

doi:10.3969/j.issn.1001-4616.2023.02.013

基于机器学习的短期规上行业工业增加值预测

项晓宇¹,朱敏捷¹,周灵刚¹,钟磊¹,闵富红²

(1.国家电网台州供电公司,浙江 台州 318000)

(2.南京师范大学南瑞电气与自动化学院,江苏 南京 210023)

[摘要] 针对现有工业增加值预测存在经济数据统计滞后、单一模型预测精度差的问题,本文提出一种基于 Stacking 框架下短期规上行业的工业增加值预测模型,实现了预测时效性与精度的提升.通过皮尔逊相关性系数检验,对浙江省某市 4 个重点规上行业的用电量、工业增加值进行分析,发现两者具有中强度的相关性,表明了基于行业用电量预测工业增加值方法的可行性.接着,以随机森林算法、自适应增强算法、极致梯度增强算法 3 种模型作为基学习器,支持矢量回归机算法作为元学习器,搭建双层 Stacking 融合模型框架对规上行业用电量、工业增加值、当地气温数据进行模型训练测试,得到最终预测模型.最后,将本文所提出的 Stacking 模型与单一模型预测误差指标进行实例对比分析,结果表明,该模型具有更高的预测精度,且采用月度收集的实时电力消费数据提升了预测时效性,能被更好地应用在“双碳”背景下工业增加值的预测场景中,也有利于政府分析经济发展趋势.

[关键词] 规上工业,工业增加值预测,机器学习,Stacking 算法

[中图分类号] TP301.6 [文献标志码] A [文章编号] 1001-4616(2023)02-0099-08

Short-Term Industrial Added Value Prediction of the Above-Scale Industry Based on Machine Learning

Xiang Xiaoyu¹, Zhu Minjie¹, Zhou Linggang¹, Zhong Lei¹, Min Fuhong²

(1. Taizhou Electric Power Company, Taizhou 318000, China)

(2. School of Electrical and Automation Engineering, Nanjing Normal University, Nanjing 210023, China)

Abstract: Considering the lagging indicators with the economic statistics and poor accuracy of the single model in the prediction of industrial added value, this paper proposes a short-term model (STM), via the Stacking algorithm, to improve the timeliness and accuracy of the prediction. With Random Forest Regressor, AdaBoost Regressor and Xgboost Regressor as the base learners and Support Vector Regression as the meta learners, STM is established in a two-layer stacking framework. Via the Pearson correlation analysis, the data of the electricity consumption, the local temperature, and the industrial added value of the above-scale industries in a city of Zhejiang province are investigated, the result of which shows that these data are quite related to each other. Therefore, via the training test of these data, STM can effectively predict the industrial added value through the industrial electricity consumption. The prediction results via STM are compared with that of the single model. The comparison shows that the monthly collected real-time electric consumption in Stacking model can improve the time efficiency of the prediction. Therefore, STM is quite suitable for the prediction of the industrial added value, and useful for the dual control system of the total carbon emissions and carbon emission intensity, which is helpful for the government to analyze the trend of economic development.

Key words: above-scale industry, industrial added value prediction, machine learning, Stacking algorithm

电力消费是国民经济发展的风向标,工业生产活动的开展离不开电力消费,且工业用电增长与国民经济的发展有长期稳定的联系并伴有明显趋势性和季节性^[1].而工业增加值是衡量工业经济成效的重要指标,准确及时地预测工业增加值能帮助政府把握经济趋势并做出应对,保障经济稳定发展^[2].智能电网大数据的优势为用电量数据月度采集和分析提供了优质平台^[3],相对于 GDP 每季度测算一次,工业增加值、产值等经济数据也需要次月下旬才能完成统计,用电量指标具有指标发布时间早、电网采集便捷的优势,

收稿日期:2022-09-27.

基金项目:国家自然科学基金项目(61971228).

通讯作者:项晓宇,工程师,研究方向:新能源. E-mail:418892683@qq.com

能够更为及时有效地观测经济基本运行状况,第一时间反映瞬息万变的市场信息.因此,基于用电量数据能有效提高预测时效性和准确性,但数据采样频率的提升对预测模型适配性提出了更高要求.

目前,关于规上行业的短期工业增加值预测,主要分为传统方法和人工智能方法两类.传统预测方法有回归模型法^[4]、趋势外推法模型^[5]以及时间序列模型^[6],大多为线性模型,算法简单,但均采用宏观经济数据预测,且在基于月度数据的短期预测中精度会下降.人工智能方法包括随机森林、自适应增强、极致梯度增强、支持向量回归机模型、神经网络等^[7],在工业增加值预测方面的研究主要集中在长短期记忆(long-short term memory, LSTM)神经网络^[8]上, LSTM属于非线性模型,能较好地应对工业增加值等复杂非线性数据的预测问题,但经济数据滞后性的问题仍然存在.除了LSTM模型外,其他人工智能算法在工业增加值预测领域研究鲜有报道,主要被应用在用电量预测^[9]、负荷预测^[10]领域.文献[11]通过随机森林(random forest regressor, RF)回归算法有效识别不同用户群体的用电关联因素,并预测用电量,该方法适用于大数据分析处理并具有较高的预测精度,但对某些含特定噪声的数据训练时容易出现过拟合.文献[12]采用自适应增强(adaptive boosting, AdaBoost)回归算法提升微电网光伏发电在不同天气条件下的预测准确度,对提高微电网经济调度水平具有指导意义,算法模型灵活度高且不易过拟合,但对异常样本敏感.文献[13]利用极致梯度增强(extreme gradient boosting, Xgboost)回归算法进一步提升了负荷预测准确度,对需求侧电能管理极具意义,模型高效可扩展,处理大规模数据集时效果好,但调参复杂,不适合处理高维数据.文献[14]引入支持向量回归机(support vector regression, SVR)算法预测建筑冷负荷,能有效满足建筑能耗预测的工程应用需求,模型计算复杂度低,但对参数和核函数的选择敏感.上述人工智能方法在非线性时间序列的数据挖掘上更具优势,但单一算法很难适用于所有预测场景,因此,学者们尝试提出基于多个不同模型的融合算法以适应更复杂的非线性数据和不同应用场景需求^[15].文献[16]基于Stacking模型融合Xgboost、SVR、RF等6种算法提出一种高精度的负荷预测方法,有助于满足能源分配需求,综合多个模型结果精度明显高于单一模型且鲁棒性强,但元学习器的选择对预测结果影响较大.根据上述研究成果,人工智能算法对用电量、负荷等时间序列数据预测精度的提升效果显著,工业增加值预测可以借鉴相关方法,但数据本身具有更复杂的非线性,且规上工业种类繁多.因此,为了满足不同规上行业短期工业增加值预测需求,研究一种基于用电量数据的多模型融合预测方法,能为政府提供更具实效性和置信度的支撑数据,具有较高研究价值.

本文为了提升以规上行业月度用电量对工业增加值进行短期预测的有效性和精度,提出一种基于Stacking框架的多算法融合模型并分析误差性能指标和预测结果.采用随机森林回归算法、自适应增强回归算法和极致梯度增强回归算法为基学习器,以支持向量回归机算法为元学习器,对浙江省某市橡胶与塑料制品业、电气机械与器材制造业、金属制品业、专用设备制造业4个重点规上行业的历史用电量、工业增加值、气温数据进行模型训练,搭建双层Stacking融合框架下的工业增加值预测模型.模型能够实现以月度即时采集的用电量数据和天气状况对上月工业增加值预测,对比单一模型的预测结果,预测精度取得了明显提升.研究成果可以有效解决以往工业增加值预测数据统计滞后和单一模型短期预测精度差的问题,也为政府分析工业用电与经济变化趋势关系和制定相关政策提供有效的数据支撑.

1 规上行业工业增加值预测相关性分析

目前大多数工业增加值预测研究都是以季度、年度数据建立模型,而利用某个具体规上行业月度用电量建立的工业增加值预测模型,由于数据时间跨度变小,具有更高敏感性.为了确保以月度用电量数据预测工业增加值的可行性,需要对用电量与工业增加值的相关性进行量化分析.

1.1 电力基础要素分析

为了更直观地观察和分析用电量与工业增加值的趋势性联系,这里选取浙江省某地市2017年—2022年工业用电量与工业增加值数据.用电量来自浙江省某地市国家电网统计数据,工业增加值来自统计局发布的统计数据,两者时间变化趋势如图1所示.图1(a)为2017年3月—2022年5月的月度用电量,用电量从19.25亿kWh逐步上升至26.02亿kWh,在15~25亿kWh的区间内不断上升.图1(b)为2017年3月—2022年5月的月度工业增加值,从85亿元上升至140亿元.工业用电量与工业增加值都呈现出了波动性地稳步上升趋势,工业电力投入的增加从一定程度上反映了工业经济发展水平的提升.工业涵盖制造业种类繁多,进

一步选取浙江省某地市经济贡献度高的橡胶与塑料制品业、电气机械与器材制造业、金属制品业、专用设备制造业 4 个重点规上行业 2020—2022 年用电量与工业增加值数据,量化分析两者关联性。

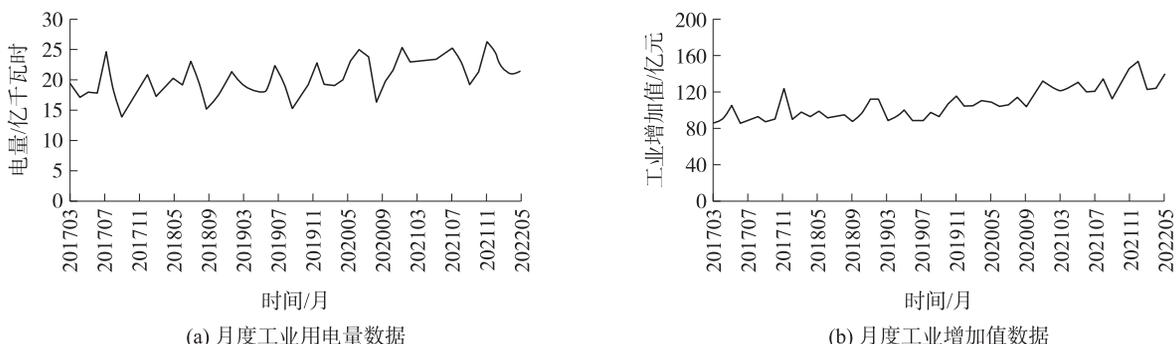


图 1 电力-经济数据

Fig. 1 Electricity-economy data

通常量化电力指标与宏观经济指标之间联系,可以采用皮尔逊相关系数分析.皮尔逊相关系数^[17]是分布在 $[-1, 1]$ 区间上体现要素相关性的数值,一般认为, $0.8 \leq |\rho_{XY}| \leq 1$ 可认为两个要素极强相关, $0.6 \leq |\rho_{XY}| < 0.8$ 则强相关, $0.4 \leq |\rho_{XY}| < 0.6$ 则中度相关, $0.2 \leq |\rho_{XY}| < 0.4$ 则弱相关,而 $|\rho_{XY}| < 0.2$ 则极弱相关或不相关.总体皮尔逊相关系数计算如式(1)所示:

$$\rho_{xy} = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^n \frac{(X_i - E(X)) (Y_i - E(Y))}{\sigma_x \sigma_y}}{n}, \quad (1)$$

式中, σ 是样本标准差, E 为总体数据均值.

按式(1)计算浙江省某地市 2020—2022 年 5 月期间 4 个重点规上行业国家电网统计的月度用电量数据与统计局发布的月度工业增加值数据的皮尔逊相关性系数,结果如表 1 所示.橡胶与塑料制品业、电气机械与器材制造业、专用设备制造业和金属制品业用电量与工业增加值的相关系数分别为 0.618(强相关)、0.79(强相关)、0.609(强相关)、0.582(中度相关).因此,规上行业用电量可以作为基础要素预测工业增加值,反映经济发展形势,但工业设备在不同天气情况下工作效率不同,此外,加入天气扰动要素也能更好地解释用电量和工业增加值的季节性波动.

表 1 规上行业用电量与工业增加值相关性系数表

Table 1 Correlation coefficient between electricity consumption of industries above designated size and added value

规上行业	用电量与工业增加值皮尔逊相关系数
橡胶与塑料制品业	0.618
电气机械和器材制造业	0.79
专用设备制造业	0.609
金属制品业	0.582

1.2 天气扰动要素分析

天气扰动是影响工业用电和经济的重要因素,由图 1 所示,用电量和工业增加值的增速具有明显的季节性,夏季和冬季用电量偏高,工业增加值也一般在年末达到全年顶峰.天气指标主要包括温度、湿度、风力等指标,温度是最具代表性的季节性指标,温度过高过低都会影响工业生产投入的用电量,进而影响工业增加值.浙江省某地市 2017 年 3 月—2022 年 5 月气象局公布的月度气温变化情况如图 2 所示.

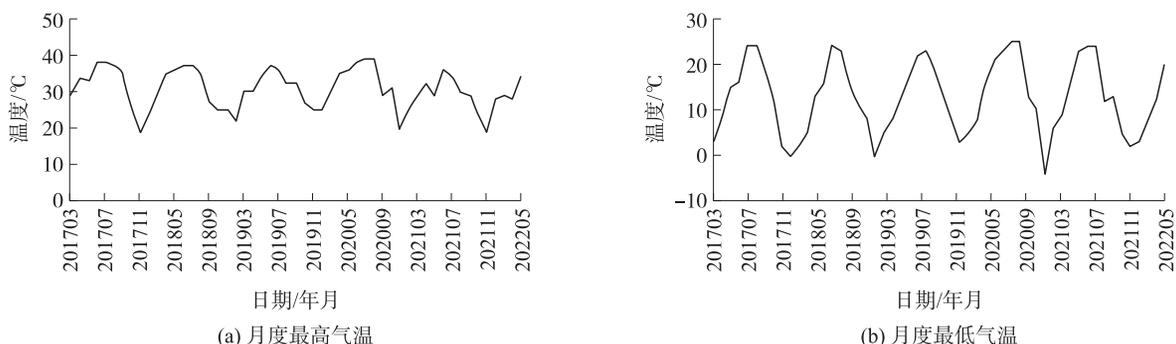


图 2 天气情况

Fig. 2 Weather condition

图 2(a)、(b)为月度最高气温和月度最低气温,夏季最高气温可达到 39 ℃,冬季最高气温达到 18 ℃,夏季最低气温为 23 ℃,冬季最低气温为-4 ℃,整体呈现出随年月的周期性变化. 本文在对浙江省某地市规上工业增加值预测模型的建立中,加入平均最高、最低温度作为扰动要素用于消除温度对预测结果的季节性影响,以提升预测模型的准确度,更好地解释用电量与工业增加值关系.

2 规上行业工业增加值预测模型

基于上述分析,规上行业用电量与工业增加值具有较好相关性,国家电网完成月度用电量数据统计时间短,用于预测工业增加值会更具时效性. 为了建立预测性能更好的行业用电量-工业增加值预测模型,利用机器学习算法挖掘规上行业用电量、工业增加值、天气等历史数据的特征信息. 机器学习采用 Stacking 算法这一集成学习方法融合 4 种算法模型进一步提升模型泛化能力. 本文提出一种基于 Stacking 框架的短期规上行业工业增加值预测模型,融合随机森林回归、自适应增强回归、支持向量回归机和极致梯度增强回归 4 种机器学习方法,充分挖掘用电量、工业增加值、温度数据特征,有效降低预测结果的平均绝对误差、均方根误差及均方误差.

基于双层 Stacking 工业增加值预测模型算法框架如图 3 所示,相应的伪代码由算法 1 给出,具体模型训练步骤如下:

(1) 数据训练测试

选取近 2 年橡胶与塑料制品业、电气机械与器材制造业、金属制品业、专用设备制造业的电力数据和天气数据 x_m 为预测输入,工业增加值数据 y_m 为预测输出目标构成完整数据集 D ,按 6:4 比例划分数据集 D 为训练数据和测试数据,训练数据用于建立模型,预测数据用于检验模型效果,实现基于电力数据和天气数据对工业增加值的预测.

(2) 模型框架构建

双层 Stacking 算法框架的第一层采用随机森林回归 \mathcal{L}_1 、自适应增强回归 \mathcal{L}_2 和极致梯度增强回归 \mathcal{L}_3 3 种算法,以 D 训练集中的电力数据和天气数据 x_i 为输入,以工业增加值数据 y_i 为输出,训练生成基学习器 h_i ($i=1,2,3$). 将 D 训练集输入基学习器 h_i 得到的工业增加值预测结果 z_i 作为输入,以对应工业增加值实际数据 y_i 为预测目标,组合构成新一轮训练数据 D' ,用于二次学习. 第二层数据训练时,采用短期数据预测效果更好的支持向量回归机算法 \mathcal{L} ,并以非线性径向基函数(Radial Basis Function, RBF)为核函数构成元学习器 h' ,由此建立起双层架构的 Stacking 模型 H . 以需要预测月份的电力和天气数据 x 为输入,得到月度工业增加值预测的输出结果为 $H(x)$.

下面给出 Stacking 模型融合的 4 种学习器算法简要说明:随机森林回归是一种非线性基于树的学习方法,最终预测结果由互不相关的多棵回归树共同决定;自适应增强回归是一种迭代算法,其核心思想是针对同一个训练集训练不同弱学习器,并集合构成一个更强的学习器;极致梯度增强回归算法通过不断学习新函数并进行特征分裂,拟合前次预测残差,最后将所有分支预测值相加即是最终预测值;支持向量回归机是支持向量在函数回归领域的应用,以回归线为中心作宽度为 2ϵ 的间隔带,训练样本落入其中则认为预测准确,否则计算损失值. 通过 Stacking 框架融合 4 种学习器,综合不同算法优势并提升模型泛化能力.

(3) 预测误差检验

采用平均绝对误差(mean absolute error, MAE)、均方根误差(root mean square error, RMSE)及均方误差(mean square error, MSE)作为工业增加值预测模型的评价指标. 其值越小,则表示模型精度越高,MAE、

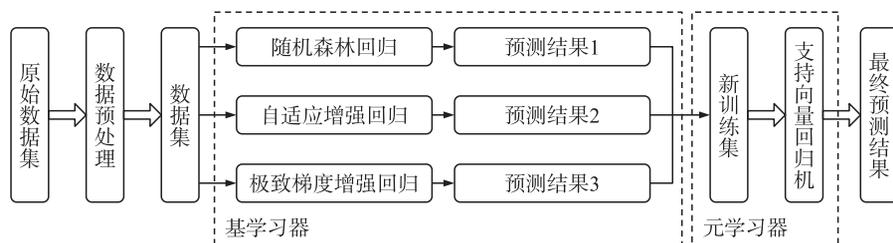


图 3 Stacking 算法模型框图

Fig. 3 Flow chart of Stacking model

RMSE 和 MSE 的定义如(2)-(4)所示:

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - f(x_i)|, \quad (2)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - f(x_i))^2}, \quad (3)$$

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - f(x_i))^2, \quad (4)$$

式中, m 表示待预测序列长度, y_i 表示实际值, $f(x_i)$ 表示预测值.

算法 1 Stacking 回归算法

输入:数据集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$; 基学习器算法 $\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_T$; 元学习器算法 \mathcal{L} ;
 过程:
 1: for $t = 1, 2, \dots, T$ do
 2: $h_t = \mathcal{L}_t(D)$; /* h_t 表示基学习器 \mathcal{L}_t 对数据集 D 训练得到的模型 */
 3: end for
 4: $D' = \emptyset$; /* 初始化二次学习训练集 D' */
 5: for $i = 1, 2, \dots, m$ do
 6: for $t = 1, 2, \dots, T$ do
 7: $z_{it} = h_t(x_i)$; /* z_{it} 表示训练好的基学习器模型 h_t 对输入数据 x_i 的预测结果 */
 8: end for
 9: $D' = D' \cup ((z_{i1}, z_{i2}, \dots, z_{iT}), y_i)$; /* 基学习器算法预测结果 z_{it} 与对应实际结果 y_i 构成新的训练集 D' */
 10: end for
 11: $h' = \mathcal{L}(D')$; /* h' 表示元学习器 \mathcal{L} 对新构建的数据集 D' 训练得到的模型 */
 输出: $H(x) = h'(h_1(x), h_2(x), \dots, h_T(x))$ /* 对预测输入 x 的双层 stacking 模型最终预测结果 */

3 算例分析

选取浙江省某市 2020 年至 2022 年的四大重点规上行业(橡胶与塑料制品业、电气机械与器材制造业、金属制品业、专用设备制造业)的行业用电量、工业增加值及当地天气数据进行算例分析,利用提出的 Stacking 工业增加值预测模型,为政府分析工业经济变化趋势提供有效的预测数据支撑.

3.1 数据预处理

(1) 异常值检测

本文使用箱型图对预测模型训练数据中的规上行业用电量、工业增加值进行数据检测与分析,主要是对于数据样本集中的离群点进行搜索并处理,防止离群点的出现造成预测误差偏大.箱型图分析依据实际数据,设定一组上下界,脱离此区间的数即为异常值.规上行业的工业增加值、用电量箱型图如图 4 所示.图 4(a)中 4 个规上行业工业增加值数据区间分别为 $[5.9, 7.8]$ 、 $[3.7, 4.9]$ 、 $[7.5, 8.1]$ 、 $[8.3, 11.2]$,实际数据均在数据区间内.图 4(b)规上行业用电量 4 个数据区间为 $[1, 1.2]$ 、 $[2.1, 2.6]$ 、 $[0.7, 0.8]$ 、 $[3.6, 3.9]$,实际

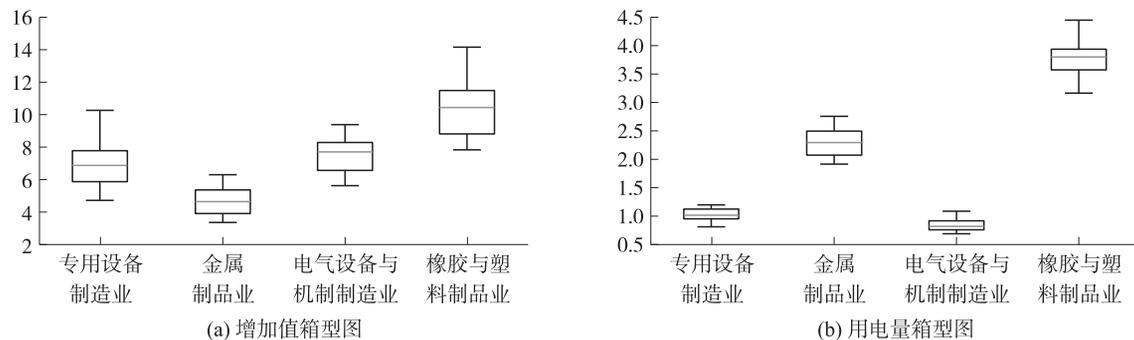


图 4 规上行业工业箱型图

Fig. 4 Box diagram of industries above designated size

数据也没有超出数据区间. 结果表明,实例分析所采用的规上行业工业增加值和用电量数据没有异常值.

(2)数据 Z-Score 标准化处理

由于用电量、工业增加值、天气具有不同量纲单位,依据式(5)对输入数据进行 Z-Score 标准化处理,排除模型训练中量纲对预测误差的影响.

$$x^* = \frac{x - \bar{x}}{\sigma}, \tag{5}$$

式中, x^* 为标准化结果, x 为样本值, \bar{x} 为样本均值, σ 为样本标准差.

3.2 参数设置

在双层 Stacking 模型训练中,以随机森林回归算法、自适应增强、极致梯度增强算法为基学习器,支持矢量回归机为元学习器进行模型融合. 使用的学习器参数设置如表 2 所示,融合的不同模型中树的数目、学习率、基函数等都将影响模型预测精度.

表 2 学习器参数设置
Table 2 Parameter setting of learners

模型名称	参数配置
随机森林回归	树的数目为 200,分支所需的最小样本量 2,叶子节点存在所需最小样本量 1
自适应增强回归	树的数目为 50,学习率 1.0,取线性函数
极致梯度回归	树的数目为 100,每棵树划分的特征比重 1.0,学习率 0.3,随机采样比例为 1,最小叶子节点样本权重和为 1,最大树深度 6,节点的最小损失函数值 0
支持向量回归机	采用径向基函数作为核函数

3.3 预测结果对比分析

选取上文提到的浙江省某市 2020 年至 2022 年 5 月国家电网统计的橡胶与塑料制品业、电气机械与器材制造业、金属制品业、专用设备制造业 4 个重点规上行业用电量、气象局统计的天气数据、工业增加值数据,分别采用随机森林回归(RF)算法、自适应增强回归(AdaBoost)算法,极致梯度增强回归(Xgboost)算法和支持矢量回归机(SVR)算法进行单一模型的预测训练,并以 Stacking 算法融合多个模型进行预测训练,5 种不同方法在测试集上的 MAE、RMSE 和 MSE 误差指标如表 3 所示.

表 3 预测误差表
Table 3 Forecast error table

规上工业	模型	误差指标		
		MAE/亿元	RMSE/亿元	MSE/亿元
橡胶与塑料制品业	RF	1.337	1.426	2.043
	Adaboost	1.198	1.267	1.616
	Xgboost	1.892	2.066	4.325
	SVR	1.189	1.322	1.774
	Stacking	0.824	0.911	0.858
电气机械与器材制造业	RF	0.605	0.685	0.491
	Adaboost	0.565	0.616	0.398
	Xgboost	0.752	0.891	0.795
	SVR	0.500	0.585	0.369
	Stacking	0.500	0.562	0.321
金属制品业	RF	0.494	0.541	0.318
	Adaboost	0.416	0.475	0.249
	Xgboost	0.612	0.686	0.478
	SVR	0.453	0.527	0.304
	Stacking	0.399	0.457	0.210
专用设备制造业	RF	0.857	1.010	1.023
	Adaboost	0.663	0.868	0.758
	Xgboost	1.150	1.392	1.976
	SVR	0.805	0.988	0.976
	Stacking	0.678	0.755	0.583

以橡胶与塑料制品业为例,本文提出的 Stacking 模型的 MAE、RMSE、MSE 分别为 0.824、0.911、0.858,

采用单个模型预测时 3 个误差指标最好结果分别为 1.189(SVR)、1.267(AdaBoost)、1.616(AdaBoost)。通过 MAE、MSE、RMSE 对用电量短期现时预测误差进行评估,结果表明,Stacking 融合模型对橡胶与塑料制品业工业增加值预测模型的 3 种误差指标分别下降了 0.365、0.356、0.758,模型预测误差明显提升,在其他 3 个规上行业中也是预测误差表现最好的。为了进一步验证提出模型在实际预测中的效果,分析 2022 年 6 月以上 5 种模型工业增加值预测结果与统计局公布的工业增加值实际结果作绝对百分比误差分析。

5 种模型以浙江省某地市 2022 年 6 月国家电网公布的即时月度用电量数据和气象局天气数据预测 4 个重点规上行业工业增加值,预测结果与统计局公布的实际结果作绝对百分比误差如图 5 所示。橡胶与塑料制品业、电气机械与器材制造业、金属制品业、专用设备制造业通过 Stacking 模型预测的绝对百分比误差分别为 1.70%、0.45%、0.29%和 0.24%,RF 模型的预测绝对百分比误差 7.30%、5.45%、1.13%和 3.73%,Adaboost 的预测绝对百分误差分别为 4.17%、7.29%、1.25%和 0.29%,Xgboost 的预测绝对百分比误差为 21.54%、2.91%、11.96%和 19.37%,SVR 的预测绝对百分比误差为 2.04%、8.54%、2.87%和 2.04%。对比上述结果,Xgboost 预测稳定和预测精度均较差,RF、Adaboost、SVR 3 种方法预测精度也不理想。Stacking 融合算法对 4 个行业的预测误差最小,预测稳定性最高。结合模型在测试集上的 MAE、MSE 和 MAPE 误差指标对比分析结果和 2022 年 6 月份预测结果与实际结果的绝对百分比误差,本文提出的基于 Stacking 融合算法的规上行业短期工业增加值预测模型具有较高的预测精度,能为政府预判规上行业经济运行趋势提供有效的预测数据支撑。

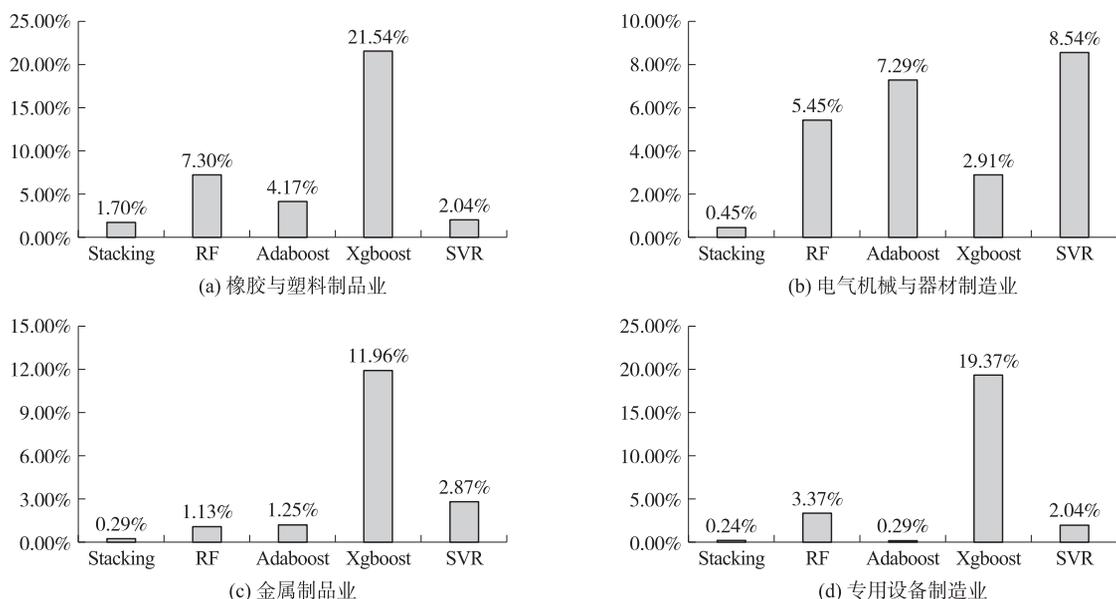


图 5 绝对百分误差对比

Fig. 5 Absolute percentage error comparison

4 结论

本文针对工业增加值预测存在的数据滞后性和单一模型预测精度低的问题,提出一种基于 Stacking 融合模型的规上行业工业增加值预测模型,融合随机森林回归、自适应增强回归、极致梯度增强回归、支持矢量回归机 4 种学习器,提升模型泛化能力,并充分利用智能电网大数据平台优势,以浙江省某地市国家电网月度即时统计的用电量数据、气象局公布的温度数据为输入,统计局公布的工业增加值为预测目标进行工业增加值预测模型训练。通过 Stacking 融合模型与单一模型预测误差指标的实例对比分析,得出如下结论:

(1) 规上行业工业增加值预测,考虑多种预测模型的结合,能够综合不同模型优势并降低预测误差,对不同类型的规上工业具有较强鲁棒性。本文提出 Stacking 框架融合多个模型预测工业增加值,通过随机森林回归、自适应增强回归、极致梯度增强回归 3 种基学习器进行第一轮预测训练,再作为元学习器的支持向量回归机对第一轮预测结果二次训练,得到最终预测值,显著降低了预测误差。

(2) 大数据时代,以用电量为数据预测工业增加值具有更好的时效性。用电量具有获取即时、便捷的优势,能为政府经济数据分析,根据经济变化形势制定相关应对政策提供前瞻性的数据支撑。

[参考文献]

- [1] 刘自敏,邓明艳,杨丹,等.降低企业用能成本可以提高能源效率与社会福利吗——基于交叉补贴视角的分析[J].中国工业经济,2020(3):100-118.
- [2] 顾光同,许冰.中国工业增加值的半月预报:基于宏观月度数据[J].系统工程理论与实践,2018,38(8):1983-1993.
- [3] DILEEP G. A survey on smart grid technologies and applications[J]. Renewable energy,2020,146(2):2589-2625.
- [4] ZHU J M, LIU J P, WU P, et al. A novel decomposition-ensemble approach to crude oil price forecasting with evolution clustering and combined model[J]. International journal of machine learning and cybernetics,2019,10(12):3349-3362.
- [5] MENG M, WANG L X, SHANG W. Decomposition and forecasting analysis of China's household electricity consumption using three-dimensional decomposition and hybrid trend extrapolation models[J]. Energy,2018,165(A):143-152.
- [6] CONEJO A J, PLAZAS M A, ESPINOLA R, et al. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models[J]. IEEE transactions on power systems,2021,20(2):1035-1042.
- [7] CHOU J S, TRAN D S. Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders[J]. Energy,2018,165(B):709-726.
- [8] TAN M, YUAN S P, LI S H, et al. Ultra-short-term industrial power demand forecasting using LSTM based hybrid ensemble learning[J]. IEEE transactions on power systems,2020,35(4):2937-2948.
- [9] RAMOS D, FARIA P, VALE Z, et al. Short time electricity consumption forecast in an industry facility[J]. IEEE transactions on industry applications,2022,58(1):123-130.
- [10] CHEN K J, CHEN K L, WANG Q, et al. Short-term load forecasting with deep residual networks[J]. IEEE transactions on smart grid,2018,10(4):3943-3952.
- [11] 赵腾,王林童,张焰,等.采用互信息与随机森林算法的用户用电关联因素辨识及用电量预测方法[J].中国电机工程学报,2016,36(3):604-614.
- [12] 谭津,邓长虹,杨威,等.微电网光伏发电的 Adaboost 天气聚类超短期预测方法[J].电力系统自动化,2017,41(21):33-39.
- [13] WANG Y Y, SUN S F, CHEN X Q, et al. Short-term load forecasting of industrial customers based on SVM and Xgboost[J]. International journal of electrical power and energy systems,2021,129(1):106830.
- [14] CHENG R Y, YU J Q, ZHANG M. Short-term hybrid forecasting model of ice storage air-conditioning based on improved SVR[J]. Journal of building engineering,2022,50:104194.
- [15] 顾海艳,曹林,朱涛,等.神经网络二次集成预测算法研究[J].南京师大学报(自然科学版),2022,45(2):136-141.
- [16] ZHANG Q, WU J J, MA Y, et al. Short-term load forecasting method with variational mode decomposition and stacking model fusion[J]. Sustainable energy grids and networks,2022,30:100622.
- [17] LIU Y H, NA X, YIN C C, et al. 3-D Joint inversion of airborne electromagnetic and magnetic data based on local Pearson correlation constraints[J]. IEEE transactions on geoscience and remote sensing,2022,60(7):5910813.

[责任编辑:陆炳新]