

DOI: 10.19783/j.cnki.pspc.201485

# 基于 XGBoost 的配电网线路峰值负荷预测方法

姜建<sup>1</sup>, 刘海琼<sup>2</sup>, 李衡<sup>1</sup>, 赵斌炎<sup>1</sup>, 鲍威<sup>2</sup>, 郑梦娜<sup>2</sup>

(1. 国网浙江省电力有限公司杭州供电公司, 浙江 杭州 310020;

2. 杭州沃瑞电力科技有限公司, 浙江 杭州 310012)

**摘要:** 为实现电网平稳迎峰度夏, 需要在夏季负荷高峰前提前 1~2 个月对配电网线路进行峰值负荷预测, 为设备部门有计划地制订和实施增容和改扩建方案提供数据支撑。提出一种基于 XGBoost 的配电网线路峰值负荷预测方法。该方法综合考虑气象因素、时间因素、春季基础负荷因素, 分析各类因素与夏季负荷高峰的相关性, 确定预测样本特征值。通过 K-means 算法对线路负荷增长趋势进行聚类分析, 筛选出未来可能负荷较重的目标线路, 进而使用 XGBoost 算法进行线路峰值负荷预测。使用所提方法对某实际城区局部配网进行预测, 算例结果验证了该算法的预测准确性。与其他算法的对比结果体现了该算法计算规模小、预测速度快的优点。

**关键词:** 负荷预测; 电力峰值负荷; XGBoost; K-means 聚类; 配电网

## Peak load forecasting method of distribution network lines based on XGBoost

JIANG Jian<sup>1</sup>, LIU Haiqiong<sup>2</sup>, LI Heng<sup>1</sup>, ZHAO Binyan<sup>1</sup>, BAO Wei<sup>2</sup>, ZHENG Mengna<sup>2</sup>

(1. Hangzhou Power Supply Company, State Grid Zhejiang Electric Power Co., Ltd., Hangzhou 310020, China;

2. Hangzhou Worui Power Technology Co., Ltd., Hangzhou 310012, China)

**Abstract:** To enable the power grid to smoothly pass through the summer load peak, it is necessary to predict the peak load of distribution network lines 1-2 months in advance before the summer load peak. This provides data support for the equipment department to develop and implement capacity expansion and reconstruction projects in a planned manner. This paper proposes a distribution network line peak load forecasting method based on XGBoost. The method comprehensively considers various factors including meteorological, time and spring load factors, and analyzes the correlation coefficients between summer load and various factors, and accordingly determines the eigenvalues of prediction samples. The K-means algorithm is used to cluster the line load growth trends to identify the target lines with heavy load in the future. The XGBoost algorithm is used to predict the peak load of the lines. The predice results on an actual urban distribution network verify the prediction accuracy of method. Comparisons with other methods demonstrate that the proposed method has the advantages of smaller computation scale and faster prediction speed.

This work is supported by the Science and Technology Project of State Grid Corporation of China "Load Forecast and Promotion of Distribution Network Power Supply Potential Based on Ubiquitous Power Internet of Things and Big Data Analysis Application" (No. 5211HZ19014P).

**Key words:** load forecasting; electrical peak load; XGBoost; K-means clustering; distribution network

## 0 引言

夏季用电高峰时段, 城市配电网供电形势较为严峻, 若出现线路重过载现象, 将对配电网构成冲

击。为平稳迎峰度夏, 在每年夏季到来之前, 配网管理人员通常提前 1~2 个月对配电网线路进行负荷峰值预测, 准确识别夏季将出现重过载现象的线路, 并预测其峰值负荷大小, 这对于电力部门有计划地制订和实施电网规划、增容和改扩建方案至关重要<sup>[1]</sup>。配电网线路通常多达数千条, 预测配电网线路未来 1~2 个月后的夏季负荷峰值, 需提取大量预测对象的历史数据, 同时需考虑不同线路可能有不

**基金项目:** 国家电网公司科技项目资助“基于泛在电力物联网与大数据分析应用的负荷预测及配网供电能力挖潜提升”(5211HZ19014P)

同程度的负荷影响因素，且预测时间跨度介于短期预测和中期预测之间，对模型的泛化性提出了更高的要求，使得负荷预测更加困难。因此，如何对未来某一时段线路的高峰负荷走势展开精确预测成为了亟待解决的问题<sup>[2]</sup>。

目前国内外常用的负荷预测方法主要分为两类：只考虑负荷自身时序规律<sup>[3]</sup>的方法、考虑相关因素的预测方法<sup>[4]</sup>。电力负荷往往受到政策宏观调控、经济、气象等多方面因素的影响<sup>[5-6]</sup>，这些因素随机多变，电力负荷与各影响因素之间并非线性关系，难以用精确的关系式表达两者之间的关系，而机器学习算法能深度挖掘各类特征因素对线路负荷的映射关系，因此基于神经网络<sup>[7-10]</sup>、支持向量机(SVM)<sup>[11-13]</sup>、随机森林<sup>[6,14]</sup>等考虑多因素的负荷预测方法涌现出来。近年来，以 XGBoost(Extreme Gradient Boosting)算法为典型的树集成算法涌现，已在许多预测问题中取得了良好的应用效果<sup>[15-19]</sup>。文献[16]提出一种基于贝叶斯优化 XGBoost 的模型用于短期峰值负荷预测。文献[17]使用 XGBoost 算法进行短期负荷预测，并采用网格搜索方法来寻找 XGBoost 的最优超参数。文献[18]提出了一种结合 XGBoost 和 LSTM 的充电负荷预测方法。文献[19]提出基于 K-折交叉验证和 Stacking 融合的短期负荷预测方法。在峰值负荷预测方面，文献[20]提出了一种共轭梯度 BP 算法，应用于峰值负荷预测。文献[21]建立基于时间序列的台区配变负荷峰值预测方法，其模型包括自相关和移动平均两部分。但是上述方法大部分应用于输电网的负荷预测，而应用到配电网线路峰值负荷预测方面的却很少<sup>[22-31]</sup>。

XGBoost 算法具有防过拟合效果更好、损失函数更精确、处理稀疏矩阵效率更高等优点，因此本文采用 XGBoost 算法进行配电网线路峰值负荷预测，预测算法示意图如图 1 所示。在负荷预测前先

进行两方面的分析：一方面结合气象因素、时间因素等，分析各类因素与线路夏季负荷的相关系数分布特征，以确定预测所需的样本特征值；另一方面对线路夏季负荷峰值增长趋势进行聚类分析，筛选出未来可能发生重过载的线路作为目标线路。然后再对目标线路进行 XGBoost 模型训练，预测夏季负荷高峰周的线路负荷曲线，输出重过载线路。并通过实际算例分析证明该算法能大大减少计算量，且具有预测指标好、预测速度快等特点。

### 1 影响因素分析

由于每条配电网线路所接入的负荷类型、负荷所属行业不同，气象因素、是否工作日等时间因素对每条线路的夏季日负荷峰值影响程度也存在差异。除气象因素、时间因素以外，线路的空调负荷在春季可以忽略，相同年份线路春季基础负荷值的高低也可能会影响夏季峰值负荷的大小。

为研究各类因素与夏季峰值负荷的关系，本文提取某地区配电网 1 000 条 10 kV 线路 2016 年至 2019 年的春季负荷数据、夏季负荷高峰周的负荷数据和相关影响因素数据进行相关系数分析。相关系数  $P$  用于衡量两个变量之间的相关程度， $P$  的范围是  $[-1.0, 1.0]$ ，若  $P > 0$ ，表明 2 个变量正相关；若  $P < 0$ ，表明 2 个变量负相关。 $P$  的绝对值越大表明相关性越强， $P \geq 0.6$  为强相关， $0.4 \leq |P| < 0.6$  为中等相关。

#### 1.1 气象因素

提取 2016 年至 2019 年夏季负荷高峰周的负荷数据与天气数据，计算日负荷峰值与最高温度、最低温度、温差、风力的 Pearson 相关系数，并绘制直方图及核密度估计曲线如图 2 所示。经统计可得线路夏季日负荷峰值与各类气象因素的相关系数分布情况如下：

- 1) 与最高温度的相关系数分布范围大致为  $[-0.5, 0.85]$ ，其中 54.3% 的相关系数绝对值超过 0.4，23.5% 的相关系数绝对值超过 0.6。
- 2) 与最低温度的相关系数分布范围大致为  $[-0.85, 0.85]$ ，其中 37.2% 的相关系数绝对值超过 0.4，16.2% 的相关系数绝对值超过 0.6。
- 3) 与温差的相关系数分布范围大致为  $[-0.25, 0.75]$ ，其中 47.3% 的相关系数绝对值超过 0.5，5.1% 的相关系数绝对值超过 0.6。
- 4) 与风力的相关系数分布范围大致为  $[-0.5, 0.25]$ ，可见日负荷峰值与风力相关度低，且 87% 的线路日负荷峰值与风力负相关。

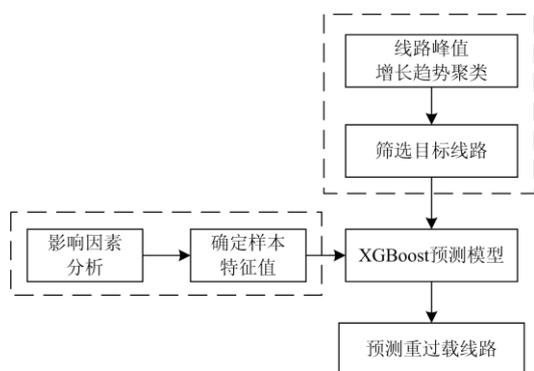


图 1 基于 XGBoost 的配电网线路峰值预测算法示意图

Fig. 1 Schematic diagram of peak load prediction method of distribution network based on XGBoost

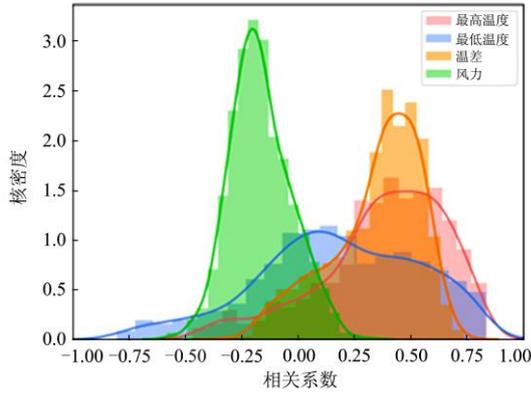


图2 夏季日负荷峰值与各类气象因素的相关系数分布图

Fig. 2 Distribution of correlation coefficient between peak daily load and various meteorological factors in summer

### 1.2 是否工作日

将夏季每日分为工作日与非工作日两类, 计算夏季日负荷峰值与是否工作日的 Spearman 相关系数, 并绘制直方图及核密度估计曲线如图 3 所示。可见, 线路夏季日负荷峰值与是否工作日的相关系数分布范围大致为 $[-0.4, 0.7]$ , 其中 23% 的线路与是否工作日的相关系数绝对值超过 0.4, 11.8% 的线路与是否工作日的相关系数绝对值超过 0.6。

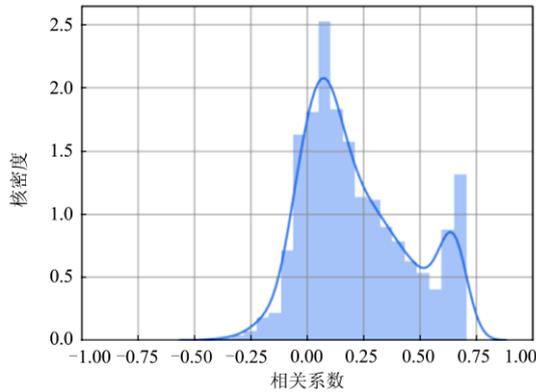


图3 夏季日负荷峰值与是否工作日的相关系数分布图

Fig. 3 Distribution of correlation coefficient between peak daily load in summer and working day or weekend

### 1.3 春季负荷峰值

提取线路 2016 年至 2019 年春季负荷峰值和夏季负荷峰值, 计算春季峰值与夏季峰值的 Pearson 相关系数, 并绘制直方图及核密度估计曲线如图 4 所示, 其中, 70.9% 的线路春夏季峰值的相关系数绝对值超过 0.4, 57.7% 的线路春夏季峰值的相关系数绝对值超过 0.6。

综上所述, 每条配电网线路夏季负荷峰值与最

高气温、最低气温、温差、风力、是否工作日、春季负荷峰值存在不同程度的相关性。为了提高预测未来 1~2 个月后的线路负荷峰值的准确度, 本文将以上影响因素数据加入 XGBoost 预测算法中。

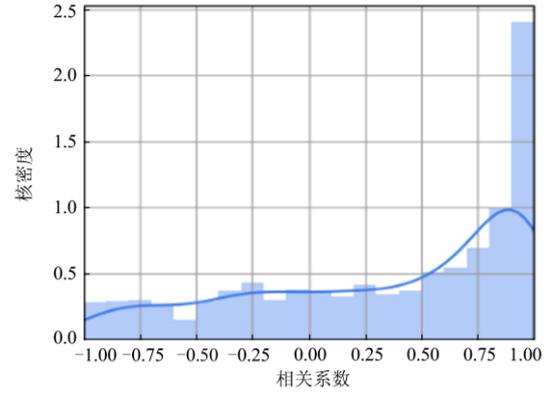


图4 春夏季负荷峰值的相关系数分布图

Fig. 4 Distribution of correlation coefficient of peak load in spring and summer

## 2 XGBoost 算法

### 2.1 算法原理

XGBoost 是一种 Boosting 集成学习算法, XGBoost 算法通过不断迭代, 生成新树来拟合前一棵树的残差, 因此随着迭代次数的增多, 精度在不断提高。同时, XGBoost 算法在 GBDT(梯度提升决策树)的基础上对 Boosting 算法进行改进, 解决了 GBDT 算法模型难以并行计算的问题, 实现对模型过拟合问题的有效控制。

XGBoost 所用树模型是 CART 回归树模型。XGBoost 模型为

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (1)$$

式中:  $x_i$  为输入的第  $i$  个样本数据;  $\hat{y}_i$  为第  $i$  个样本的模型预测值;  $K$  为树的数量;  $F$  为树的集合空间;  $f_k$  为集合空间  $F$  中的一个函数, 且  $f_k(x) = w_{q(x)}$ ,  $f_k$  对应第  $k$  棵独立的树的结构  $q$  和叶子权重  $w$  的相关状况。

XGBoost 的目标函数为

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (2)$$

式(2)中的目标函数由两部分构成, 第一部分用来衡量预测值  $\hat{y}_i$  和真实值  $y_i$  的误差, 第二部分是正则化项, 代表每棵树的复杂度之和, 用于控制模型的复杂度, 计算公式为

$$\Omega(f_k) = \gamma T + \lambda \frac{1}{2} \sum_{j=1}^T \omega_j^2 \quad (3)$$

式中： $T$  表示叶子节点的个数； $\omega$  表示叶子节点的分数； $\gamma$  可以控制叶子节点的个数； $\lambda$  可以控制叶子节点的分数不会过大，防止过拟合。

模型每一次迭代时不影响原模型，并将一个新的函数添加到模型里。一个函数对应一棵树，新生成的树拟合上次预测的残差，对目标函数进行二阶泰勒展开后再求偏导，可得目标函数的最优解为

$$\omega_j^* = -\frac{G_j}{H_j + \lambda} \quad (4)$$

式中： $\omega_j^*$  为第  $j$  个叶子节点的最优得分值； $G_j$  为所有数据在损失函数上的一阶导数； $H_j$  为所有数据在损失函数上的二阶导数。将式(4)代入式(2)，则式(2)的目标函数可简化为

$$Obj = \frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \lambda T \quad (5)$$

除了在目标函数中引入正则项，XGBoost 对每个叶子节点应用了缩减(shrinkage)算法，降低每个树和叶子对结果的影响，提升模型的泛化能力。同时，XGBoost 通过预排序的方式进行分桶，找到多个分割点并将相邻分位点之间的样本分割为一个区域。在遍历特征时，仅需遍历分割点即可计算出最佳分区。在寻找最佳增益节点时，将所有数据放入内存进行计算，得到预排序结果，然后在计算分裂增益的时候直接调用。对于每一层的节点，可以同时计算，从而加快模型训练速度。

XGBoost 是一种有监督学习算法，在电力负荷预测时，它的输入样本数据可以由多个特征值组成，样本标签为负荷数据，通过对样本数据进行多次训练，能够对负荷数据进行较好拟合。

### 2.2 算法样本特征值的选取

对需要进行负荷预测的线路，以夏季负荷高峰周每个负荷值作为样本值，对应的样本特征值数据包括三部分：气象特征数据、时间特征数据、春季负荷特征数据。其中，气象特征数据指每个夏季负荷样本值对应时刻的最高温度、最低温度、温差、风力。时间特征数据指每个夏季负荷样本值对应时刻所属的年、月、日、时、分、是否工作日。春季负荷特征数据指每个夏季负荷样本值对应时刻所在年份的春季负荷峰值。具体包括以下七类数据：

- 1) 最高温度  $T_h$ ，即当日的最高气温。
- 2) 最低温度  $T_l$ ，即当日的最低气温。
- 3) 温差  $T_t$ ，即当日的最高气温与最低气温的

差值。

4) 风力  $W_d$ ，即当日的最大风力。

5) 是否工作日  $W_k$ ，当日若为工作日，则标记为 0，否则标记为 1。

6) 当前日期与时间。包括：当前年份  $T_y$ 、当前月份  $T_m$ 、当前日期  $T_d$ 、当前小时  $T_h$ 、当前分钟  $T_{mm}$ 。

7) 春季负荷峰值  $L_p$ 。即当前年份春季的负荷峰值。

以此七类数据为基础，组成每个时刻的样本输入，即  $x = [T_h, T_l, T_t, W_d, W_k, T_y, T_m, T_d, T_h, T_{mm}, L_p]$  将对应时刻的负荷值作为输出  $y = L$ ，则训练数据的输入数据为  $X = [x_1, x_2, x_3, \dots, x_n]$ ，训练数据的输出数据为  $Y = [y_1, y_2, y_3, \dots, y_n]$ 。其中， $n$  为数据点个数。

### 3 预测模型

地区配电网线路条数常达数千条，而每年出现重过载的线路约数十条，若对所有线路进行负荷预测，必然需要加载大量的历史数据，因此为减少历史数据加载量与计算量，本文提出的基于 XGBoost 的配电网线路峰值预测流程如图 5 所示。

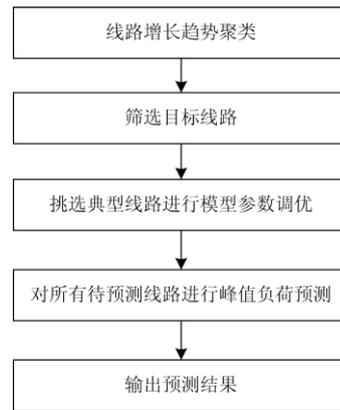


图 5 配电网线路峰值负荷预测流程图

Fig. 5 Flow chart of peak load forecasting for distribution network lines

具体流程说明如下。

#### 1) 负荷增长趋势聚类

线路负荷峰值变化趋势存在一定的共性，有些线路长年处于低负载运行水平，且增长趋势平缓；有些线路负荷峰值逐年增长或逐年下降。因此，本预测算法首先从近几年的线路负荷数据中提取出线路的夏季负荷峰值，组成每条线路的负载率增长趋势曲线，再对线路负载率增长趋势进行 K-means 聚类分析。

## 2) 筛选目标线路

从聚类结果中, 将近年负载率低且增长趋势平缓、未来出现重过载现象可能性低的线路从预测线路列表中剔除, 以减少预测计算工作量。

## 3) 模型参数调优

选取几条典型线路, 设置不同的算法参数, 对每条线路进行单独训练及预测。对比不同的算法参数情况下预测指标的差异, 从而确定所有目标线路在模型训练时需要设置的统一的算法参数。

其中, 本文选用的四个预测指标分别为: 模型训练耗时、平均相对误差(Mean Absolute Percentage Error, MAPE)  $E_{MAPE}$ 、相对均方误差(Root Mean Square error, RMSE)  $E_{RMSE}$  和峰值误差  $E_{PEAK}$ 。

$$E_{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (6)$$

$$E_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

$$E_{PEAK} = \frac{|y_{peak} - \hat{y}_{peak}|}{I_{control}} \quad (8)$$

式中:  $y_i$  为实际负荷值;  $\hat{y}_i$  为预测负荷值;  $n$  为预测点个数;  $y_{peak}$  为预测时段线路实际峰值;  $\hat{y}_{peak}$  为预测时段线路预测峰值;  $I_{control}$  为线路夏控电流。

## 4) 峰值负荷预测

用统一的算法参数, 对目标线路进行负荷模型训练。再将待预测时段的特征值组成输入数据, 调用训练好的模型, 计算待预测时段的线路负荷值。从而得到线路在预测时段的最大负载率, 输出预测结果清单。

## 4 算例分析

为验证算法效果, 本文对某省会城市城区配电网 2019 年夏季负荷数据进行统计得知, 其负荷高峰时段为 2019 年 7 月 29 日至 8 月 2 日, 因此本文对此时段进行配电网 10 kV 线路的峰值负荷预测。由于历史负荷数据距离预测时段的时间越久, 其对预测时段负荷的影响越小, 从而选取用于分析计算的历史数据负荷数据包括: 2016 年至 2018 年共计 3 年夏季 7 月至 8 月的负荷数据、2016 年至 2019 年共计 4 年春季 3 月至 4 月的负荷数据, 线路负荷采样间隔为 5 min, 即每天有 288 个时间点的负荷。所用实验平台条件为 Windows10 X64 操作系统、Inter i5-8265 CPU, 使用 Python 语言编程实现, XGBoost 采用框架 py-xgboost 框架, 为了对比效果, 分别与 BP 神经网络、LSTM 两种算法进行计算比

较, 其中 BP 算法、LSTM 算法采用 keras 框架。

## 4.1 增长趋势聚类分析

从 2016 年至 2018 年历史 3 年的线路负荷数据中提取出每年夏季的线路负荷峰值, 组成线路负载率增长趋势曲线, 再对线路负载率增长趋势进行 K-means 聚类分析。

由于有些线路历史数据不全, 因此仅对数据完整的 889 条线路进行聚类计算。拟定聚类中心个数范围为[5,12], 依次进行聚类分析, 最终确定聚类中心个数为 10 时, 负荷增长趋势分类较为合理。图 6 所示为 10 类线路负荷增长趋势曲线及相应的增长趋势中心曲线。由图 6 可见, 第 1、9 类线路呈递增趋势, 未来出现重过载的可能性较高; 第 2、3、4 类线路处于较低水平, 且增长趋势平缓, 第 6 类呈下降趋势, 未来出现重载的可能性小; 第 5 类线路历年最大负载率均处于较高水平, 未来出现重过载的风险最大; 第 7、8、10 类呈现波动趋势。因此第 2、3、4、6 类可不参与预测。

表 1 所示为各类增长趋势所包含的线路条数分布, 可见第 2、3、4 类线路条数最多, 共计 458 条, 总比例为 51.52%, 说明一半以上线路属于低负载运行, 且增长趋势平缓。第 6 类共 66 条, 说明 7.42% 的线路呈负载率下降趋势。

表 1 各类增长趋势所含线路条数分布情况

Table 1 Distribution of the number of lines in various

growth trends					
类别	第 1 类	第 2 类	第 3 类	第 4 类	第 5 类
线路条数	43	153	134	171	70
类别	第 6 类	第 7 类	第 8 类	第 9 类	第 10 类
线路条数	66	59	31	101	61

## 4.2 峰值负荷预测

采用 XGBoost 算法对需要预测的 476 条线路进行 2019 年夏季峰值负荷周的曲线预测, 并将预测峰值结果与线路实际负荷峰值进行对比, 预测结果中 73.7% 的线路峰值预测误差小于 15%。

线路实际重载(负载率  $\geq 80\%$  为重载)且预测误差小于 15% 或预测结果也为重载, 则认为线路重载预测正确。重载线路预测结果如表 2 所示, 预测时段实际出现重载的线路共 44 条, 线路重载预测正确的线路共 32 条, 占总重载线路条数的 72.73%。其中, 第 5 类线路中出现重载的比例最高, 达 25.71%, 为 18 条, 预测正确率为 100%。可见, 第 5 类线路出现重载的概率最高, 预测结果最好。

同时, 为验证线路聚类筛选的可靠性, 对属于第 2、3、4、6 类的 524 条线路采用 XGBoost 算法

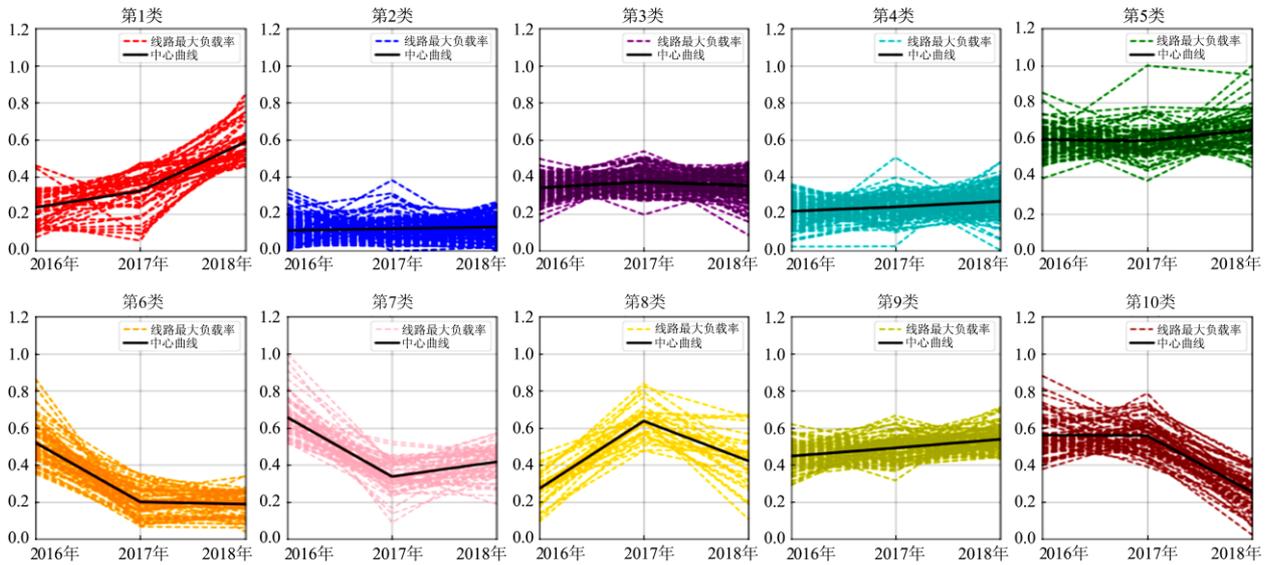


图 6 线路负荷峰值增长趋势聚类结果

Fig. 6 Clustering results of growth trend of lines peak load

进行峰值预测，预测结果中无重载线路。而此 524 条线路实际重载线路条数为 2 条，其中第 4 类线路 1 条，第 6 类线路 1 条。可见线路聚类筛选方法有效减少了预测计算量，但被剔除的线路中也有可能

表 2 重载线路预测结果

Table 2 Prediction results of heavy load lines

增长类型	实际重载条数	实际重载占比/%	预测正确条数	预测正确占比/%	线路总条数
未知类	3	2.70	2	66.67	111
第 1 类	5	11.63	4	80.00	43
第 5 类	18	25.71	18	100.00	70
第 7 类	3	5.08	2	66.67	59
第 8 类	4	12.90	2	50.00	31
第 9 类	9	8.91	3	33.33	101
第 10 类	2	3.28	1	50.00	61
合计	44	9.24	32	72.73	476

### 4.3 预测结果对比

为了对比预测效果，在模型参数调优分析时，本文算法分别与 BP 神经网络、LSTM 算法进行对比。以目标线路中的两条典型线路为例，线路预测负荷曲线与实际负荷曲线对比如图 7、图 8 所示，预测指标对比如表 3 所示。由图 7、图 8 所示负荷曲线图对比可见，两条线路采用 XGBoost 算法的预测负荷曲线与实际负荷曲线最接近，而 BP 与 LSTM 算法的预测负荷曲线与实际负荷曲线偏差较大。

由表 3 所示预测指标对比可见：

1) 模型训练耗时方面，两条线路采用 XGBoost

算法耗时均少于 500 ms，而采用 BP 算法耗时均超过 2 s，少于 10 s，采用 LSTM 算法耗时超 100 s；

表 3 线路负荷预测指标对比

Table 3 Comparison of line load forecasting indexes

指标	XGBoost	BP	LSTM	
线路 1	训练耗时/s	0.164	2.404	105.3
	$E_{RMSE}/\%$	18.853	47.644	62.205
	$E_{MAPE}/\%$	7.737	16.616	35.882
	$E_{PEAK}/\%$	3	6	10
线路 2	训练耗时/s	0.482	5.12	133.6
	$E_{RMSE}/\%$	22.589	55.289	55.756
	$E_{MAPE}/\%$	6.866	17.859	18.652
	$E_{PEAK}/\%$	6	13	18

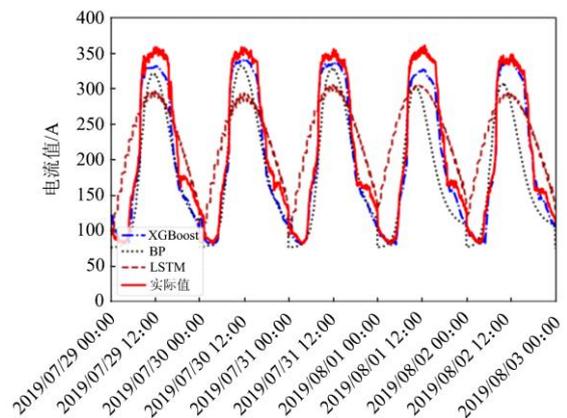


图 7 线路 1 预测负荷曲线与实际负荷曲线对比

Fig. 7 Comparison between predicted load curve and real load curve of line 1

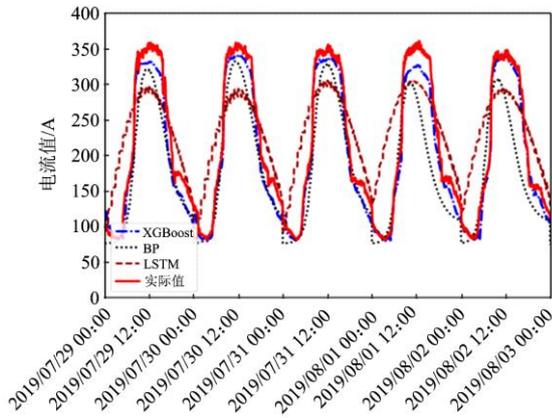


图 8 线路 2 预测负荷曲线与实际负荷曲线对比

Fig. 8 Comparison between predicted load curve and real load curve of line 2

2) 预测指标方面, XGBoost 算法的  $E_{RMSE}$ 、 $E_{MAPE}$ 、 $E_{PEAK}$  均优于 BP 与 LSTM 算法。

而在算法预处理方面: BP、LSTM 算法计算前需对数据进行归一化处理, 而 XGBoost 进行归一化处理。

在算法参数设定方面:

1) XGBoost 参数设定: 树的最大个数  $n_{estimators}$  设置为 100,  $subsample$  设置为 1.0, 学习率设置为 0.1, 其他参数均设置为默认值。其中  $n_{estimators}$  个数的变化对峰值指标的影响最明显, 如表 4 所示, 随着  $n_{estimators}$  从 10 增加到 60, 两条线路的峰值误差快速下降,  $n_{estimators} \geq 70$  以后误差不变, 因此将  $n_{estimators}$  保守设置为 100。

表 4 XGBoost 参数对峰值指标影响对比

Table 4 Comparison of the influence of XGBoost parameters on peak index

序号	$n_{estimators}$ 值	线路 1	线路 2
		峰值误差/%	峰值误差/%
1	10	37	31
2	20	20	15
3	30	12	9
4	40	9	6
5	50	8	5
6	60	7	4
7	70	6	4
8	80	6	4
9	90	6	4
10	100	6	4

2) BP 参数设定: 第一层隐含层神经元个数为 12, 第二层隐含层神经元个数为 24, 训练次数为 400, 学习速率为 0.1。

3) LSTM 参数设定: 第一层隐含层神经元个数

为 16, 第二层隐含层神经元个数为 32, 训练次数为 1000, 批处理样本数为 288。

综上所述, BP 与 LSTM 均需通过配置多个参数项、训练次数多才能达到比较好的预测效果, 而采用 XGBoost 参数设置相对简单, 且具有计算速度快、拟合效果好、预测误差小的优点。

## 5 结论

针对配电网线路夏季峰值负荷预测, 本文提出一种基于 XGBoost 的配电网线路峰值负荷预测方法, 并通过某地区配电网进行实际算例分析进行验证, 算例结果表明该方法具有较好的应用价值。本文提出的基于 XGBoost 的配电网线路峰值负荷预测方法的特点在于:

1) 面对大量的预测线路对象, 而各影响因素对各条线路的影响程度不同, 该预测方法基于气象因素、时间因素、春季基础负荷与线路夏季负荷的相关系数分布特征分析, 再将各影响因素作为预测算法的样本特征值, 使算法具有更好的泛化性。

2) 对于预测对象历史数据量大的问题, 该预测方法先对线路夏季负荷峰值增长趋势进行聚类分析, 识别发生重过载可能性低的线路, 筛选出未来可能发生重过载的线路作为目标线路, 以减少待预测线路, 避免不必要的计算。

3) 在影响因素分析与线路聚类分析筛选的基础上, 提出了基于 XGBoost 算法的配电网线路峰值负荷预测方法, 与 BP、LSTM 算法预测效果相比, 具有预测指标好、预测速度快、参数易于设置的特点。

## 参考文献

- [1] 康重庆, 夏清, 刘梅, 等. 电力系统负荷预测[M]. 2 版. 北京: 中国电力出版社, 2007.
- [2] 王文秀, 田世明, 王泽忠, 等. 一种基于贝叶斯网络的电力负荷峰值预测方法[J]. 供用电, 2019, 36(7): 57-64. WANG Wenxiu, TIAN Shiming, WANG Zezhong, et al. A power peak load forecasting method based on Bayesian network[J]. Distribution & Utilization, 2019, 36(7): 57-64.
- [3] HUANG L, YANG Y, ZHAO H, et al. Time series modeling and filtering method of electric power load stochastic noise[J]. Protection and Control of Modern Power Systems, 2017, 2(3): 269-275. DOI: 10.1186/s41601-017-0059-8.
- [4] 常小强, 宋政湘, 王建华. 基于蒙特卡罗算法的电动汽车充电负荷预测及系统开发[J]. 高压电器, 2020, 56(8): 1-5. CHANG Xiaoqiang, SONG Zhengxiang, WANG Jianhua.

- Electric vehicle charging load prediction and system development based on Monte Carlo algorithm[J]. High Voltage Apparatus, 2020, 56(8): 1-5.
- [5] 刘翊枫, 周辉, 刘昕, 等. 基于气象成分分解的夏季短期负荷预测[J]. 电测与仪表, 2019, 56(21): 129-135.  
LIU Yifeng, ZHOU Hui, LIU Xin, et al. Short-term load forecasting in summer based on meteorological factors decomposition[J]. Electrical Measurement & Instrumentation, 2019, 56(21): 129-135.
- [6] 李焱, 贾雅君, 李磊, 等. 基于随机森林算法的短期电力负荷预测[J]. 电力系统保护与控制, 2020, 48(21): 117-124.  
LI Yan, JIA Yajun, LI Lei, et al. Short term power load forecasting based on a stochastic forest algorithm[J]. Power System Protection and Control, 2020, 48(21): 117-124.
- [7] 何川, 舒勤, 贺含峰. ICA 特征提取与 BP 神经网络在负荷预测中的应用[J]. 电力系统及其自动化学报, 2014, 26(8): 40-46.  
HE Chuan, SHU Qin, HE Hanfeng. Application of BP neural network and ICA feature extraction in power load forecasting[J]. Proceedings of the CSU-EPSA, 2014, 26(8): 40-46.
- [8] 祝学昌. 基于 IFOA-GRNN 的短期电力负荷预测方法研究[J]. 电力系统保护与控制, 2020, 48(9): 121-127.  
ZHU Xuechang. Research on short-term power load forecasting method based on IFOA-GRNN[J]. Power System Protection and Control, 2020, 48(9): 121-127.
- [9] 黄少雄, 王璨, 孔庆竹, 等. 含短期预测的光伏配电网智能调压策略[J]. 热力发电, 2020, 49(7): 21-27.  
HUANG Shaoxiong, WANG Can, KONG Qingzhu, et al. Intelligent voltage regulation strategy of photovoltaic distribution network considering short-term forecasting[J]. Thermal Power Generation, 2020, 49(7): 21-27.
- [10] 骆小满, 皇甫成, 阮江军, 等. 基于神经网络的热电联产机组热负荷和电负荷预测[J]. 热力发电, 2019, 48(9): 46-50.  
LUO Xiaoman, HUANG Fucheng, RUAN Jiangjun, et al. Prediction of heat and electric load of cogeneration unit based on neural network[J]. Thermal Power Generation, 2019, 48(9): 46-50.
- [11] 宫毓斌, 滕欢. 基于 GOA-SVM 的短期负荷预测[J]. 电测与仪表, 2019, 56(14): 12-16.  
GONG Yubin, TENG Huan. Short-term load forecasting based on GOA-SVM[J]. Electrical Measurement & Instrumentation, 2019, 56(14): 12-16.
- [12] 陈勇, 李鹏, 张忠军, 等. 基于 PCA-GA-LSSVM 的输电线路覆冰负荷在线预测模型[J]. 电力系统保护与控制, 2019, 47(10): 110-119.  
CHEN Yong, LI Peng, ZHANG Zhongjun, et al. Online prediction model for power transmission line icing load based on PCA-GA-LSSVM[J]. Power System Protection and Control, 2019, 47(10): 110-119.
- [13] 张兴科, 魏朝阳, 李征, 等. 包含高渗透率分布式电源的母线负荷区间预测[J]. 电网与清洁能源, 2020, 36(12): 101-106.  
ZHANG Xingke, WEI Chaoyang, LI Zheng, et al. Bus load interval prediction of the distributed generation with high penetration[J]. Power System and Clean Energy, 2020, 36(12): 101-106.
- [14] 邢书豪, 高广玲, 张智晟. 基于双层随机森林算法的短期负荷预测模型[J]. 广东电力, 2019, 32(9): 160-166.  
XING Shuhao, GAO Guangling, ZHANG Zhisheng. Short-term load forecasting model based on double-layer random forest algorithm[J]. Guangdong Electric Power, 2019, 32(9): 160-166.
- [15] 杨修德, 王金梅, 张丽娜. XGBoost 在超短期负荷预测中的应用[J]. 电气传动自动化, 2017, 39(4): 21-25.  
YANG Xiude, WANG Jinmei, ZHANG Lina. Application of XGBoost in ultra-short load forecasting[J]. Electric Drive Automation, 2017, 39(4): 21-25.
- [16] 龚雪娇, 朱瑞金, 唐波. 基于贝叶斯优化 XGBoost 的短期峰值负荷预测[J]. 电力工程技术, 2020, 39(6): 76-81.  
GONG Xuejiao, ZHU Ruijin, TANG Bo. Short-term peak load forecasting based on Bayesian optimization XGBoost[J]. Electric Power Engineering Technology, 2020, 39(6): 76-81.
- [17] OH J Y, HAM D H, LEE Y G, et al. Short-term load forecasting using XGBoost and the analysis of hyperparameters[J]. Transactions of the Korean Institute of Electrical Engineers, 2019, 68(9): 1073-1078.
- [18] XUE M, WU L, ZHANG Q P, et al. Research on load forecasting of charging station based on XGBoost and LSTM model[J]. Journal of Physics: Conference Series, 2021, 1757(1).
- [19] 朱文广, 李映雪, 杨为群, 等. 基于 K-折交叉验证和 Stacking 融合的短期负荷预测[J]. 电力科学与技术学报, 2021, 36(1): 87-95.  
ZHU Wenguang, LI Yingxue, YANG Weiqun, et al. Short-term load forecasting based on the K-fold cross-validation and stacking ensemble[J]. Journal of Electric Power Science and Technology, 2021, 36(1): 87-95.
- [20] SAINI L M, SONI M K. Artificial neural network based peak load forecasting using conjugate gradient methods

- model[J]. IEEE Transactions on Power Systems, 2002, 17(3): 907-912.
- [21] 刘升. 基于时间序列的台区配电负荷峰值预测[J]. 电力科学与工程, 2018, 34(7): 56-60.  
LIU Sheng. Peak value forecasting for district distribution load based on time series[J]. Electric Power Science and Engineering, 2018, 34(7): 56-60.
- [22] LAHOUAR A, SLAMA J B H. Day-ahead load forecast using random forest and expert input selection[J]. Energy Conversion & Management, 2015, 103: 1040-1051.
- [23] 王增平, 赵兵, 纪维佳, 等. 基于GRU-NN模型的短期负荷预测方法[J]. 电力系统自动化, 2019, 43(5): 53-62.  
WANG Zengping, ZHAO Bing, JI Weijia, et al. Short term load forecasting method based on GRU-NN model[J]. Automation of Electric Power Systems, 2019, 43(5): 53-62.
- [24] ZHAO Dongbo, GE Qian, TIAN Yuting, et al. Short-term load demand forecasting through rich features based on recurrent neural networks[J]. IET Generation, Transmission & Distribution, 2020, 15(5).
- [25] 刘翊枫, 周国鹏, 刘昕, 等. 基于智能相似日识别及偏差校正的短期负荷预测方法[J]. 电力系统保护与控制, 2019, 47(12): 138-145.  
LIU Yifeng, ZHOU Guopeng, LIU Xin, et al. A short-term load forecasting method based on intelligent similar day recognition and deviation correction[J]. Power System Protection and Control, 2019, 47(12): 138-145.
- [26] KONG Weicong, DONG Zhaoyang, JIA Youwei, et al. Short-term residential load forecasting based on LSTM recurrent neural network[J]. IEEE Transactions on Smart Grid, 2019, 10(1): 841-851.
- [27] 李富鹏, 沈秋英, 王森, 等. 基于大数据和多因素组合分析的单元制配电网精细化负荷预测[J]. 智慧电力, 2020, 48(1): 55-62.  
LI Fupeng, SHEN Qiuying, WANG Sen, et al. Refined load forecasting method for unit distribution network based on big data and multiple factors[J]. Smart Power, 2020, 48(1): 55-62.
- [28] 陈潇雅, 刘志坚, 刘晓欣, 等. 基于长期负荷预测和联络分析的配电网规划[J]. 广东电力, 2019, 32(3): 59-65.  
CHEN Xiaoya, LIU Zhijian, LIU Xiaoxin, et al. Distribution network planning based on long-term load forecasting and tie-line analysis[J]. Guangdong Electric Power, 2019, 32(3): 59-65.
- [29] 徐璐, 熊天军, 杨素, 等. 计及新电改政策的配电网变化影响因素及其规划方法综述[J]. 智慧电力, 2020, 48(9): 108-117.  
XU Lu, XIONG Tianjun, YANG Su, et al. Survey of distribution network change factors & planning methods under new electricity reform policy[J]. Smart Power, 2020, 48(9): 108-117.
- [30] 马愿, 张倩, 李国丽, 等. 基于频域分解的短期负荷预测研究分析[J]. 中国电力, 2020, 53(4): 114-121.  
MA Yuan, ZHANG Qian, LI Guoli, et al. Research and analysis of short-term load forecasting based on frequency domain decomposition[J]. Electric Power, 2021, 54(3): 114-121.
- [31] 王睿, 高欣, 李军良, 等. 基于聚类分析的电动汽车充电负荷预测方法[J]. 电力系统保护与控制, 2020, 48(16): 37-44.  
WANG Rui, GAO Xin, LI Junliang, et al. Electric vehicle charging demand forecasting method based on clustering analysis[J]. Power System Protection and Control, 2020, 48(16): 37-44.

收稿日期: 2020-12-01; 修回日期: 2021-05-06

作者简介:

姜建(1977—), 男, 高级工程师, 研究方向为电力系统配电网智能运行技术;

刘海琼(1984—), 女, 通信作者, 硕士, 中级工程师, 研究方向为电力系统调度自动化技术。E-mail: 185206695@qq.com

(编辑 周金梅)