

DOI:10.12158/j.2096-3203.2023.01.024

基于双层 XGBoost 和数据增强的空间负荷预测方法

黄冬梅¹, 张宁宁¹, 胡安锋¹, 胡伟², 肖勇³, 陈岸青⁴

(1. 上海电力大学电子与信息工程学院, 上海 201306; 2. 上海电力大学经济与管理学院, 上海 201306; 3. 上海电力大学继续教育学院, 上海 200090; 4. 国网信通亿力科技有限责任公司, 福建 福州 350001)

摘要:为了解决空间负荷预测面临的特征变量众多和数据匮乏问题,文中提出一种基于双层极端梯度提升(extreme gradient boosting, XGBoost)和数据增强的空间负荷预测方法。该方法首先将待预测区域按馈线供电范围划分为若干子区域;其次构建基于双层 XGBoost 的特征选择模型,第一层 XGBoost 对特征进行评分及排序,将组合特征和负荷输入第二层 XGBoost 并进行子区域负荷预测,根据子区域负荷预测结果选择每个子区域的最佳特征变量;然后利用生成对抗网络(generative adversarial network, GAN)增强每个子区域的训练集样本,并通过极限学习机(extreme learning machine, ELM)实现子区域预测;最后将每个子区域的预测值相加得到待预测区域的预测值。以上海市局部区域为例,对文中方法进行仿真实验和对比分析。结果表明,文中方法可同时解决特征变量选择和数据匮乏问题,具有更高的预测精度。

关键词:空间负荷预测;极端梯度提升(XGBoost);特征选择;生成对抗网络(GAN);数据增强;极限学习机(ELM)

中图分类号: TM715

文献标志码: A

文章编号: 2096-3203(2023)01-0201-08

0 引言

空间负荷预测是待预测区域的负荷总量和负荷增长位置同时进行预测,在城市配电网规划中具有非常重要的作用,可为供电设备容量及最佳配置位置的选择提供依据,促进配电网规划的精益化管理^[1-2]。

空间负荷的时空变化受待预测区域的社会、经济、人口、气候等众多特征因素影响。文献[3]融合非结构化信息、建设情况、发展定位等元胞特征信息,利用深度置信神经网络算法充分挖掘负荷元胞的多源信息特征,提升空间负荷预测精度。文献[4]建立计及经济、人口、气温等因素影响的多变量预测模型,实现城市电网的负荷总量预测,并采用基于用电信息的负荷密度指标法,结合地理信息系统平台分配到各个供电小区。文献[5-6]提出基于灰色关联分析(grey relational analysis, GRA)和最小二乘支持向量机(least square support vector machine, LSSVM)的空间负荷预测方法,综合用地属性、人口密度、人均年用电负荷等指标,并将其作为特征因素输入,进而对负荷密度值进行预测。上述文献均考虑了较多的负荷特征变量,但并未针对性地进行选择,忽略了特征变量对空间负荷影响程度的差异性。

随着机器学习的发展,基于神经网络模型的空

间负荷预测方法得以应用。文献[7]以元胞的历史负荷数据为基础,结合径向基函数(radial basis function, RBF)神经网络在非线性拟合方面的优势,提出基于元胞负荷特性分析的 RBF 神经网络空间负荷预测方法。文献[8]提出考虑本位元胞接受能力和相邻元胞负荷影响的空间负荷预测方法,利用空间卷积建立反映相邻元胞间负荷相互影响的量化模型。文献[9]提出一种数据驱动方法,建立 Softmax 多元概率分类模型,对未知地块的负荷水平类型进行匹配。然而,神经网络实现高精度空间负荷预测需要大量数据,而空间负荷预测面临数据匮乏的问题。针对该问题,文献[10-11]利用用户业扩报装计划和其他用户的历史负荷数据建立模型,文献[12]利用生成对抗网络(generative adversarial network, GAN)建立数据生成模型,提升空间负荷预测精度。但上述文献只增强了负荷数据,没有对与空间负荷密切相关的社会、经济、人口等特征变量进行数据增强。

文中针对上述特征变量选择与数据增强问题,提出基于双层极端梯度提升(extreme gradient boosting, XGBoost)和数据增强的空间负荷预测方法。首先构建双层 XGBoost 模型筛选出各个子区域的最佳特征组合,利用 GAN 增强各子区域的训练集样本。然后引入极限学习机(extreme learning machine, ELM)实现预测,提高模型的运算速度。最后通过实例分析对文中方法提高空间负荷预测精度的可行性进行验证。

收稿日期:2022-07-13;修回日期:2022-09-29

1 基于双层 XGBoost 的特征选择模型

1.1 XGBoost 原理

XGBoost 是一个树集成模型,可以看成对 K 棵树的样本预测值进行求和,其定义函数为:

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

式中: \hat{y} 为预测值; x_i 为第 i 个样本; $\theta(x_i)$ 为 x_i 的函数; f_k 为第 k 棵树模型; K 为树的总数; F 为由决策树组成的函数空间。

XGBoost 的目标函数为:

$$L(\theta) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (2)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (3)$$

式中: \hat{y}_i 为第 i 个样本的预测值; y_i 为第 i 个样本的类别标签; $l(\hat{y}_i, y_i)$ 为损失误差函数; $\Omega(f_k)$ 为第 k 棵树的目标函数; T 为叶子的个数; γ, λ 为系数; ω 为叶子节点的权重向量。

式(2)为损失函数项,即训练误差。式(3)为正则化项,具有防止过拟合的作用。整体的目标是在 $L(\theta)$ 取得最小值时得出相对应的模型 f , f 不是具体值,因此无法在欧式空间中进行优化。通常将式(1)进行展开,函数 $f_k(x_i)$ 须学习,每次学习结束后原模型不变,同时向原模型加入一个新的函数,目的是使目标函数最小化。

1.2 特征选择模型

在进行空间负荷预测时,特征数量并非越多越好,有效特征才能提高模型预测精度。为了确定最合适的特征,文中构建了一个双层 XGBoost 模型。

第一层 XGBoost 网络作为特征评价层。用特征工程进行特征重要性排序,通过梯度提升算法衡量特征在模型提升决策树构建中的价值。模型在构建决策树时,某一属性被用到的次数越多,其重要性就相对越高。文中各子区域收集了负荷和多个特征变量组成数据集,经过对数据集进行特征提取并调用权重和增益这 2 个取值,最终将一个属性在所有提升树中的结果进行加权求和然后求平均,得到重要性得分,从高到低直观显示每个特征的重要性。

第二层 XGBoost 网络作为预测层。将各子区域数据集分成训练集和测试集,排好序的特征变量由高分到低分取不同的数量进行组合,网络训练完成后通过测试集负荷预测结果的平均绝对百分比误差(mean absolute percentage error, MAPE) 确定最佳特征变量。MAPE 越小,此时的特征组合对应的负

荷预测精度越高,进而确定各子区域的特征变量。

XGBoost 代价函数里加入正则项防止过拟合,在小样本数据上仍能取得较好的预测效果,且通过嵌入式特征选择方法选出来的特征变量更加合理有效。

2 数据增强与预测模型

2.1 GAN 生成模型

GAN 能在训练过程中模拟原始样本分布,批量生成新样本,增强原始样本集的多样性和规模^[13],在处理小样本数据时性能突出,广泛应用于图像处理、计算机视觉、自然语言处理等领域。

GAN 由生成网络和判别网络构成,均为多层感知神经网络,其整体结构如图 1 所示。生成网络和判别网络均采用 LeakyReLU 函数作为激活函数,可以解决训练过程的神经饱和问题,提高计算效率。

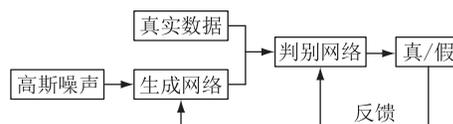


图 1 GAN 结构

Fig.1 Structure of GAN

生成网络通过接收一个随机高斯噪声 z 生成新的数据^[13]。判别网络则是判别生成数据的真假,判别网络的输入参数包括真实数据 x 和生成网络的输出 $G(z)$,判别网络的输出为 $D(\cdot)$,若结果为 1,则为真实数据,若结果为 0,则为假数据。通过生成网络与判别网络不断地对抗博弈,最终达到纳什均衡。

GAN 模型训练过程中先是固定生成网络,通过损失函数优化判别网络,可表示为:

$$\max_D V(D, G) = E(\log_2 D(x)) + E(\log_2(1 - D(G(z)))) \quad (4)$$

式中: $E(\cdot)$ 为数学期望; x 服从概率分布 $p_x(x)$; z 服从概率分布 $p_z(z)$ 。

然后固定判别网络,通过损失函数优化生成网络,可表示为:

$$\min_G V(D, G) = E(\log_2(1 - D(G(z)))) \quad (5)$$

将 2 个过程整合后得到总体目标函数,即:

$$\min_G \max_D V(G, D) = E(\log_2 D(x)) + E(\log_2(1 - D(G(z)))) \quad (6)$$

文中首先用 Adam 优化器对原始参数进行更新优化,再利用此时的判别网络对生成网络进行多次无监督训练来优化其网络参数,使生成网络生成更加真实的数据,直至判别网络判别生成的数据为真实数据。GAN 模型用于增强各个子区域的训练集,

根据其原始训练集样本来生成足量的新样本并合并到原来的训练集中去,以此达到数据增强的目的。该模型能够把握负荷与特征变量之间的映射关系,符合原始数据之间的变化规律。

2.2 ELM 预测模型

ELM 通过改进反向传播算法提升学习效率、简化学习参数,不仅可以应用于监督学习,也可以处理非监督学习问题,在图像处理及时间序列预测等方面应用广泛。

ELM 是一种单隐含层的前馈神经网络,其结构如图 2 所示。输入层和隐含层的连接权值以及隐含层的阈值可随机设定且不再调整,大幅减少了运算量^[14-15]。隐含层与输出层的连接权值不用迭代调整,可通过解方程组一次性确定。

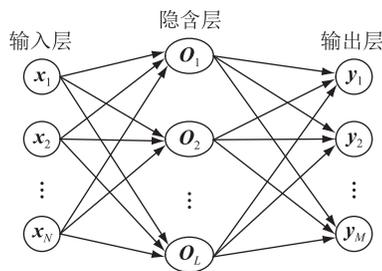


图 2 ELM 的网络结构

Fig.2 Network structure of ELM

假设输入层 \mathbf{x} 有 N 组样本 $\mathbf{x}_i = [x_{i1} \ x_{i2} \ \dots \ x_{in}]^T, i=1, 2, \dots, N$; 隐含层 \mathbf{O} 有 L 个神经元; 输出层 \mathbf{y} 为 $\mathbf{y}_j = [y_{j1} \ y_{j2} \ \dots \ y_{jm}]^T, j=1, 2, \dots, M$ 。则 ELM 的输出为:

$$\mathbf{y} = \sum_{l=1}^L [\beta_l h(\omega_l \mathbf{x} + b_l)] \quad (7)$$

式中: $h(\cdot)$ 为激活函数; ω_l 为输入层与隐含层第 l 个神经元的输入权重; β_l 为隐含层第 l 个神经元与输出层的输出权重; b_l 为隐含层第 l 个神经元的偏置。

文中 ELM 是最终的预测模型,将各子区域增强后的训练集输入训练,模型训练完成后,再将其测试集输入预测,即可得到各子区域的预测值。

3 空间负荷预测过程

首先将待预测区域划分为子区域,其次用双层 XGBoost 模型确定特征变量,然后用 GAN 增强训练样本,最后采用 ELM 进行预测,空间负荷预测流程如图 3 所示,具体步骤如下。

(1) 子区域划分及数据收集。依据馈线的供电范围将待预测区域划分为不同子区域^[16-20],结合区域用电构成和《城市用地分类与规划建设用地标

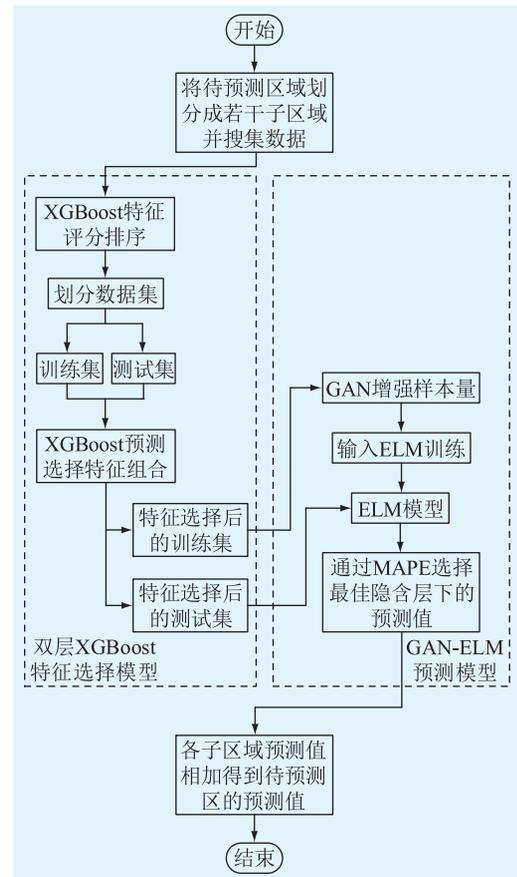


图 3 空间负荷预测流程

Fig.3 Flow chart of spatial load forecasting

准》进行负荷影响因素分析。从统计年鉴中搜集各子区域 10 a 内的用电量以及 10 个影响负荷的主要特征变量数据。10 个主要特征分别为人口密度、居住房屋面积、工厂面积、人均可支配收入、工业总产值、一般公共预算收入、年平均气温、年总降水量、年极端最高气温和年极端最低气温^[21-22]。分布式光伏和储能可能会对空间负荷需求产生一定影响,但考虑到光伏和储能使用空间限制、政策因素影响以及在大城市电网电能供给中占比较低的实际情况,并未将二者作为影响负荷的主要特征。

(2) 特征选择。分别在每个子区域内用 XGBoost 对 10 个特征变量进行评价,按照得分将其由高到低排序。然后将样本按照 3:2 的比例划分训练集和测试集,取不同的特征进行组合,再通过 XGBoost 进行预测,结果中 MAPE 最小的一组特征即是每个子区域预测效果最佳的特征组合。

(3) 数据增强。通过 GAN 将各子区域的训练集分别生成新样本,然后与原训练集样本合并实现数据增强。GAN 生成网络和判别网络通过选取合适的参数使其彼此之间不断优化,充分把握负荷与特征之间的映射关系,生成可用于空间负荷预测的数据。

(4) 子区域负荷预测。将增强后的训练集输入 ELM 模型进行训练,训练完成后再将测试集输入进行预测。为防止隐含层神经元数量对精度造成影响,设置不同数量的神经元,初始值为 5,依次增加 5 个,终值为 30,通过 MAPE 选择最佳隐含层下的预测结果。

(5) 负荷计算。将各子区域每年的预测结果相加^[23],即可得到待预测区域每年的总负荷预测值 P_N 。

4 实例分析

以上海市局部区域为例,将该待测区域划分为 4 个子区域,以 A、B、C 和 D 表示,然后搜集 2010 年—2019 年的数据,每年数据包含该区域用电量和 10 个特征变量,采用文中方法进行实验分析。

4.1 特征选择分析

特征选择的过程如下。以子区域 A 为例,将样本中用电量作为标签值,计算其对数损失函数的一阶导数和二阶导数,并与样本中的各个特征一一对应,损失函数中初始预测概率默认设置为 0.5。建立第一棵树时,每个特征均有与 10 个样本对应的 10 个取值,因此按 10 种方式将取值划分成左子节点集合和右子节点集合。计算第一个特征每种划分方式对应的一阶导数之和以及二阶导数之和,进而计算增益,并选择出最大增益。按照上述方式依次计算其余特征对应的最大增益。选择 10 个特征分别对应的最大增益中的最大值,树节点按照该最大值对应特征的集合划分方式进行分裂。对左、右子节点集合分别进行分裂操作,循环上述过程。树构建完成后,计算叶子节点的权重值。更新损失函数中的预测概率,按照上述方式建立其余的树。最后根据特征分裂的次数(即为权重)和该特征在分裂节点产生的增益之和求得特征平均增益。

XGBoost 参数设置如表 1 所示。表中, g 为正则化项控制叶子节点数量复杂度的系数。

表 1 XGBoost 参数设置
Table 1 Parameter setting of XGBoost

参数	第一层数值	第二层数值
g	0	0
子样本	0.8	0.8
最大深度	10	4
采样比例	0.9	0.9
正则化惩罚系数	1	1
正则化项	0	0
最少样本权重	1	1

对每个子区域的 10 个特征变量通过 XGBoost 评分排序,权重和增益取值越高,特征变量分数越高。文中根据权重绘制特征得分,排序结果如图 4 所示,其中一些得分为 0 的特征未显示。

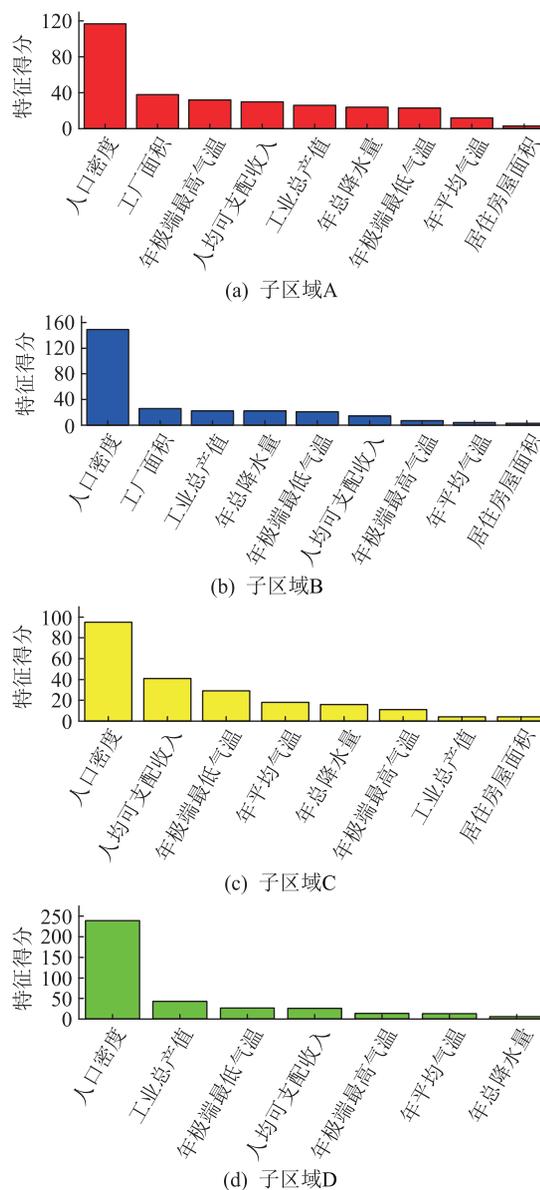


图 4 各子区域特征得分排序

Fig.4 Ranking of feature scores of each sub region

将各子区域 10 a 的数据划分为训练集和测试集,再通过 XGBoost 回归模型的预测结果筛选出最合适的特征组合,预测结果的 MAPE 越小,则该特征组合预测精度越高,特征变量个数与 MAPE 之间的关系如图 5 所示。

由图 5 可以确定各子区域最终的特征变量,具体如表 2 所示。表 2 中各子区域最佳的特征变量个数不同,这是由于每个子区域经济发展情况以及人口、气候等因素对负荷影响的重要程度不同。在空间负荷预测过程中,如果考虑的特征变量不足或过

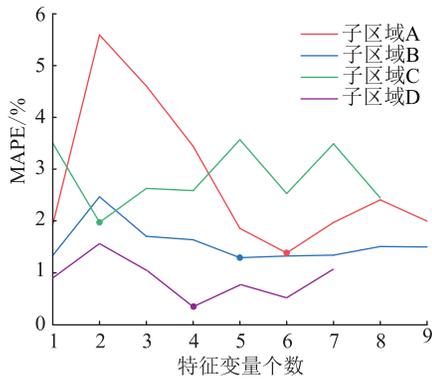


图5 特征变量个数与 MAPE 间的关系
Fig.5 Relationship between the number of characteristic variables and MAPE

多,都会影响空间负荷预测精度,因此需要根据实际情况选择特征变量。

表2 各子区域选择的特征变量
Table 2 Characteristic variables selected for each sub region

子区域	特征变量
A	人口密度、工厂面积、年极端最高气温、人均可支配收入、工业总产值、年总降水量
B	人口密度、工厂面积、工业总产值、年总降水量、年极端最低气温
C	人口密度、人均可支配收入
D	人口密度、工业总产值、年极端最低气温、人均可支配收入

4.2 GAN 数据增强分析

为了确定生成训练样本的最佳数量,文中以子区域 A 为例进行测试。经特征选择之后,用无数据生成的原始训练集与加入不同倍数生成样本后的增强训练集进行对比。以 MAPE 和均方误差^[24-31] (mean square error, MSE) 作为指标评价结果的优劣,如表 3 所示。

由表 3 可知,与无数据增强的原始训练集相比,增强训练集的样本数量可以提高模型的预测精度。当加入的生成样本增加到 10 倍时,MAPE 从原始的 6.87% 降到了 1.16%,MSE 也从 275.03 (kW·h)² 大幅降到了 4.94 (kW·h)²。而继续增加生成样本后,

表3 加入不同倍数生成样本的 ELM 预测精度对比
Table 3 Comparison of ELM prediction accuracy with different multiple generated sample sizes

加入生成样本的倍数	MAPE/%	MSE/(kW·h) ²
0	6.87	275.03
2	5.59	144.26
4	4.16	92.10
6	2.57	29.16
8	1.54	9.52
10	1.16	4.94
12	1.24	10.13
14	1.77	17.03
16	1.55	11.15

MAPE 和 MSE 的值又小幅度上升。因此,文中选择 10 倍的生成样本在 4 个子区域上分别增强训练集样本。

4.3 空间负荷预测结果

经过上述处理后,再按照第 3 章的步骤(4)——步骤(5),即可实现空间负荷预测。将文中预测结果与未用 XGBoost 特征选择的 GAN-ELM 模型和无 GAN 增强训练集样本的预测结果进行对比,如表 4 所示。表 4 中, *e* 表示相对误差。

由表 4 可以看出,文中方法预测值的相对误差分别为 0.09%、0.56%、1.40%、0.51%,文中方法预测精度最高且其 MAPE 最小,分别为 0.37%、1.51%、1.27%、1.69%。由此可以看出,用 XGBoost 进行特征选择对模型的预测精度有着重要影响,同时用 GAN 增强训练样本也能提升网络的学习效果,提高模型的预测精度。

文中方法是在双层 XGBoost 预测模型特征选择的基础上实施预测的,为了验证文中方法具有更好的预测精度,与双层 XGBoost 预测模型、文献[5]中的灰色关联分析-最小二乘支持向量机 (grey relational analysis-least squares support vector machine, GRA-LSSVM) 方法进行对比,2 种方法的预测结果如表 5 所示。

由表 4 中的文中方法和表 5 中的 2 种方法对比

表4 不同方法的负荷预测结果

Table 4 Load prediction results of different methods

年份	实际值/ 10 ⁸ (kW·h)	文中方法			无 XGBoost 的 GAN-ELM			无 GAN 增强		
		预测值/ 10 ⁸ (kW·h)	<i>e</i> /%	MAPE/%	预测值/ 10 ⁸ (kW·h)	<i>e</i> /%	MAPE/%	预测值/ 10 ⁸ (kW·h)	<i>e</i> /%	MAPE/%
2019	725.81	726.01	0.09	0.37	718.68	0.98	1.27	720.69	0.71	1.72
2018	688.17	684.35	0.56	1.51	700.80	1.84	9.78	696.41	1.20	5.09
2017	655.29	664.47	1.40	1.27	682.50	4.15	4.30	643.69	1.77	3.37
2016	639.09	642.36	0.51	1.69	668.45	4.59	3.84	633.97	0.80	1.20

表 5 2 种方法的预测结果

Table 5 Prediction results of two methods

年份	双层 XGBoost			GRA-LSSVM		
	预测值/ 10 ⁸ (kW·h)	e/%	MAPE/%	预测值/ 10 ⁸ (kW·h)	e/%	MAPE/%
2019	716.05	1.35	1.19	707.66	2.50	20.20
2018	683.97	0.61	1.64	710.81	3.29	15.73
2017	640.96	2.18	1.45	689.49	5.22	16.37
2016	649.41	1.62	1.71	669.22	4.71	16.40

可以看出,文中方法每年的相对误差和 MAPE 均低于其他 2 种方法,整体预测精度最高,其次是双层 XGBoost 模型,最后是 GRA-LSSVM 方法。文中所提方法相比于其他方法具有更高的预测精度,在特征选择和数据不足的场景中更具实用性。

5 结语

空间负荷预测受到众多特征变量的影响,同时面临着数据匮乏的问题,文中提出了一种基于双层 XGBoost 和数据增强的空间负荷预测方法。其中双层 XGBoost 模型可以在小样本情况下进行特征筛选,能根据各子区域中的实际情况选择特征变量,具有更强的泛化能力。同时 GAN 可以充分学习负荷与特征之间的映射关系,生成符合历史数据的新样本,有效地增强各子区域的训练样本。实验结果表明,文中方法可以同时解决最佳特征选择和样本量不足的问题,具有更高的预测精度,可以应用于数据匮乏或新建成区域的配电网精益化管理中。

致 谢

本文得到上海市科委地方院校能力建设项目(20020500700)资助,谨此致谢!

参考文献:

[1] 黄庆键,欧周,林佳亮. 关于空间电力负荷预测方法综述及展望[J]. 自动化应用,2017(2):79-81.
HUANG Qingjian,OU Zhou,LIN Jialiang. Review and prospect of space power load forecasting methods[J]. Automation Application,2017(2):79-81.

[2] 王梦蔚,晏阳. 空间负荷预测法的应用实例及要点分析[J]. 电气自动化,2019,41(2):42-44,55.
WANG Mengwei,YAN Yang. Application examples and main point analysis for the spatial load forecasting[J]. Electrical Automation,2019,41(2):42-44,55.

[3] 梁荣,杨波,马润泽,等. 利用多源信息和深度置信神经网络的配电系统空间负荷预测[J]. 电力建设,2018,39(10):12-19.
LIANG Rong,YANG Bo,MA Runze,et al. Spatial electric load forecasting for distribution systems using multi-source information and deep belief network-deep neural network [J].

Electric Power Construction,2018,39(10):12-19.

[4] 肖白,杨修宇,穆钢,等. 基于多变量分析的城市电网空间负荷预测方法[J]. 东北电力大学学报,2013,33(S1):39-44.
XIAO Bai,YANG Xiuyu,MU Gang,et al. Spatial electric load forecasting of urban power system based on multivariate analysis [J]. Journal of Northeast Dianli University,2013,33(S1):39-44.

[5] 唐玮,钟士元,舒娇,等. 基于 GRA-LSSVM 的配电网空间负荷预测方法研究[J]. 电力系统保护与控制,2018,46(24):76-82.
TANG Wei,ZHONG Shiyuan,SHU Jiao,et al. Research on spatial load forecasting of distribution network based on GRA-LSSVM method [J]. Power System Protection and Control,2018,46(24):76-82.

[6] 刘业峰,王婷. 基于 GRA-LSSVM 密度法的配电网空间负荷预测方法研究[J]. 计算机测量与控制,2018,26(11):256-260.
LIU Yefeng,WANG Ting. Research on spatial load forecasting of distribution network based on GRA-LSSVM density method [J]. Computer Measurement & Control,2018,26(11):256-260.

[7] 肖白,刘庆永,牛强,等. 基于元胞负荷特性分析的 RBF 神经网络空间负荷预测方法[J]. 电网技术,2018,42(1):301-307.
XIAO Bai,LIU Qingyong,NIU Qiang,et al. A spatial load forecasting method based on RBF neural network and cellular load characteristics analysis[J]. Power System Technology,2018,42(1):301-307.

[8] 肖白,张小娜,姜卓,等. 考虑本位元胞接受能力和相邻元胞负荷影响的空间负荷预测[J]. 电力系统自动化,2021,45(12):57-64.
XIAO Bai,ZHANG Xiaona,JIANG Zhuo,et al. Spatial load forecasting considering acceptability of standard cell and influence of load of adjacent cells[J]. Automation of Electric Power Systems,2021,45(12):57-64.

[9] 郑伟民,叶承晋,张曼颖,等. 基于 Softmax 概率分类器的数据驱动空间负荷预测[J]. 电力系统自动化,2019,43(9):117-124.
ZHENG Weimin,YE Chengjin,ZHANG Manying,et al. Data-driven spatial load forecasting method based on Softmax probabilistic classifier [J]. Automation of electric power system,2019,43(9):117-124.

[10] 肖白,穆冠男,姜卓,等. 计及城市发展程度的多阶段空间负荷预测方法[J]. 电网技术,2019,43(7):2251-2257.
XIAO Bai,MU Guannan,JIANG Zhuo,et al. Multi-stage spatial load forecasting method based on urban development degree[J]. Power System Technology,2019,43(7):2251-2257.

[11] 马临超,齐山成,牛赛,等. 考虑小区发展不均衡性和不确定性的多阶段空间负荷预测[J]. 电力系统保护与控制,2021,49(1):91-97.
MA Linchao,QI Shancheng,NIU Sai,et al. Multi-stage spatial load forecasting considering the imbalance and uncertainty of

- the development of the sub-area[J]. *Power System Protection and Control*, 2021, 49(1):91-97.
- [12] 肖白, 黄钰茹, 姜卓, 等. 数据匮乏场景下采用生成对抗网络的空间负荷预测方法[J]. *中国电机工程学报*, 2020, 40(24):7990-8001, 8236.
- XIAO Bai, HUANG Yuru, JIANG Zhuo, et al. The method of spatial load forecasting based on the generative adversarial network for data scarcity scenarios[J]. *Proceedings of the CSEE*, 2020, 40(24):7990-8001, 8236.
- [13] 郭毅博, 牛猛, 王海迪, 等. 基于生成对抗网络的飞机燃油数据缺失值填充方法[J]. *浙江大学学报(理学版)*, 2021, 48(4):402-409.
- GUO Yibo, NIU Meng, WANG Haidi, et al. An aircraft fuel data missing value filling method with generative adversarial network[J]. *Journal of Zhejiang University (Science Edition)*, 2021, 48(4):402-409.
- [14] 张睿, 谭江浩, 杨瑞, 等. 基于相空间重构和 EMD-ELM 的短期电力负荷预测[J]. *电气时代*, 2022(1):103-107.
- ZHANG Rui, TAN Jianghao, YANG Rui, et al. Short-term power load forecasting based on phase space reconstruction and EMD-ELM[J]. *Electric Age*, 2022(1):103-107.
- [15] 邓燕国, 王冰, 曹智杰, 等. 基于熵权法与 GRA-ELM 的配电网空间负荷预测[J]. *电力工程技术*, 2021, 40(4):136-141.
- DENG Yanguo, WANG Bing, CAO Zhijie, et al. Spatial load forecasting of distribution network based on entropy weight method and GRA-ELM[J]. *Electric Power Engineering Technology*, 2021, 40(4):136-141.
- [16] 肖白, 刘庆永, 房龙江, 等. 基于模糊粗糙集理论和时空信息的空间负荷预测[J]. *电力建设*, 2017, 38(1):58-67.
- XIAO Bai, LIU Qingyong, FANG Longjiang, et al. Spatial load forecasting based on fuzzy rough set theory with spatial and temporal information[J]. *Electric Power Construction*, 2017, 38(1):58-67.
- [17] 肖白, 张婕, 姜卓, 等. 基于秩次集对分析理论的空间负荷预测方法[J]. *电力自动化设备*, 2020, 40(4):153-158.
- XIAO Bai, ZHANG Jie, JIANG Zhuo, et al. Spatial load forecasting method based on rank set pair analysis[J]. *Electric Power Automation Equipment*, 2020, 40(4):153-158.
- [18] 肖白, 赵晓宁, 姜卓, 等. 利用模糊信息粒化与支持向量机的空间负荷预测方法[J]. *电网技术*, 2021, 45(1):251-260.
- XIAO Bai, ZHAO Xiaoning, JIANG Zhuo, et al. Spatial load forecasting method using fuzzy information granulation and support vector machine[J]. *Power System Technology*, 2021, 45(1):251-260.
- [19] 肖白, 杨欣桐, 田莉, 等. 计及元胞发展程度的空间负荷预测方法[J]. *电力系统自动化*, 2018, 42(1):61-67.
- XIAO Bai, YANG Xintong, TIAN Li, et al. Spatial load forecasting method based on development degree of cell[J]. *Automation of Electric Power Systems*, 2018, 42(1):61-67.
- [20] EVANGELOPOULOS V, KARAFOTIS P, GEORGILAKIS P. Probabilistic spatial load forecasting based on hierarchical trending method[J]. *Energies*, 2020, 13(18):4643.
- [21] DB A, MVC A, SP B, et al. Forecasting peak electricity demand for Los Angeles considering higher air temperatures due to climate change[J]. *Applied Energy*, 2019, 236:1-9.
- [22] 杨军胜, 彭石, 王承民, 等. 基于城市用地性质的配网空间负荷预测研究[J]. *电测与仪表*, 2018, 55(11):30-34.
- YANG Junsheng, PENG Shi, WANG Chengmin, et al. Research of spatial load forecasting in distribution network based on grid partition of urban land-use property[J]. *Electrical Measurement & Instrumentation*, 2018, 55(11):30-34.
- [23] 吴争荣, 孔祥玉, 董旭柱, 等. 基于配用电信息分区分类的短期空间负荷预测[J]. *电力系统及其自动化学报*, 2019, 31(2):26-31.
- WU Zhengrong, KONG Xiangyu, DONG Xuzhu, et al. Short-term spatial load forecasting based on partition and classification of power distribution information[J]. *Proceedings of the CSU-EPSC*, 2019, 31(2):26-31.
- [24] VIEIRA D A G. Large scale spatial electric load forecasting framework based on spatial convolution[J]. *International Journal of Electrical Power & Energy Systems*, 2020, 117:105582.
- [25] 王秋实, 杨明, 李鹏, 等. 基于 LSTM-SAE 与支持向量机的窃电识别方法研究[J]. *电力信息与通信技术*, 2022, 20(9):51-58.
- WANG Qiushi, YANG Ming, LI Peng, et al. Studies on electricity theft detection approach based on LSTM-SAE and support vector machine[J]. *Electric Power Information and Communication Technology*, 2022, 20(9):51-58.
- [26] 郑乐, 徐青山, 冯小峰. 基于层次聚类算法与 ISA-LSSVM 的短期负荷预测研究[J]. *电力需求侧管理*, 2022, 24(5):51-57.
- ZHENG Le, XU Qingshan, FENG Xiaofeng. Short-term load forecasting based on hierarchical clustering algorithm and ISA-LSSVM[J]. *Power Demand Side Management*, 2022, 24(5):51-57.
- [27] 刘焯煊, 杨学良, 陶晓峰, 等. 基于改进 Prophet 算法的短期日负荷预测方法研究[J]. *电力需求侧管理*, 2022, 24(5):58-63.
- LIU Niexuan, YANG Xueliang, TAO Xiaofeng, et al. Short-term power load forecasting based on improved Prophet algorithm[J]. *Power Demand Side Management*, 2022, 24(5):58-63.
- [28] 马临超, 齐山成, 牛赛, 等. 考虑小区发展不均衡性和不确定性的多阶段空间负荷预测[J]. *电力系统保护与控制*, 2021, 49(1):91-97.
- MA Linchao, QI Shancheng, NIU Sai, et al. Multi-stage spatial load forecasting considering the imbalance and uncertainty of the development of the sub-area[J]. *Power System Protection and Control*, 2021, 49(1):91-97.
- [29] 付文杰, 李化, 杨伯青, 等. 基于集合卡尔曼滤波与相空间重构的负荷预测方法研究[J]. *电力需求侧管理*, 2022, 24(1):49-54.
- FU Wenjie, LI Hua, YANG Boqing, et al. Research of load forecasting method based on ensemble Kalman filter and phase-space reconstruction[J]. *Power Demand Side Management*,

2022,24(1):49-54.

Power Demand Side Management,2022,24(1):55-62.

[30] 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.

[31] 孙志翔,丁彬,孙晓燕. 基于迁移学习和 GRU 网络的新建小区负荷预测[J]. 电力需求侧管理,2022,24(1):55-62.

SUN Zhixiang, DING Bin, SUN Xiaoyan. New community load prediction based on transfer learning and GRU network [J].

作者简介:



黄冬梅

黄冬梅(1964),女,硕士,教授,博士生导师,研究方向为海洋与电力时空信息技术 (Email:dmhuang_dl@163.com);

张宁宁(1998),男,硕士在读,研究方向为空间负荷预测;

胡安铎(1983),男,博士,讲师,研究方向为电力时空信息技术。

Spatial load forecasting method based on double-layer XGBoost and data enhancement

HUANG Dongmei¹, ZHANG Ningning¹, HU Anduo¹, HU Wei², XIAO Yong³, CHEN Anqing⁴

(1. College of Electronic and Information Engineering, Shanghai University of Electric Power, Shanghai 201306, China;

2. College of Economics and Management, Shanghai University of Electric Power, Shanghai 201306, China;

3. College of Continuing Education, Shanghai University of Electric Power, Shanghai 200090, China;

4. State Grid Info-Telecom Great Power Science and Technology Co., Ltd., Fuzhou 350001, China)

Abstract: Spatial load forecasting faces the problems of multiple characteristic factors and data shortage. A spatial load forecasting method based on double-layer extreme gradient boosting (XGBoost) and data enhancement is proposed. Firstly, the area to be predicted is divided into several sub regions according to the supply range of feeder power. Secondly, a feature selection model based on double-layer XGBoost is constructed. The first layer XGBoost scores and sorts the features. The combined features are loaded into the second layer XGBoost for sub regional load forecasting. The best feature variables of each sub region are selected according to the load forecasting results. Then, the training set samples of each sub region are enhanced by the generative adversarial network (GAN), and the load of sub regions is forecasted through the extreme learning machine (ELM). Finally, the predicted values of sub regions are added to obtain the load of the region to be predicted. Taking local areas of Shanghai as an example, the simulated experiment and comparative analysis are carried out. The results show that the proposed method can solve the problems of characteristic variable selection and data shortage at the same time, and has high prediction accuracy.

Keywords: spatial load forecasting; extreme gradient boosting (XGBoost); feature selection; generative adversarial network (GAN); data enhancement; extreme learning machine (ELM)

(编辑 吴楠)