

改进型 XGBoost 算法模型的企业电力负荷电费优化

仲立军, 杨玉锐, 周晓琴, 牛中伟, 周子誉
(国网嘉兴供电公司 浙江 嘉兴 314000)

摘要: 针对现有技术中企业电力负荷电费优化效果差的问题, 提出了新型的企业电力负荷电费优化方法, 构建了改进型 XGBoost 算法模型。通过构建多个决策树提高了企业电力负荷的预测精度。在 XGBoost 算法模型的基础上, 应用逐步回归算法模型, 通过持续调整 XGBoost 算法模型中决策树的权重, 不断地训练输出的单个弱学习器。实现 XGBoost 算法模型输出的多个决策树学习器的加权求和, 从而实现了电力负荷电费预测数值的高精度输出。试验表明, 研究的方法精确度高。

关键词: 电力负荷; 电费优化; XGBoost 算法模型; 逐步回归算法模型; 权重; 弱学习器

DOI: 10.3969/j.issn.1000-3886.2021.03.023

[中图分类号] TM714 [文献标志码] A [文章编号] 1000-3886(2021)03-0072-04

Optimization of Enterprise Power Load and Electricity Bill Based on Improved XGBoost Algorithm Model

Zhong Lijun, Yang Yurui, Zhou Xiaoqin, Niu Zhongwei, Zhou Ziyu
(State Grid Jiaxing Power Supply Company, Jiaxing Zhejiang 314000, China)

Abstract: Aiming at the problem of poor optimization effect of enterprise power load and electricity bill in the prior art, a new method of enterprise power load and electricity bill optimization was proposed, and an improved XGBoost algorithm model was constructed. By constructing multiple decision trees, the forecasting accuracy of the power load of the enterprise was improved. On the basis of the XGBoost algorithm model, the stepwise regression algorithm model was applied, and the weight of the decision tree in the XGBoost algorithm model was continuously adjusted. Furthermore, the output of a single weak learner was steadily trained, and the multiple decision tree learners' output by the XGBoost algorithm model were weighted and summed, thereby achieving a high-precision output of the predicted value of the electricity load and electricity bill. The experiments show that the research method has high accuracy.

Keywords: power load; electricity charge optimization; XGBoost algorithm model; stepwise regression algorithm model; weight; weak learner

0 引言

在企业工业用电以及用户的用电中, 电费的整体架构通常包括基本电费、电度电费和功率因数调整电费等不同的形式。化工企业多数实行 24 h 连续生产, 对供电可靠性要求高, 多数采用双电源以保证供电质量^[1], 因此无法有效通过生产负荷分时段调整, 以利用分时电价差降低电度电费。现有技术中的企业功率因数在一定程度上虽然能够满足企业用电的需求, 但是无功率太多, 电费的优化空间相对比较小^[2]。在计算基本电费时, 需要根据企业用户的实际用电容量或者最大用电需量进行计算。在众多用电过程中, 企业用户难以对未来的用电负荷情况进行判断, 使得在选择基本电费计收方式时, 难免与自身实际用电负荷存在偏差, 容易引起基本电费虚高。这就需要一种方法实现企业电费优化。

1 负荷预测大数据模型设计

本文通过研究浙江某工业园区化工企业用电负荷特征, 根据企业月度用电最大需量波动率分组, 采用改进型 XGBoost 算法模型^[3] 构建化工企业月度用电最大需量精准预测模型。为企业合理选择基本电费计算方式提供指导, 帮助企业合理降低用电成本。采用的大数据模型如图 1 所示, 下面分别对不同的算法模型

进行说明。

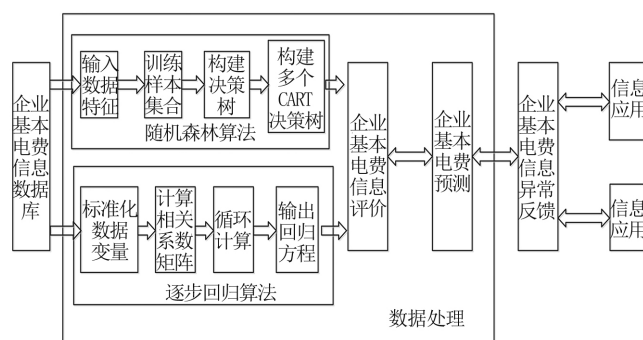


图1 负荷预测大数据模型设计

1.1 改进型 XGBoost 算法模型

本文采用改进型 XGBoost 算法模型, 在传统 XGBoost 算法模型的基础上加入逐步回归算法模型^[4], 具有以下技术优势。

- (1) 对大量的负荷数据比较敏感, 计算速度比较快, 数据效率输出比较高。
- (2) 模型构建较为快捷, 适用范围广, 在电力负荷电费预测和计算过程中表现出较好的适应性。
- (3) 克服了现有技术分类算法效率低和分类性差的问题, 提高了评估企业电力负荷电费的能力。

定稿日期: 2021-02-02

基金项目: 浙江省重大科技项目 (ZJ2020JKH00095395001)

本文采用改进型 XGBoost 算法模型的核心意义在于使用 CART 模型。首先对负荷预测模型进行定义,设定其输出的目标函数,目标函数由训练损失 $L(\theta)$ 、正则化 $\Omega(\theta)$ 两种不同的函数表示式构成,数学表达式如式(1)所示。

$$\text{Obj}(\theta) = L(\theta) + \Omega(\theta) \quad (1)$$

对上述函数求最小化,其本质是对训练损失 $L(\theta)$ 和正则化 $\Omega(\theta)$ 求最小化。通过图形使拟合曲线达到最佳,最终预测的负荷方差最小^[5]。输出的结果较为稳定。

评估目标为:将出现的损失函数在拟合曲线图中呈现欠拟合的形态,输出的优化正则化项图形表现为过拟合形态,使最后的输出函数达到最小值。这样就能够使用输出函数的模型实现较佳的评估效果。流程示意图如图 2 所示。

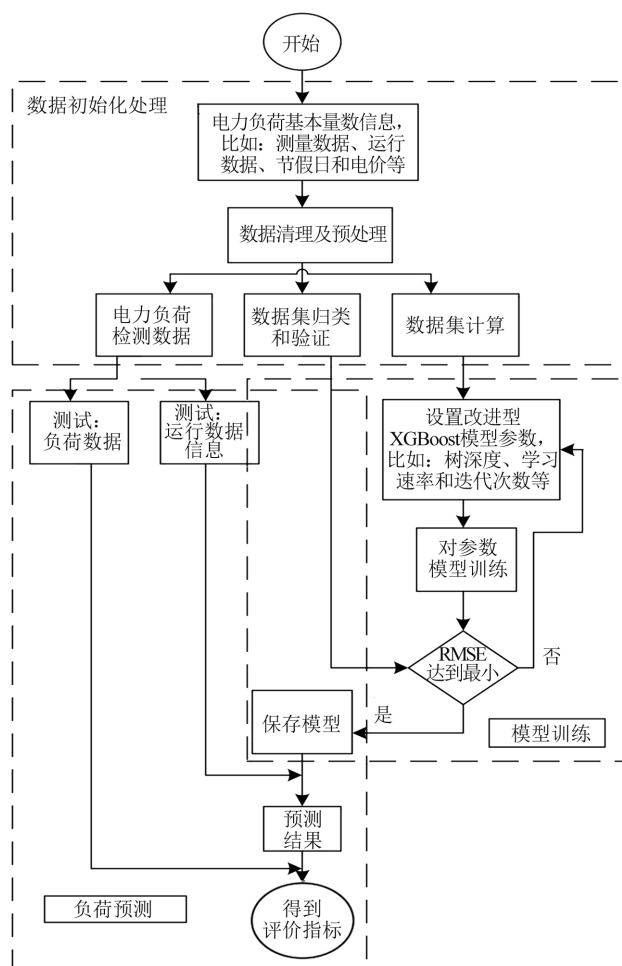


图 2 改进型 XGBoost 模型构建流程示意图

下面分步骤说明上述函数。

(1) 构建多个决策树,反复进行迭代计算,将每个决策树通过设置根节点和叶节点的方式,构建出最佳决策树模型。

(2) 在每次迭代开始之前,计算损失函数在每个训练样本点的一阶导数 g_i 和二阶导数 h_i 。

(3) 采用贪心算法将不同的决策树模型对不同的叶节点和根节点的计算输出的预测值进行评价。

(4) 将每次计算输出的决策树 $f_i(x)$ 通过迭代模型进行计

算,迭代模型公式为:

$$\hat{y}_i' = \hat{y}_i^{t-1} + f_i(x_i) \quad (2)$$

当出现多个决策树时,为了提高决策树的精度,通过以下函数进行优化:

$$\hat{y}_i' = \hat{y}_i^{t-1} + \alpha f_i(x_i) \quad (3)$$

式中: \hat{y}_i 为构建出的决策树; f 为所构建出的决策树中的单个决策树; x_i 为评估电力负荷中的不同参数值,比如电力测量数据、运行数据、节假日和电价等类型; α 为学习率参数^[6]。根据具体的选择需要设置 α 的数值,比如介于 0~5 之间。在进行电网负荷预测时,通过均方误差(mean squared error, MSE) 值对损失函数进行评价。MSE 评估方法包括一阶、二次两种不同的项,构建出的损失函数表达式为:

$$L(\theta) = \sum_i (y_i - \hat{y}_i)^2 \quad (4)$$

改进型 XGBoost 算法模型的负荷预测模型可以用以下公式表示。

$$\text{Obj}^{(t)} = \sum_{i=1}^n L(y_i - [\hat{y}_i^{t-1} + f_i(x_i)])^2 + \Omega(f_i) = \sum_{i=1}^n [2(\hat{y}_i^{t-1} - y_i)f_i(x_i) + f_i(x_i)^2] + \Omega(f_i) \quad (5)$$

式中: $(\hat{y}_i^{t-1} - y_i)$ 为预测输出残差。式(5)表示在进行第 t 步目标时,构建出的 t 棵树要输出的目标函数。

通常,在进行负荷评估时,将改进型 XGBoost 算法模型内的各种决策树模型进行固定设置,令 $q(x) = K$,对 $\text{Obj}^{(t)}$ 进行求导,其输出的一阶导数等于 0,决策树上的叶子节点 j 的参数值可通过以下函数表示。

$$\omega_j^* = \frac{G_j}{H_j + \lambda} \quad (6)$$

式中: G_j 为叶子节点 j 的所有输入样本的一阶导之和; H_j 为叶子节点 j 所有输入样本的二阶导之和; λ 为正则化系数。 ω 为构建出的决策树输出的分数向量。

则对电力负荷电费评估和优化的目标函数为:

$$\text{Obj} = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (7)$$

式中: T 为构建出的决策树中每个叶的节点数量。

通过该方法,能够将电力负荷运行情况转化为大数据模型的方式表示,提高了数据计算的直观能力。通过不断地调整数据权重能够不断地训练单个弱学习器,有助于纠正和调整弱学习器输出的数据残差,将原始构建出的多个不同决策树学习器进行加权求和,最终输出较为准确的预测数值。

为了进一步提高上述计算精度,本文采用逐步回归算法模型对上述算法进一步修正,以提高评估精度。

1.2 逐步回归算法模型

对逐步回归算法模型的构建方法进行说明,假设存在 n 个不同的企业基本电费类型。企业基本电费类型中的种类为 m ,每种 m 个数据类型中的数据量为 p ,反映用户实际用电情况的矩阵为 $A_{n \times p}$,用户理论用电矩阵为 $C_{n \times m}$,评估误差矩阵为 $E_{n \times m}$,则引出关系式为:

$$C_{n \times m} = A_{n \times p} P_{p \times m} + E_{n \times m} \quad (8)$$

$$P_{p \times m} = (A_{p \times n}^T A_{n \times p})^{-1} A_{p \times n}^T C_{n \times m} \quad (n \geq p) \quad (9)$$

在对个别企业基本电费进行数据分析时,可以令 $m = 1$,此时,式(10)可以转化为:

$$\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} 1 & a_{11} & \cdots & a_{p1} \\ 1 & a_{12} & \cdots & a_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & a_{1n} & \cdots & a_{pn} \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ \vdots \\ p_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (10)$$

式中: a_{ki} 为企业基本电费的变量数据之间是相关系数为 1 的线性相关量。当 $r(A^T A) < p$ 时, $(A^T A)^{-1}$ 为无解^[7]。如果 a_{ki} 中企业基本电费变量中的两个具有比较大的相关性时,则需要核对算法模型中的数据变量,去除无用数据,以免数据模型不准确^[8]。

通过上述模型的构建,逐步回归算法模型内的总离差平方和与上述回归方差平方和关系可以用以下关系式表示。

$$S_{\text{总}}(t) = S_{\text{回}}(t) + S_{\text{残}}(t) \quad (11)$$

$$S_{\text{回}}(t) = S_{\text{总}}(t) - S_{\text{残}}(t) \quad (12)$$

其中:

$$S_{\text{总}}(t) = \sum_{i=1}^n (c_i - \bar{c})^2; S_{\text{回}}(t) = \sum_{i=1}^n (\hat{c}_i - \bar{c})^2; S_{\text{残}}(t) = \sum_{i=1}^n (c_i - \hat{c}_i)^2 \quad (13)$$

式中: \bar{c} 为在利用上述数据模型进行回归时,相同属性存在平均值的大小量; c_i 为第 i 种企业基本电费属性; \hat{c}_i 为第 i 种企业基本电费属性回归的所属属性。

当在企业基本电费的数据信息中剔除一个数据 a_i ,则函数中的 $S_{\text{回}}(t)$ 转变为 $S_{\text{回}}(t-1)$,再通过公式 $\Delta_i(t) = S_{\text{回}}(t) - S_{\text{回}}(t-1)$ 表示异常的数据信息 a_i 对分类属性 c 的总方差贡献值。然后调取数据统计量:

$$F_i = \frac{\Delta_i(t)}{S_{\text{回}}(t) / (n - t - 1)} \quad (14)$$

通过融入回归算法模型后,可以核对 a_i 对分类属性 c 的影响程度。通过引入异常数据去除变量 \tilde{F} ,使得企业基本电费的最终统计量 F_i 能够按照 $F(t, n-t-1)$ 进行分布,最终使得 $F_a(t, n-t-1) = \tilde{F}$,在 $F_i < \tilde{F}$ 的情况下,表示在众多的企业基本电费中的数据信息剔除了异常数据。这种方法使得评估精度大大提高^[9]。由线性回归原理可知,当研究对象变动较为平稳,受随机因素影响较小,使用该模型效果较好。

2 案例分析及验证

本文主要验证改进型 XGBoost 算法模型的工作效率。仿真试验时,工作环境为 Python3.5。该算法模型的参数为:每个决策树的树深度取 6,学习效率为 0.43,进行 400 次迭代计算^[10]。决策树剪枝后的数值为 0.3,决策树每个叶子输出的权重值为 6,其中对决策树进行随机采样输出的数值比例为 0.7。改进型 XGBoost 算法模型的正则化被命名为 L2,将本文的改进型 XGBoost 算法模型与 BP、GRNN 算法分别进行对比验证。

本文实例数据为浙江某工业园区内化工企业 2015 年 1 月至

2018 年 8 月最大

需量及影响因素

指标数据。以需

量变异系数 0.08

为阈值,变异系数

>0.08 为波动较

大组,变异系数

≤0.08 为波动较

小组。园区两组

类型企业 2018 年

8 月需量数据如表 1 所示。

通过设置如表 1 所示的数据信息,进行数据分析。

本文以园区企业 2015 年 1 月至 2017 年 12 月的需量及其相

关影响指标数据作

为训练集,进行模

型训练;以 2018

年 1 月至 8 月的

数据为验证集,对

模型效果进行评

价。样本数据表

如表 2 所示。

通过数据采

集和设置,首先需

要对获取的数据

信息进行预处理,归一化处理函数为:

$$\bar{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (15)$$

在进行精确度评价时,采用均方根误差 (root mean square error, RMSE) 和平均绝对误差百分比 (mean absolute percent error, MAPE),其中均方根 RMSE 误差数据模型为:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (16)$$

平均绝对误差百分比 MAPE 模型可以为:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (17)$$

式中: y_i 为在评估时间为 i 时刻下的真实电网负荷值; \hat{y}_i 为在 i 时刻下所采用模型的预测负荷值。RMSE 和 MAPE 值越小表示预测值越准,所采用的模型效果越好。

通过对改进型 XGBoost 算法模型进行参数设置,在具体应用时,需要调用 XGBoost 函数库中的 CV 函数进行数据信息计算,通过 400 次的迭代计算,输出如图 3 所示的误差曲线示意图。

通过图 3 可以看到,在经历一段时间的运行后,为了避免模型在运行过程中出现过拟合现象,将改进型 XGBoost 算法模型的迭代次数取值 230。

下面以几种简单的影响因素作为示例性分析,如图 4 所示。

通过图 4 可以看到,负荷影响因素不同,则均方根值不同,各个影响因素的分布比例都不相同。进而可以直观地看出不同因素的影响。

表 1 样本 1 园区部分企业

2018 年 8 月最大需量数据

波动分组	园区企业	变异系数	最大需量
平稳组 (min)	企业 min1	0.013 56	4 241.60
	企业 min2	0.033 81	5 980.00
	企业 min3	0.052 93	2 438.00
波动组 (max)	企业 max1	0.195 60	96.48
	企业 max2	0.315 68	184.80
	企业 max3	0.393 09	944.80

表 2 样本 2 试验数据表

参数	含义
X_1	记录的负荷值
$X_2 \sim X_4$	当天温度 (平均、高、低)
$X_5 \sim X_7$	时间时刻
输入 $X_8 \sim X_9$	季节、温湿度
$X_{10} \sim X_{11}$	日期、假期
$X_{12} \sim X_{13}$	天气条件
X_{14}	电价
输出 Y	评估模型输出的具体时刻的负荷

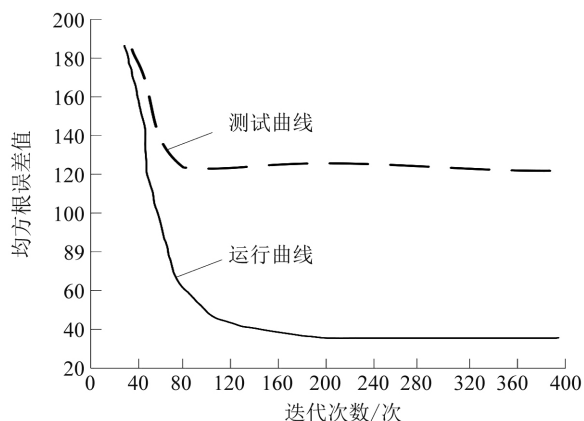


图 3 负荷预测均方根误差曲线示意图

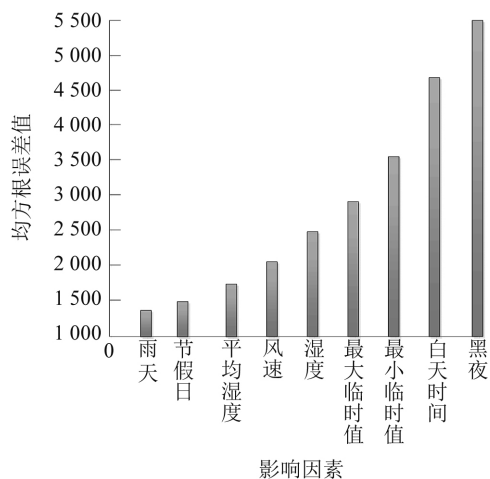


图 4 改进型 XGBoost 算法模型中不同影响因素的特征重量级分布图

将 BP、GRNN 算法分别与本文的改进型 XGBoost 算法模型进行对比分析,得出如图 5 所示的对比示意图。

分析图 5 可知,改进型 XGBoost 算法模型预测准确度较高,且较为稳定,更适用于电力超短期负荷预测。

3 结束语

本文根据当前电力负荷预测的需要,提出了新型的负荷预测模型,应用了改进型 XGBoost 算法模型,通过调整权值的方式实现弱分类器和学习器的训练和学习,提高了数据输出的精确度。通过试验,本文方法能够直观地看出负荷影响因素,误差低,能为企业电力负荷的预测做出一定的贡献,但是尚且存在其他不足,

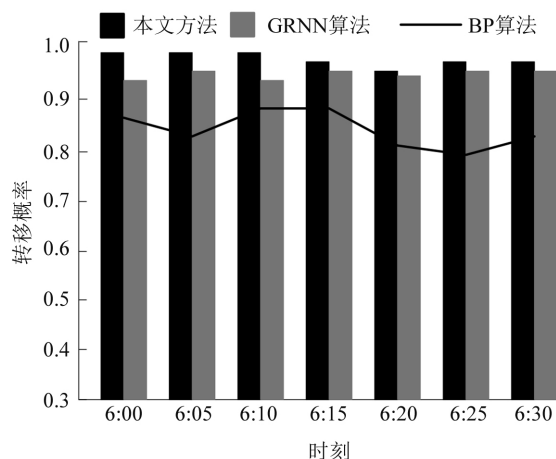


图 5 不同方法的预测准确性分析对比示意图

需要进行进一步的研究。

参考文献:

- [1] 拜润卿,何欣,赵红,等. 基于粒子群优化方法的短期电力负荷组合预测[J]. 电子测试,2015,2(4): 46-48.
- [2] 刘南艳,牟丰. NRS 和 PSO 算法优化最小二乘支持向量机的短期电力负荷预测[J]. 现代电子技术,2019,42(7): 115-118;124.
- [3] 张震,徐子怡,张龙,等. 基于小波神经网络的热负荷预测方法[J]. 自动化技术与应用,2016,35(5): 6-9.
- [4] 孔祥玉,郑锋,鄂志君,等. 基于深度信念网络的短期负荷预测方法[J]. 电力系统自动化,2018,42(5): 133-139.
- [5] 黄毅成,杨洪耕. 改进遗传 K 均值算法在负荷特性分类的应用[J]. 电力系统及其自动化学报,2014,26(7): 70-75.
- [6] 金士琛,薛会,林霞,等. 基于 FDNN 的电力系统短期负荷预测模型研究[J]. 青岛大学学报(工程技术版),2017,32(4): 12-16.
- [7] CHEN T, GUESTRIN C. XGBoost: a scalable tree boosting system [C]. //ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016: 785-794.
- [8] 张智晟,于道林. 考虑需求响应综合影响因素的 RBF-NN 短期负荷预测模型[J]. 中国电机工程学报,2018,38(6): 1631-1638;1899.
- [9] 刘畅,刘天琪,陈振寰,等. 基于 KPCA 和 BP 神经网络的短期负荷预测[J]. 电测与仪表,2016,53(10): 57-61.
- [10] TAIEB S B, HYNDMAN R J. A gradient boosting approach to the Kaggle load forecasting competition [J]. International Journal of Forecasting, 2014, 30(2): 382-394.

【作者简介】仲立军(1975—)男,江西赣州人,高级经济师,主要研究方向:企业运营监测及分析管理通信。

为了你和你的家人的健康,请不要吸烟。
吸烟是继战争、饥饿和瘟疫之后,对人类生存的最大威胁。