特征选择方法进行有效性验证, 选取与电力负荷相关性较大的 27 个影响因素作为特征, 如表 5 所示。选择“ABL + XGB”模型对 6 种特征选择方法进行训

表 5 数据集包含因素
Tab.5 Factors contained in the dataset

特征	描述
F1	日类型
F2	节假日类型
F3	时刻
F4	电价
F5	2 m处的温度
F6	2 m处的露点/冰点
F7	2 m处的湿球温度
F8	2 m处的比湿度
F9	2 m处的相对湿度
F10	降水矫正
F11	表面压力
F12	10 m处的风速
F13	10 m处的风向
F14	50 m处的风速
F15	50 m处的风向
F16	全天空表面标准杆数
F17	全天空表面短波向下辐照度
F18	晴空表面短波向下辐照度
F19	大气顶短波向下辐照度
F20	全天空表面反照率
F21	太阳天顶角
F22	全天空表面标准杆数合计
F23	晴空表面标准杆数合计
F24	全天空表面UVA辐照度
F25	全天空表面UVB辐照度
F26	全天空表面UV指数
F27	全天空表面长波向下辐照度

练和测试, 进行结果比较和误差分析, 并加入不进行特征选择算法的模型进行对照, 以证明特征选择的必要性, 进而比较性能最佳的特征选择方法。先使用上述 6 种方法分别对包含 27 个影响因素的数据集进行特征选择; 然后为了验证不同特征选择算法的性能, 将 6 种特征选择方法分别融入 Attention-Bi-LSTM 模型和 XGBoost模型对特征选择后的数据进行预测, 使用误差倒数法计算 Attention-Bi-LSTM 模型和 XGBoost 模型的权重, 形成最终预测模型, 进行预测结果比较。

4.1.1 预测结果和实际数据比较

a. 图 7 显示经过 6 种特征选择方法进行特征提取后“ABL+XGB”预测模型的预测结果和实际值之间的比较。从图 7 中可以看出, 经过 Lasso 回归算法特征选择后的“ABL+XGB”预测模型的预测精度最高, 而不经特征选择算法处理的“ABL+XGB”预测模型预测精度最低。

b. 图 8 为新加坡 2020 年 12 月 30 日 24 h 预测

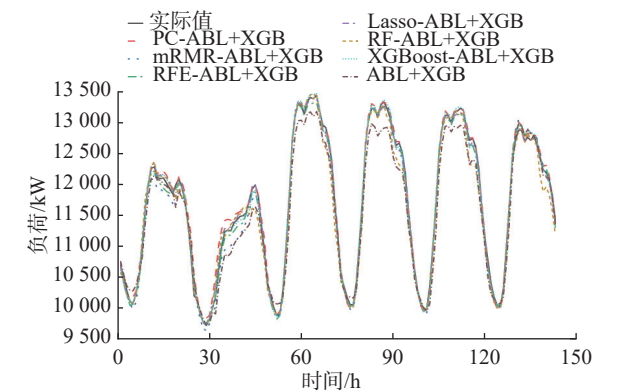


图 7 新加坡电力市场特征选择后预测结果和实际值趋势比较图
Fig.7 Trend comparison chart of prediction results and real values after feature selection in Singapore electricity market

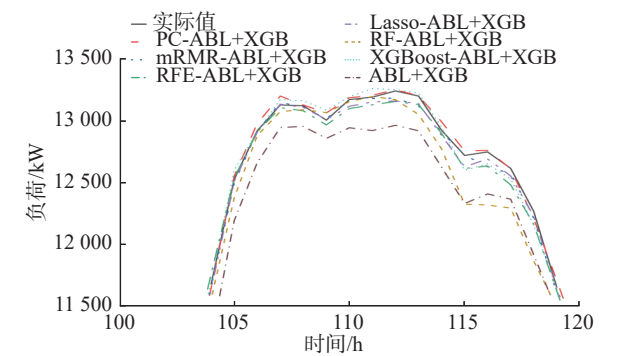


图 8 新加坡电力市场 2020 年 12 月 30 日各模型预测结果的局部放大图
Fig.8 Partial enlarged view of models' forecasting results in Singapore electricity market on December 30, 2020

结果的局部放大图，从中可以看出，经过 Lasso 回归算法特征选择后的“ABL+XGB”预测模型的预测值曲线在走势和数值上均与实际值曲线最接近，即在本文使用的 6 种特征选择方法中，Lasso 回归算法在提高模型精度方面具有最优的性能。

4.1.2 误差比较分析

上述 6 种特征选择后的结果如表 6 所示，预测模型的 MAPE、MAE 和 RMSE 值如表 7 所示，通过比较上述两表结果可知：

表 6 新加坡电力市场特征选择结果

Tab.6 Singapore electricity market feature selection results						
特征	特征选择模型					
	PC	mRMR	RFE	Lasso	RF	XGBoost
F1	×	√	×	×	×	×
F2	×	√	√	×	×	×
F3	√	×	×	√	×	×
F4	×	√	×	√	√	√
F5	√	×	√	×	√	√
F6	×	√	√	×	×	×
F7	√	×	√	×	×	×
F8	×	×	√	×	×	×
F9	√	√	√	×	√	√
F10	×	√	×	×	×	×
F11	×	×	√	×	√	√
F12	×	×	√	×	√	×
F13	×	√	×	×	×	×
F14	×	√	×	×	√	√
F15	×	×	×	×	×	×
F16	√	×	×	×	×	×
F17	√	×	×	×	×	×
F18	√	×	×	√	×	×
F19	√	×	×	√	×	√
F20	√	√	×	×	×	×
F21	√	×	×	√	×	×
F22	×	×	×	×	×	×
F23	√	×	×	×	×	×
F24	√	×	×	×	×	×
F25	×	×	√	×	×	×
F26	×	×	√	×	×	×
F27	√	√	×	√	√	√

a. 通过表 6 可以观察到，不同的选择机制会产生不同的结果。其中，过滤型方法 PC 和 mRMR 选择了较多与太阳能有关的特征，但它们的过度选择会造成信息冗余，降低模型的预测性能。而基于封装型和嵌入型的方法消除了一些相关性较小的特征，预测精度都有不同程度的提高。

表 7 新加坡电力市场各特征选择方法的误差对比
Tab.7 Error comparison of feature selection methods in Singapore electricity market

模型	$X_{MAPE} / \%$	X_{MAE} / kW	X_{RMSE} / kW
无特征选择-ABL+XGB	1.606	190.968	227.794
PC- ABL+XGB	0.691	77.935	104.575
mRMR- ABL+XGB	0.704	80.925	106.158
RFE- ABL+XGB	0.522	60.916	72.821
Lasso- ABL+XGB	0.426	49.267	58.887
RF- ABL+XGB	0.531	61.140	76.713
XGBoost- ABL+XGB	0.463	54.433	66.760

b. 当将影响因素增加到 27 个后，不进行特征选择的“ABL+XGB”预测模型误差大幅增加。从表 6 中可以看出，6 种特征选择方法所筛选出的重要特征各不相同，对模型的优化能力也存在较大的差异。因此，伴随着输入特征数量的增加，大量冗余信息导致模型的预测精度大幅下降，对数据进行特征选择是有必要的。

c. 表 7 显示，经过 Lasso 回归算法特征选择后“ABL+XGB”预测模型的 MAPE、MAE 和 RMSE 值分别为 0.426、49.267 和 58.887，均小于其他 5 种特征选择方法处理后的结果，且小于没有经过特征选择的结果。因此，Lasso 回归算法在特征选择方面优势显著，能够有效消除冗余信息，改进模型预测性能。

4.2 挪威电力市场

为了进一步验证和评估 Lasso 回归算法在特征选择方面的有效性，本文采用挪威的小时负荷数据作为另一个测试数据集，特征选择结果如表 8 所示。图 9、图 10 以及表 9 描述了 Lasso-“ABL+XGB”预测模型与其他特征处理方法的图形对比和误差结果。

从表 6 和表 8 可见，天气相关特征(F5~F9、F12~F14)在 6 种特征选择方法中的选中频率显著低于时间因素(F1~F3)和经济因素(F4)。在新加坡电力市场特征选择结果中，F5(温度)仅被 3 种

表 8 挪威电力市场特征选择结果

Tab.8 Norway electricity market feature selection results

特征	特征选择模型					
	PC	mRMR	RFE	Lasso	RF	XGBoost
F1	×	√	×	×	√	×
F2	×	×	√	√	×	√
F3	×	√	√	×	√	√
F4	√	×	√	√	√	×
F5	×	√	√	×	√	√
F6	√	√	√	√	√	√
F7	×	√	×	×	√	√
F8	√	√	√	×	×	×
F9	√	×	√	√	×	×
F10	×	×	√	×	√	×
F11	√	×	×	√	√	×
F12	√	×	×	√	×	×
F13	√	×	×	×	×	×
F14	√	×	×	×	×	×
F15	×	×	×	×	×	×
F16	×	×	×	×	×	×
F17	×	×	×	×	×	×
F18	×	×	×	×	×	×
F19	×	×	×	×	×	×
F20	√	×	×	×	×	×
F21	√	×	×	×	×	×
F22	×	×	×	×	×	×
F23	×	×	×	×	×	×
F24	×	×	×	×	×	×
F25	×	×	×	×	×	×
F26	×	×	×	×	×	×
F27	√	√	×	√	×	×

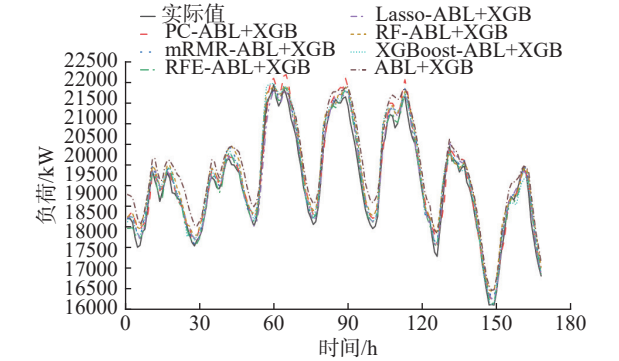


图 9 挪威电力市场特征选择后预测结果和实际数据趋势比较图

Fig.9 Trend comparison chart of prediction results and real values after feature selection in Norway electricity market

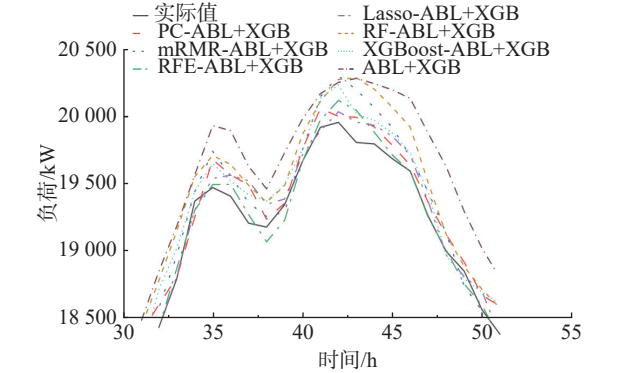


图 10 挪威电力市场 2021 年 12 月 26 日各模型预测结果的局部放大图

Fig.10 Partial enlarged view of models' forecasting results in Norway electricity market on December 26, 2021

表 9 挪威电力市场各特征选择方法的误差对比

Tab.9 Error comparison of feature selection methods in Norway electricity market

模型	$X_{MAPE} / \%$	X_{MAE} / kW	X_{RMSE} / kW
无特征选择- ABL+XGB	2.228	423.713	460.472
PC- ABL+XGB	1.045	203.516	244.851
mRMR- ABL+XGB	1.043	200.653	246.385
RFE- ABL+XGB	0.774	149.182	189.285
Lasso- ABL+XGB	0.741	142.555	182.764
RF- ABL+XGB	1.480	282.304	324.868
XGBoost- ABL+XGB	1.087	210.808	252.809

方法选中, F6(露点)仅被 2 种方法选中, 而时间因素 F3(时刻)和经济因素 F4(电价)分别被 5 种和 4 种方法保留; 在挪威电力市场特征选择结果中, F4(电价)和 F3(时刻)被 6 种方法全部选中, 而天气特征 F5(温度)仅被 3 种方法选中, F7(湿球温度)未被任何方法保留。进一步结合表 7 和表 9, 当天气特征被剔除时, 模型预测误差未显著上升, 如挪威市场中 F7 未被选中, Lasso 回归后 MAPE 仅降低 0.033, 表明天气相关因素对负荷预测的贡献度相对有限。

4.2.1 预测结果和实际数据比较

通过比较图 9 和图 10 中加入 6 种特征选择方法后的预测模型, 可以从图形趋势的拟合程度观察模型的准确性。从图 10 中可以看出, 与其他特征选择方法相比, 实际值曲线与不进行特征选择的预测值曲线之间偏差最大, 使用 Lasso 回归算法进行特征选择的预测值曲线之间偏差最小, 即 Lasso-“ABL+XGB”预测模型具备最优的预测结果。

4.2.2 误差比较分析

表 9 列出了 7 次实验的 3 次精确度测试的数值结果,从中可以看出,挪威电力市场数据预测的误差结果普遍高于新加坡电力市场,除不进行特征选择的预测模型外,RF 特征选择方法的优化性能最差。而 Lasso-“ABL+XGB”预测模型实现了 MAPE、MAE 和 RMSE 的最小结果(分别为 0.741、142.555 和 182.764)。根据误差比较,同样可以得出结论:使用 Lasso 回归算法特征选择后的预测模型比另外 5 种特征选择方法具备更加准确的预测性能,与 4.1 节实验所得结果相同。

5 结 论

通过将影响电力负荷的特征增加到 27 个,并使用 6 种特征选择算法分别对数据集进行特征选择,然后使用“ABL+XGB”组合预测模型对 6 种特征选择算法进行预测和评估,并加入不进行特征选择的预测结果作为对照,最后使用新加坡和挪威两个电力市场的数据进行适应性验证,得到一种基于最优特征选择的电力负荷混合预测方法,主要结论如下:

- a. “ABL+XGB”组合预测模型通过双数据集验证确定其预测结果的准确性。
- b. 不同特征选择方法产生的结果作为模型输入对“ABL+XGB”模型预测结果影响显著。
- c. 天气相关因素与时间因素和经济因素相比,重要性相对较低
- d. 与其他 5 种特征选择方法相比, Lasso 回归算法特征选择优势显著,能够有效降低特征量增加带来的不利影响,更好地改善“ABL+XGB”预测模型性能。

尽管本文所使用的预测模型性能相当可观,但对本文之外的更多特征选择方法可以进行更多测试,以深入了解不同特征如何影响预测精度,从而为研究人员和电力行业提供更多的可能性。

参考文献:

[1] LI X M, LI N, DING S, et al. A novel data-driven seasonal multivariable grey model for seasonal time series forecasting[J]. *Information Sciences*, 2023, 642: 119165.

[2] YIN C, MAO S H. Fractional multivariate grey Bernoulli model combined with improved grey wolf algorithm: application in short-term power load forecasting[J].

Energy, 2023, 269: 126844.

[3] AFZAL S, SHOKRI A, ZIAPOUR B M, et al. Building energy consumption prediction and optimization using different neural network-assisted models; comparison of different networks and optimization algorithms[J]. *Engineering Applications of Artificial Intelligence*, 2024, 127: 107356.

[4] ZHANG L J, JÁNOŠÍK D. Enhanced short-term load forecasting with hybrid machine learning models: CatBoost and XGBoost approaches[J]. *Expert Systems with Applications*, 2024, 241: 122686.

[5] FAN G F, HAN Y Y, LI J W, et al. A hybrid model for deep learning short-term power load forecasting based on feature extraction statistics techniques[J]. *Expert Systems with Applications*, 2024, 238: 122012.

[6] BHAMBU A, GAO R B, SUGANTHAN P N. Recurrent ensemble random vector functional link neural network for financial time series forecasting[J]. *Applied Soft Computing*, 2024, 161: 111759.

[7] GAO R B, ZHANG X C, LIANG M H, et al. Wave energy forecasting: a state-of-the-art survey and a comprehensive evaluation[J]. *Applied Soft Computing*, 2025, 170: 112652.

[8] ABAD A R B, MOUSAVI S, MOHAMADIAN N, et al. Hybrid machine learning algorithms to predict condensate viscosity in the near wellbore regions of gas condensate reservoirs[J]. *Journal of Natural Gas Science and Engineering*, 2021, 95: 104210.

[9] TANG L, LI J Y, DU H C, et al. Big data in forecasting research: a literature review[J]. *Big Data Research*, 2022, 27: 100289.

[10] CHEN H H, ZHU M Y, HU X, et al. Research on short-term load forecasting of new-type power system based on GCN-LSTM considering multiple influencing factors[J]. *Energy Reports*, 2023, 9: 1022–1031.

[11] DAI Y M, YANG X Y, LENG M M. Optimized Seq2Seq model based on multiple methods for short-term power load forecasting[J]. *Applied Soft Computing*, 2023, 142: 110335.

[12] WANG J Z, LV M Z, LI Z W, et al. Multivariate selection-combination short-term wind speed forecasting system based on convolution-recurrent network and multi-objective chameleon swarm algorithm[J]. *Expert Systems with Applications*, 2023, 214: 119129.

[13] 杨超, 冉启武, 罗德虎, 等. 基于注意力机制的 CNN-BIGRU 短期电价预测 [J]. *电力系统及其自动化学报*, 2024, 36(3): 22–29.

[14] JEON Y, HWANG G. Feature selection with scalable variational gaussian process via sensitivity analysis based on L^2 divergence[J]. *Neurocomputing*, 2023, 518: 577–592.

- [15] CHENG F, CHU F X, XU Y, et al. A steering-matrix-based multiobjective evolutionary algorithm for high-dimensional feature selection[J]. *IEEE Transactions on Cybernetics*, 2022, 52(9): 9695–9708.
- [16] WANG Q J, YUE C F, LI X Q, et al. Enhancing robustness of monthly streamflow forecasting model using embedded-feature selection algorithm based on improved gray wolf optimizer[J]. *Journal of Hydrology*, 2023, 617: 128995.
- [17] GONG H H, LI Y Y, ZHANG J N, et al. A new filter feature selection algorithm for classification task by ensembling Pearson correlation coefficient and mutual information[J]. *Engineering Applications of Artificial Intelligence*, 2024, 131: 107865.
- [18] DAI Y M, YANG X Y, LENG M M. Forecasting power load: a hybrid forecasting method with intelligent data processing and optimized artificial intelligence[J]. *Technological Forecasting and Social Change*, 2022, 182: 121858.
- [19] DAI Y M, YU W J, LENG M M. A hybrid ensemble optimized BiGRU method for short-term photovoltaic generation forecasting[J]. *Energy*, 2024, 299: 131458.
- [20] HUO Y N, XING H W, YANG Y, et al. Real-time estimation of aggregated electric vehicle charging load based on representative meter data[J]. *Energy*, 2025, 321: 135162.
- [21] NI F Y, XIANG Y, WANG S Q, et al. Charging management of electric vehicles with consumption of renewable energy[J]. *Energy*, 2025, 321: 135264.
- [22] 白冰青, 刘江涛, 王旭, 等. 基于MRMR和双重注意力机制的城市能源多元负荷短期预测[J]. *电力系统自动化*, 2022, 46(17): 44–55.
- [23] ALBASHISH D, HAMMOURI A I, BRAIK M, et al. Binary biogeography-based optimization based SVM-RFE for feature selection[J]. *Applied Soft Computing*, 2021, 101: 107026.
- [24] ZHANG N, JIAO B B, YE Y X, et al. Embedded cooling method with configurability and replaceability for multi-chip electronic devices[J]. *Energy Conversion and Management*, 2022, 253: 115124.
- [25] 张赞宁, 魏广军. 考虑特征选择的短期光伏功率组合预测模型[J]. *电力系统及其自动化学报*, 2024, 36(8): 122–132.
- [26] LU S X, XU Q F, JIANG C X, et al. Probabilistic load forecasting with a non-crossing sparse-group Lasso-quantile regression deep neural network[J]. *Energy*, 2022, 242: 122955.
- [27] BHATNAGAR M, ROZINAJ G, VARGIC R. Using crafted features and polar bear optimization algorithm for short-term electric load forecast system[J]. *Energy and AI*, 2025, 19: 100470.
- [28] JIANG Z Y, CHE J X, HE M J, et al. A CGRU multi-step wind speed forecasting model based on multi-label specific XGBoost feature selection and secondary decomposition[J]. *Renewable Energy*, 2023, 203: 802–827.
- [29] DAS S, IMTIAZ M S, NEOM N H, et al. A hybrid approach for Bangla sign language recognition using deep transfer learning model with random forest classifier[J]. *Expert Systems with Applications*, 2023, 213: 118914.
- [30] PREETI, DEEP K. A random walk Grey wolf optimizer based on dispersion factor for feature selection on chronic disease prediction[J]. *Expert Systems with Applications*, 2022, 206: 117864.
- [31] WANG J J, XU S L, SHU S Q. An optimal weight heterogeneous integrated carbon price prediction model based on temporal information extraction and specific comprehensive feature selection[J]. *Energy*, 2024, 312: 133654.
- [32] 代业明, 周琼. 基于改进Bi-LSTM和XGBoost的电力负荷组合预测方法[J]. *上海理工大学学报*, 2022, 44(2): 138–147.

(编辑: 丁红艺)