

基于 K -折交叉验证和 Stacking 融合的短期负荷预测

朱文广, 李映雪, 杨为群, 刘小春, 熊 宁, 周 成, 王 丽

(国网江西省电力有限公司经济技术研究院, 江西 南昌 330000)

摘 要:短期负荷预测对于电力系统的经济调度和稳定运行具有重要意义。为了提升短期负荷预测的精度,提出基于 K -折交叉验证和 Stacking 融合的短期负荷预测方法。首先,基于皮尔逊相关系数对影响短期负荷的多个特征进行筛选,剔除冗余特征。其次,利用 K -折交叉验证法训练第一层的各个子模型,并将各个子模型的预测结果作为新特征用于训练第二层模型。接着,将子模型的结果进行 Stacking 融合,使用第二层的模型得到短期负荷的预测结果。最后,采用新英格兰的实际数据验证所提方法的有效性。仿真结果表明,所提的 K -折交叉验证法能够有效地提高模型的泛化能力,Stacking 融合不仅可以提升预测的平均精度,还可以减小最大的预测误差,比单一模型预测更具优势。

关 键 词:短期负荷预测;皮尔逊相关系数; K -折交叉验证;Stacking 融合

DOI:10.19781/j.issn.1673-9140.2021.01.010 中图分类号:TM614 文章编号:1673-9140(2021)01-0087-09

Short-term load forecasting based on the K -fold cross-validation and stacking ensemble

ZHU Wenguang, LI Yingxue, YANG Weiqun, LIU Xiaochun,
XIONG Ning, ZHOU Cheng, WANG Li

(Economic and Technological Research Institute, State Grid Jiangxi Electric Power Company, Nanchang 330000, China)

Abstract: Short-term load forecasting is of great significance for the economic dispatching and operation of power systems. In order to improve the accuracy of short-term load forecasting, a short-term load forecasting method based on k -fold cross validation and Stacking ensemble is proposed. Firstly, the Pearson coefficient method is utilized to screen multiple features affecting short-term load, and redundant features are eliminated. Secondly, the k -fold validation cross-validation method is applied to train the sub-models of the first level, and the prediction results of each sub-model are taken as new features to train the second level model. Thirdly, the results of the sub-model are stacked, and the short-term load forecasting results obtained by the second layer model. Finally, the validity of the proposed method is verified by the actual data set from New England. The simulation results show that the proposed k -fold cross-validation method can effectively improve the generalization ability of the model. Stacking fusion can not only improve the average accuracy of prediction, but also reduce the maximum prediction error, which is more advantageous than single model prediction.

Key words: short-term load forecasting; pearson correlation coefficient; K -fold cross validation; stacking ensemble

收稿日期:2019-01-15;修回日期:2019-05-06

基金项目:国网公司总部科技项目(SGTYHT/16-JS-198)

通信作者:李映雪(1982-),男,博士,高级工程师,主要从事电力负荷预测和电网智能化技术研究;E-mail:jxliyingxue@126.com

负荷预测指的是利用收集的运行参数、环境条件、社会因素等参数,在满足一定精度的前提下,确定未来某个时刻的电力需求量。准确的负荷预测不仅可以指导用电方定制合理的用电计划,还有助于发电方构造最优的电量、电价投标策略。根据预测时间尺度的不同,负荷预测可分为超短期预测、短期预测、中期预测和长期预测,相对应的时间分辨率是分钟、小时、月以及年。

长期以来,国内外学者关于短期负荷预测方面展开了大量的研究,并取得了显著的成果。已有的方法主要包括时间序列法^[1]、灰色预测法^[2]、支持向量机(support vector machine, SVM)^[3]、深度神经网络(deep neural network, DNN)^[4]和光梯度提升机(light gradient boosting machine, LightGBM)^[5]等方法。其中,时间序列法的模型比较简单,仅能对少量影响负荷的因素以及样本数据进行处理,要求历史负荷序列平稳性较高。灰色预测理论中以灰色GM(1,1)应用最为广泛,不需要大量的样本,计算工作量小。但灰色预测要求原始离散数据是光滑离散函数,且不能考虑负荷和多种不同影响因素的内在联系。SVM的基础理论比较完善,能够解决传统方法中存在非线性、收敛于局部最优和维数高的问题,但对于大量的样本数据会有计算速度慢、预测精度低的问题。相对于传统的浅层学习,DNN具有更强大的非线性映射能力和柔化的网络结构,能够根据负荷预测的具体情况任意设定隐含层数和每层的神经元的个数,但存在着学习速度慢,容易陷入局部最优的缺陷。LightGBM是基于决策树模型的Boosting算法,在模型的训练速度和内存方面都比传统的XGBoost更具优势,但在一定程度上容易出现过拟合。

随着智能电表、智能传感器等新型电力电子设备的普及,数据库中的数据在不断地积累和增加,大数据分析的时代已经悄然来临^[6-7]。机器学习是从海量数据中挖掘有价值信息的利器,而模型融合技术是机器学习中提升数据分类和回归准确率的有效策略。目前,模型融合技术已经在变压器故障诊断、风机状态估计、电网暂态频率特征预测等多个领域得到了成功的应用^[8-9]。然而,模型融合技术在电力

系统短期负荷预测方面的研究还相对有限。如何有效地提取各个模型的优势,来提升短期负荷预测的精度还有待进一步的研究。

为了克服多层感知机(multi-layer perceptron, MLP)、长短期记忆网络(long short-term memory, LSTM)、向量空间模型(vector space model, VSM)以及LightGBM 4种单一模型存在预测精度不高的缺陷,提出了基于K-折交叉验证和Stacking融合的短期负荷预测方法。该文首先基于皮尔逊相关系数(pearson correlation coefficient, PCC)对影响短期负荷的多个特征进行筛选;其次采用K-折交叉验证法分别对4种模型中的任意3种进行训练,并将预测结果作为特征值对剩余模型进行训练;接着利用Stacking融合策略将各子模型结果进行融合;最后,采用新英格兰的实际数据验证所提方法的有效性。

1 K-折交叉验证和 Stacking 融合技术

针对单一模型存在不同的缺陷,拟采用Stacking融合技术实现各个算法的取长补短,以提高短期负荷预测的精度。如图1所示,通过原始特征利用K-折交叉验证法训练第1层的3个子模型。然后,以3个子模型的预测值作为新特征训练第2层的模型,并由第2层模型输出测试集的最终预测结果。

对于各个子模型的训练拟采用K-折交叉法以提升模型的泛化能力。如图2所示,以K-折交叉验证法训练LSTM为例展开说明。K-折交叉验证法是将原始LSTM训练集分成k组,然后以其中1组作为验证集,剩下的k-1组作为训练集得到一个预

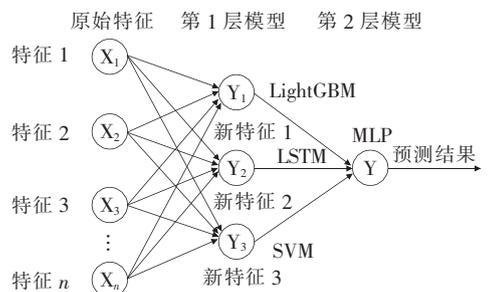


图1 Stacking融合技术的原理

Figure 1 The principle of Stacking ensemble technique

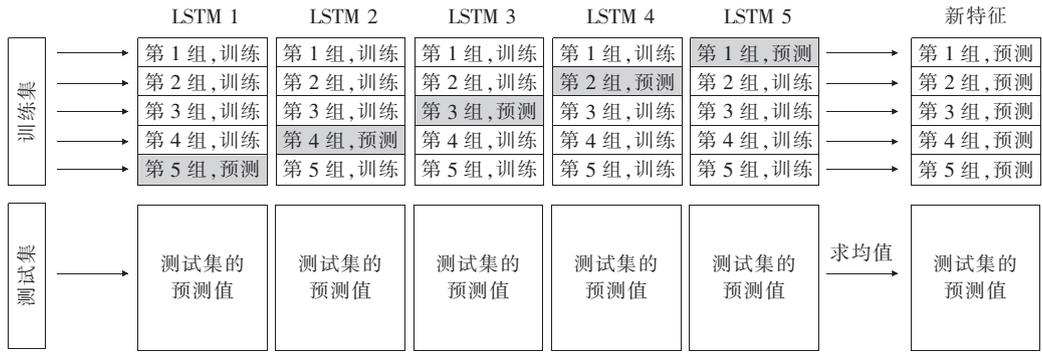


图 2 K-折交叉验证的原理

Figure 2 The principle of K-fold crosses validation

测模型,这个训练好的模型对测试集的数据进行预测。重复上述过程 k 次,得到 k 组验证集的预测结果,即原始训练集的预测值,他们将作为第 2 层模型的特征。另外,将 k 次测试集的预测值求均值作为测试集的最终预测结果。

2 预测特征的相关分析

影响短期负荷预测的因素有很多,包括日期类型、温度、天然气价格、原油价格、煤炭价格、历史电价和历史负荷等。以自回归积分滑动平均模型(auto-regressive integrated moving average model, ARIMA)和灰色 GM(1,1)为代表的部分已有方法仅依靠历史数据的趋势来预测短期负荷,而不考虑多种外部特征,这类方法明显存在不合理之处,难以适应环境的突变性,尤其对于拐点处会有较大的误差。考虑多种特征对短期负荷的影响有利于提高预测的精度,但如果把与短期负荷不相关的特征放进模型中,不仅会增加算法的复杂度,还会干扰模型的参数。因此,在训练模型前,有必要进行特征的提取。

特征的提取是机器学习中常用的数据处理技术,已有的短期负荷预测方法往往依靠经验选择主要特征。这种方法具有一定的主观性,不同时间和空间下影响短期负荷预测的特征也不一定是相同的。因此,需要采用定量的方法找到影响短期负荷的主要特征。

的相关性。相关系数越接近 1,相关性越强;相关系数越接近 0,相关性越弱。2 个特征 x 和 y 之间的 Pearson 相关系数的数学表达式^[9]:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

式中 \bar{x} 和 \bar{y} 分别为 x 和 y 的均值。

新英格兰实际数据中包含了 2004 年 1 月到 2009 年 12 月每个小时的数据。特征主要有:时刻、干球温度、露点温度、天然气价格、电价、日前价格的能源组成部分、阻塞价格、价格网损、节点边际电价、边际损耗价格和调频市场出清价格。以上述特征为基础,采用 Pearson 相关系数法对各个特征进行数据处理,得到结果表 1 所示。

从表 1 可以看出,电价对于短期负荷的影响最大,其次是日前价格的能源组成部分、时刻、节点边际电价等。考虑到露点温度、天然气价格和调频市场出清价格的 Pearson 相关系数较小,在构建模型

表 1 各个指标的 Pearson 相关系数

Table 1 Pearson correlation coefficient of each index

指标	系数	指标	系数
时刻	0.51	日前价格的能源组成部分	0.58
干球温度	0.27	阻塞价格	-0.29
露点温度	0.05	价格网损	0.33
天然气价格	0.08	节点边际电价	0.51
电价	0.66	边际损耗价格	0.13
调频市场出清价格	-0.01	—	—

Pearson 相关系数可以用来衡量 2 个特征之间

时将不考虑这几种特征。

当前时刻的用电量和前几个时刻的负荷具有一定的相关性。除了环境和电价因素外,输入的特征还应该考虑历史负荷序列。理论上历史序列长度可以是任意的,如果历史负荷序列太短会导致对历史知识学习的缺失,达不到学习前期负荷序列信息的目的。反之,历史负荷序列过长会导致增加算法复杂度甚至使得预测效果变差。因此,为了确定最优的历史负荷序列步长,需计算当前时刻 T 的负荷与 $T-1, T-2, \dots, T-15$ 时刻负荷之间的 Pearson 相关系数。

如表 2 所示,当前时刻的负荷和过去 15 h 的负荷之间的时间相关性随着时间的增加而减弱。从 $T-1$ 到 $T-7$ 的皮尔逊系数始终大于 0.1,这说明当前时刻的负荷和过去 7 个小时的负荷之间具有较强的时间相关性,故设历史负荷序列步长为 7。

表 2 历史负荷的皮尔逊系数

Table 2 Pearson correlation coefficient of historical loads

历史负荷	系数	历史负荷	系数	历史负荷	系数
$T-1$	0.96	$T-6$	0.24	$T-11$	-0.10
$T-2$	0.86	$T-7$	0.12	$T-12$	-0.11
$T-3$	0.72	$T-8$	0.02	$T-13$	-0.10
$T-4$	0.56	$T-9$	-0.04	$T-14$	-0.09
$T-5$	0.40	$T-10$	-0.08	$T-15$	-0.06

3 模型的原理和流程

3.1 LSTM 的网络结构和训练机制

LSTM 网络在 1997 年由 Sepp Hochreiter 和 Jürgen Schmidhuber 提出^[10]。LSTM 网络是通过控制输入、遗忘和输出 3 个门来实现对记忆单元的读取和修改,然后使用不同的函数计算隐含层状态。实践证明,LSTM 能够有效地处理历史序列信息和时间的依赖关系,已经应用于语音预测、电价预测、风功率预测和光伏出力预测等不同领域。图 3 为 LSTM 的记忆单元结构示意图,显示了结构单元中各个门之间的关系以及相邻时刻 LSTM 网络之间的相互作用。

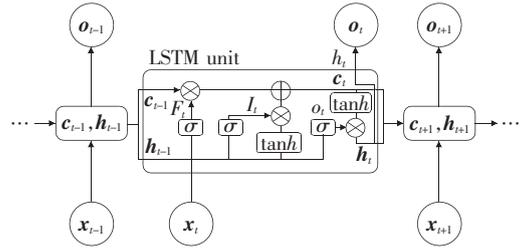


图 3 LSTM 的结构单元

Figure 3 The structure unit of LSTM

每一个时刻的输入特征有 3 个,分别是 LSTM 单元的输入向量 x_t 、神经元状态向量 c_{t-1} 和 LSTM 单元的输出向量 h_{t-1} 。每个门根据不同的输入特征进行逻辑运行和非线性变换得到输出值,各个特征的关系如下^[11]:

$$\begin{aligned}
 f_t &= \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \\
 i_t &= \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \\
 \begin{cases} o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \\ c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \end{cases} & \quad (2) \\
 h_t &= o_t \circ \sigma_c(c_t)
 \end{aligned}$$

式中 W 和 U 为权重矩阵; b 为偏置向量,这些参数在训练阶段会被动态调整; f_t 、 i_t 和 o_t 分别为遗忘门的激活向量、输入的激活向量和输出门的激活向量; σ_g 和 σ_c 分别为 sigmoid、hyperbolic tangent 函数。

确定好 LSTM 的神经网络的结构以后,采用按时间展开的反向误差传播算法^[9]对网络进行训练,该文以平均绝对百分误差(mean absolute percent error, MAPE)作为损失函数^[12]。

3.2 LightGBM 的原理

LightGBM 是微软 DMTK 团队在 2017 年提出的一种新的梯度提升框架。它是将弱分类的算法提升成强分类算法,从而在一定程度上提高分类或回归的准确率。目前,LightGBM 主要被用于分类^[13]和回归^[14-15]等机器学习领域。

相对于传统的梯度提升算法,LightGBM 主要包括互斥特征捆绑和基于梯度的单边采样技术。对于一个给定的有监督训练集 $X = \{(x_i, y_i)\}, i = 1, 2, \dots, n$, LightGBM 的目标是找到一个近似于某个函数 $\hat{f}(x)$, 该函数将指定的损失函数 $L(y, f(x))$ 的期望值最小化,即

$$\hat{f} = \arg \min_f E_{y,x} L(y, f(x)) \quad (3)$$

LighGBM 集成了许多回归树来近似最终模型,即

$$f_T(X) = \sum_{i=1}^T f_i(X) \quad (4)$$

式中 T 为回归树的个数。回归树可以表示为

$$w_{q(x)}, q \in \{1, 2, \dots, J\} \quad (5)$$

式中 J 为叶子数; q 为树的决策规则; w 为叶节点样本权重的向量。因此,LightGBM 将通过加法模型进行训练,在步骤 t 时有:

$$\Gamma_t = \sum_{i=1}^n L(y_i, F_{t-1}(x_i) + f_t(x_i)) \quad (6)$$

LightGBM 的目标函数用牛顿法快速逼近。为了简单起见,去掉式(6)中的常数项后,可以将公式转换为

$$\Gamma_t \cong \sum_{i=1}^n (g_i, f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)) \quad (7)$$

式中 g_i 和 h_i 分别为损失函数的一阶和二阶梯度。若 I_j 表示叶子 j 的样本集,则式(7)转换为

$$\Gamma_t = \sum_j \left(\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right) \quad (8)$$

对于一个确定的树结构 $q(x)$,各叶节点的最优权重 w_j^* 和 Γ_K 的极值可表示为

$$w_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (9)$$

$$\Gamma_T^* = - \frac{1}{2} \sum_{j=1}^J \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} \quad (10)$$

式中 Γ_j^* 可以被看成是用于衡量决策树结构质量的评分函数,最后目标函数可以表示为

$$G = - \frac{1}{2} \left(\frac{\left(\sum_{i \in I_L} g_i \right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i \right)^2}{\sum_{i \in I_R} h_i + \lambda} + \frac{\left(\sum_{i \in I} g_i \right)^2}{\sum_{i \in I} h_i + \lambda} \right) \quad (11)$$

式中 I_L 和 I_R 分别为样本集的左分支和右分支。

3.3 SVM 的原理

SVM 是在 20 世纪 90 年代后提出并得到快速发展的统计学算法,具有理论完善、全局优化、泛化能力和适应性强的特征。支持向量回归机将二次优

化问题转化为一个线性方程组的求解问题,在电力的风机和光伏出力预测中表现出较好的性能^[16]。

对于给定的样本集:

$$\begin{cases} z = \{(x_1, y_1), \dots, (x_t, y_t)\} \\ x_i \in R^n, y_i \in R, i = 1, \dots, t \end{cases} \quad (12)$$

这个训练样本的最小二乘向量机回归可以表示为

$$\begin{cases} \min J = \frac{1}{2} \mathbf{w} \mathbf{w}^T + C \sum_{i=1}^t e_i^2 \\ [s, ty_i = \mathbf{w} \vartheta(x_i) + \mathbf{b} + e_i \end{cases} \quad (13)$$

式中 C 为惩罚系数; \mathbf{b} 为偏置向量; e_i 为误差; \mathbf{w} 为超平面的法向量。目标函数的拉格朗日函数可以表示为

$$L = \frac{1}{2} \mathbf{w} \mathbf{w}^T + C \sum_{i=1}^t e_i^2 - \sum_{i=1}^t \alpha_i [\mathbf{w} \vartheta(x_i) + \mathbf{b} + e_i - y_i] \quad (14)$$

对该函数求极值,可以将其变换成线性方程组的求解问题:

$$\begin{bmatrix} 0 & 1 & \dots & 1 & & \\ 1 & k(x_1, x_t) + 1/C & \dots & k(x_1, x_t) & \alpha_1 & y_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & k(x_t, x_1) & \dots & k(x_t, x_t) & \alpha_t & y_t \end{bmatrix} \begin{bmatrix} 1 \\ \alpha_1 \\ \vdots \\ \alpha_t \end{bmatrix} = \begin{bmatrix} 1 \\ y_1 \\ \vdots \\ y_t \end{bmatrix} \quad (15)$$

式中 $k(x_i, x_j)$ 为核函数,该方程组的解是 t 个样本的回归系数 α 和偏置向量 \mathbf{b} , k 时刻的预测值可以表示为

$$P(x_k) = \sum_{i=1}^t \alpha_i k(x_k, x_i) + \mathbf{b} \quad (16)$$

3.4 MLP 的原理

MLP 是一种由全连接层组成的前向结构的人工神经网络,是目前应用最广泛的神经网络模型之一^[17]。如图 4 所示,一个典型 MLP 的网络结构包括输入层、隐含层和输出层。它可以被看成是一个有向图,由多个节点层组成,每一层全连接到下一层。除了输入节点,每个节点都是一个带有非线性激活函数的神经元。一种被称为反向传播算法的监督学习方法常被用来训练 MLP。MLP 是感知器的推广,克服了感知器无法实现对线性不可分数据识别的缺点。

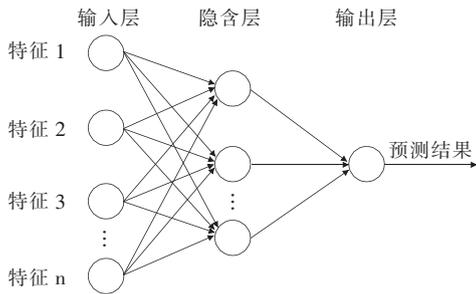


图 4 MLP 的网络结构

Figure 4 Network structure of MLP

3.5 短期负荷预测的流程

综合上述分析,基于 K -折交叉验证和 Stacking 融合的短期负荷预测方法的计算步骤可以表示如下。

1) 将输入的影响短期负荷预测的特征整理为合理的数字表达形式,并进行预处理,即使用最小最大标准化方法将数据映射到 $[0, 1]$ 区间。

2) 根据输入数据,计算各个原始特征和短期负荷的皮尔逊系数,并选出具有较大皮尔逊系数的特征作为自变量用来训练后续的子模型。

3) 反复试探和调试 LSTM 网络、SVM、MLP 以及 LightGBM 的结构和参数,直到找到最佳的参数。

4) 从 LSTM 网络、SVM、MLP 以及 LightGBM 中选取任意 3 种作为第 1 层模型,剩余一种作为第 2 层模型。利用 K -折验证交叉法训练好第 1 层中各个子模型,并将各个子模型的预测结果作为新特征训练第 2 层模型。

5) 输入测试集的数据,并将子模型的结果进行 Stacking 融合,使用第 2 层的模型得到预测的短期负荷,并统计相关误差指标。

4 算例分析

4.1 数据描述和试验环境

为了充分验证该文所提算法的有效性,将以均方根误差 (root mean square error, RMSE)、平均绝对误差 (mean absolute error, MAE) 以及平均绝对百分比误差 (mean absolute percentage error, MAPE) 作为衡量指标^[18],在新英格兰实际数据上进行仿真^[19]。设定 2004 年 1 月到 2008 年 10 月的数据作为训练集,2008 年 11 月到 2009 年 5 月的数

据作为验证集,2009 年 6 月到 2009 年 12 月的数据作为测试集。所有算法基于 Tensorflow 1.8 框架下实现,电脑环境配置为: Intel(R) Core(TM) i7-7700 M 3.60-GHz processor and 8 GB of RAM。

经过反复的调试和试探确定各个算法的最佳参数,第 1 层有 3 个子模型: LightGBM 的树深是 6,每棵树叶子是 30 片,学习率 0.1,每棵树随机采样率是 0.8,每片叶子上数据的最小数量为 90,训练次数是 16 000。目标函数是 regression; SVM 采用径向基函数作为核函数; LSTM 的输入层和输出层都是全连接层,输入层的神经元个数和输入特征个数相等,输出层的神经元个数是 1。隐含层由 3 个 LSTM 层构成,神经元个数分别是 32、18 和 10。第 2 层的模型是多层感知机,隐含层个数为 3 层,每层的神经元个数依次是 15、15 和 20。

4.2 K -折交叉验证的影响

为了验证 K -折交叉验证法对于模型训练的有效性,以多层感知机为例,设置 k 值从 0 到 14 之间变化,变化的步长是 2。每种情况都重复试验 50 次求得统计结果如表 3 所示。

表 3 不同 k 值对应的预测结果Table 3 Forecasting results in different k values

训练集误差			测试集误差			k 值
RMSE	MAE	MAPE/%	MAE	RMSE	MAPE/%	
162.2	114.9	0.76	158.1	120.2	0.74	0
202.0	142.7	0.94	166.9	126.1	0.77	2
195.3	137.1	0.90	160.7	121.8	0.74	4
191.8	133.8	0.88	159.7	120.3	0.74	6
185.9	130.6	0.86	159.7	120.8	0.74	8
181.3	127.3	0.84	157.9	119.1	0.73	10
180.4	126.5	0.84	156.9	118.1	0.72	12
179.4	125.8	0.83	154.2	117.3	0.71	14

从表 3 可以看出: ① 相对于不分组的情况而言,分组后训练集的误差普遍有所增加。另外,随着 k 值的增加,训练集的误差逐渐减小。这是因为不分组或者 k 值较大时会有更多的样本参与训练模型,再用训练好的模型去预测训练集,误差自然就比较小。② 就测试集的误差而言,随着 k 值的增加,测

试集的误差也逐渐减小。对比未分组和 k 值为 12 的情况, k 为 12 的训练集误差虽然比未分组的情况大,但测试集的误差反而比未分组的情况小,这说明增大 k 值有利于提高模型的泛化能力。

图 5 为 k 值在不同取值时多层感知机单次的训练时间,从中可以看出虽然增大 k 值有利于提高模型的泛化能力,但也导致了计算时间的增加,因此在选择 k 值的时候要权衡计算的精度和耗时,该文中后续的计算模型统一 k 值取 10。

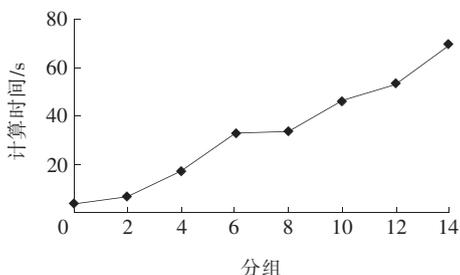


图 5 不同 k 值对应的训练时间

Figure 5 Training time corresponding to k value

4.3 单一模型的预测

在进行 Stacking 模型融合前,有必要单独对 4 个子模型进行短期负荷预测的性能测试试验。为保证结果的客观性,每种方法都进行独立重复试验 50 次,并统计误差指标如表 4 所示,图 6 为随机选中某天 24 小时负荷曲线的预测情况。

通过对测试集的误差统计发现,LightGBM、LSTM、MLP 以及 SVM 的 MAPE 值分别为 0.72%、0.68%、0.73% 以及 0.87%。从测试样本的平均误差来看,LSTM 的预测精度最高,LightGBM 和 MLP 的预测性能很接近,SVM 的表现较差。从图 6 可以看出,尽管 LightGBM 和 MLP 的平均误差都比 SVM 的小,但在某些时刻会出现较大的误差。

表 4 不同模型的预测结果

Table 4 Prediction results of different models

模型	训练样本			测试样本		
	RMSE	MAE	MAPE/%	RMSE	MAE	MAPE/%
LighGBM	137.2	96	0.64	153.8	116.7	0.72
LSTM	127	94.9	0.62	141.3	110.5	0.68
MLP	181.3	127.3	0.84	157.9	119.1	0.73
SVM	149	107	0.73	178.1	140.4	0.87

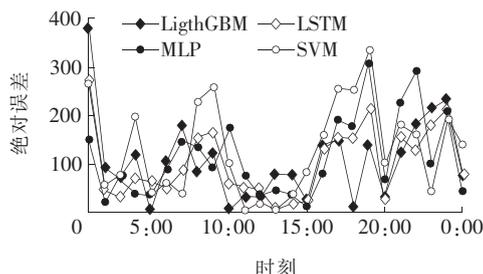


图 6 各个模型的绝对误差

Figure 6 The absolute forecasting errors of different models

4.4 Stacking 融合后的预测

为了充分测试该文所提算法的性能,依次选择 LightGBM、LSTM、MLP 和 SVM 中的其他 3 个模型作为 Stacking 融合的第 1 层,剩下的一个模型作为第 2 层。4 种方案的误差统计如表 5 所示。

表 5 4 种方案的测试集误差

Table 5 Test set error of four schemes

第 1 层	第 2 层	RMSE	MAE	MAPE/%
LSTM, MLP, SVM	LighGBM	129.1	98.7	0.60
LighGBM, MLP, SVM	LSTM	122.5	92.5	0.56
LSTM, SVM, LighGBM	MLP	155.4	121.3	0.75
LSTM, MLP, LighGBM	SVM	145.1	113.8	0.71

从表 5 可以看出:①4 个子模型在 Stacking 融合时所在的层的位置对于短期负荷预测精度有着较大的影响。MLP 在单一模型预测中表现较好,在进行模型融合后误差反而增大,表明基于 Stacking 模型融合不绝对保证融合模型的预测精度一定高于单一的模型,融合模型的预测精度受到融合的结构、子模型性能以及位置的影响。②单一模型预测的最小 MAPE 为 0.68%,前 2 种融合模型的 MAPE 分别为 0.60% 和 0.56%,这表示通过 Stacking 融合多种模型能够有效地较小预测误差,提高预测精度。

图 7 为各个单一模型和 Stacking 融合模型的关于测试集的最大 MAPE 值,融合模型 1~4 分别对应表 5 中 2~5 行的模型。就单一模型而言,LightGBM 和 LSTM 的平均 MAPE 虽然比另外 2 个模型小,但最大的 MAPE 却比 MLP 和 SVM 大。从融合模型 1 和 2 可以看出,通过 Stacking 模型融合不仅提升了预测的平均精度,而且还减小了最大预测误差。

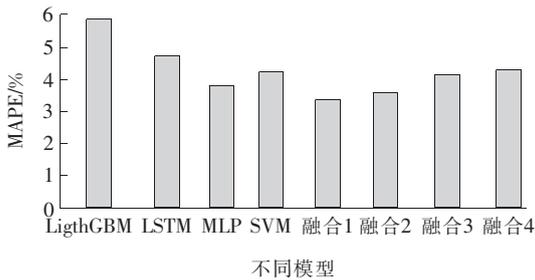


图7 各个模型的最大 MAPE

Figure 7 The maximal MAPE of different models

5 结语

为了提升短期负荷预测的精度,提出了基于 K -折交叉验证和 Stacking 融合的短期负荷预测方法。首先,基于皮尔逊相关系数对影响短期负荷的多个特征进行筛选,剔除冗余特征。其次,利用 K -折交叉验证法训练第 1 层的各个子模型,并将各个子模型的预测结果作为新特征用于训练第 2 层模型。接着,将子模型的结果进行 Stacking 融合,使用第 2 层的模型得到短期负荷的预测结果。通过新英格兰的实际数据集仿真表明:

1) 利用 K -折交叉验证法训练短期负荷预测的模型有利于提升泛化能力,但 k 值的选取不宜过大,还要考虑计算时间。

2) 基于 LighGBM、LSTM、MLP 以及 SVM 的单一模型负荷预测具有不同特点,虽然 LighGBM 和 MLP 的平均预测误差较小,但最大的预测误差较大。

3) 基于 Stacking 融合模型不仅提升了预测的平均精度,而且还减小了最大的预测误差,比单一模型预测更具优势。

参考文献:

[1] 吴佳懋,李艳,符一健. 基于粗糙集-混沌时间序列 Elman 神经网络的短期用电量预测[J]. 电力系统保护与控制,2019,47(3):23-30.
WU Jiamao, LI Yan, FU Yijian. Short-term power consumption prediction based on rough set chaotic time series Elman neural network[J]. Power System Protection and Control, 2019, 47(3): 23-30.

[2] 韩俊,韩文花,陈曦,等. 基于熵权-灰色关联分析的降低区域停电损失举措研究[J]. 中国电力,2020,53(5):10-17.
HAN Jun, HAN Wenhua, CHEN Xi, et al. Research on reducing regional outage loss based on entropy weight-gray correlation analysis [J]. Electric Power, 2020, 53(5): 10-17.

[3] 刘伟鹏,张国钢,刘亚魁,等. 基于主成分分析和支持向量机的高压断路器机械状态识别方法[J]. 高压电器,2020,56(9):267-272+278.
LIU Weipeng, ZHANG Guogang, LIU Yakui, et al. Mechanical status identification of high voltage circuit breakers based on principal component analysis and support vector machines [J]. High Voltage Apparatus, 2020, 56(9): 267-272+278.

[4] 吴迪,汤小兵,李鹏,等. 基于深度神经网络的变电站继电保护装置状态监测技术[J]. 电力系统保护与控制,2020,48(5):81-85.
WU Di, TANG Xiaobing, LI Peng, et al. State monitoring technology of substation relay protection device based on deep neural network[J]. Power System Protection and Control, 2020, 48(5): 81-85.

[5] 刘羽霄,张宁,康重庆. 数据驱动的电力网络分析与优化研究综述[J]. 电力系统自动化,2018,42(6):157-167.
LIU Yuxiao, ZHANG Ning, KANG Chongqin. A review on data-driven analysis and optimization of power grid [J]. Automation of Electric Power Systems, 2018, 42(6): 157-167.

[6] 李俊卿,王焕仲,季刚,等. 基于大数据分析的风机轴承故障预警[J]. 智慧电力,2020,48(2):25-30.
LI Junqing, WANG Huanzhong, JI Gang, et al. Fault early warning of wind turbine bearing based on big data analysis[J]. Smart power, 2020, 48(2): 25-30.

[7] 石蓉,杨海波,李永毅,等. 基于同期售电量大数据处理技术的分行业用电分析与预测[J]. 电网与清洁能源,2019,35(8):42-48.
SHI Rong, YANG Haibo, LI Yongyi, et al. Analysis and prediction of power consumption in different industries based on large data processing technology of the same-period electricity sales[J]. Power Grid and Clean Energy, 2019, 35(8): 42-48.

[8] 刘华新,苑一鸣,周沛,等. 基于融合理论的风电机组状态评价正态云模型[J]. 太阳能学报,2018,39(10):2891-2900.

- LIU Huaxin, YUAN Yiming, ZHOU Pei, et al. Normal cloud model for condition evaluation of wind turbines based on fusion theory[J]. *Acta Energetica Solaris Sinica*, 2018, 39(10): 2891-2900.
- [9] 朱乔木, 李弘毅, 王子琪, 等. 基于长短期记忆网络的风电场发电功率超短期预测[J]. *电网技术*, 2017, 41(12): 68-73.
- ZHU Qiaomu, LI Hongyi, WANG Zhiqi, et al. Short-term wind power forecasting based on LSTM[J]. *Power System Technology*, 2017, 41(12): 68-73.
- [10] Hochreiter S, Schmidhuber J. Long short-term memory [J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [11] 李冰, 张妍, 刘石. 基于 LSTM 的短期风速预测研究 [J]. *计算机仿真*, 2018, 35(11): 456-461.
- LI Bing, ZHANG Yan, LIU Shi. Wind speed short term prediction study based on LSTM[J]. *Computer Simulation*, 2018, 35(11): 456-461.
- [12] 杨甲甲, 刘国龙, 赵俊华, 等. 采用长短期记忆深度学习模型的工业负荷短期预测方法[J]. *电力建设*, 2018, 39(10): 20-27.
- YANG Jiajia, LIU Guolong, ZHAO Junhua, et al. A long short term memory based deep learning method for industrial load forecasting[J]. *Electric Power Construction*, 2018, 39(10): 20-27.
- [13] Wang D H, Zhang Y, Zhao Y. LightGBM: an effective miRNA classification method in breast cancer patients [C]//Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics. New York, USA, 2017.
- [14] Zhang W, Quan H, Srinivasan D. Parallel and reliable probabilistic load forecasting via quantile regression forest and quantile determination [J]. *Energy*, 2018, 160: 810-819.
- [15] Ma X, Sha J, Wang D, et al. Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning [J]. *Electronic Commerce Research and Applications*, 2018, 31: 24-39.
- [16] 刘纯, 范高锋, 王伟胜, 等. 风电场输出功率的组合预测模型[J]. *电网技术*, 2009, 33(13): 74-79.
- LIU Chun, FAN Gaofeng, WANG Weisheng, et al. A combination forecasting model for wind farm output power [J]. *Power System Technology*, 2009, 33(13): 74-79.
- [17] Bulhoes J S, Martins C L, Reis M R C, et al. Predicting system for flooded areas developed with MLP-type neural network // [C] 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC/ICEE/ICPS Europe). Palermo, Italy: IEEE, 2018.
- [18] 周海, 李登宣, 尹万思, 等. 基于极限学习机的光伏发电短期预测校正方法[J]. *电网与清洁能源*, 2020, 36(6): 64-69+77.
- ZHOU Hai, LI Dengxuan, YIN Wansi, et al. Short-term forecasting correction method of photovoltaic power based on extreme learning machine [J]. *Power System and Clean Energy*, 2020, 36(6): 64-69+77.
- [19] Aggarwal S K, Saini L M, Kumar A. Electricity price forecasting in deregulated markets: a review and evaluation [J]. *International Journal of Electrical Power & Energy Systems*, 2009, 31(1): 13-22.