基于GBDT算法的电力专项规划负荷预测研究

谢中元

（上海同济城市规划设计研究院有限公司，上海 200080）

摘要：负荷预测是电力专项规划的重要组成部分，是规划电力设施的基础。传统的电力规划负荷预测普遍存在准确度不高、预测结果相对误差较大等问题。采用了一种基于梯度提升决策树 (Gradient Boosted Decision Tree，GBDT) 算法的负荷预测新方法，与传统的负荷预测方法相比，该方法的预测结果整体准确度更高，在电力规划中具有较高的实际参考价值。

关键词：电力规划；负荷预测；回归树；残差

Research on load forecasting of electric power special planning based on GBDT algorithm

XIE Zhongyuan

(Shanghai Tongji Urban Planning and Design Institute Co., Ltd., Shanghai 200080, China)

Abstract: Load forecasting is an important part of power special planning and the basis for planning power facilities. Traditional power planning load forecasting generally has problems such as low accuracy and relatively large forecast results error. This paper adopts a new load forecasting method based on the GBDT algorithm. Compared with the traditional load forecasting method, the overall accuracy of the forecasting results of this method is higher, which means higher actual reference value.

Key words: power planning; load forecasting; regression tree; residual

0 引言

电力专项规划是对城市未来用电规模和用电容量的提前部署，是城市发展规划的重要组成部分，负荷预测是电力规划的重要工作内容，对高、中压电网设施建设的控制和预留起着至关重要的作用。电力规划常进行中、长期负荷预测，如何提高预测的准确率一直以来是热点问题。传统的电力规划负荷预测方法主要有产值单耗法、电力弹性系数法和自回归移动平均法等，这些方法简单考虑了地方经济发展速度、用地功能划分等因素[1]，操作比较方便，但不够准确。文献[2-3]采用了长短期记忆（Long Short Term Memory，LSTM）网络模型，考虑了不同时间尺度的历史负荷、地区经济等数据，该方法有效提高了预测的准确性，但应用范围不够广泛。

本文采用了一种基于改进 GBDT算法的负荷预测方法，以《邓州市电力设施专项规划》为案例进行研究，结果表明与传统的负荷预测方法相比，该方法预测结果更加接近电网负荷的实际值，整体效果更好。

1 GBDT负荷预测模型

GBDT算法通过训练回归树1 (弱学习器) 得出样本残差1，将残差1输入回归树2经训练后得

出残差2，再将残差2输入回归树3重复前一个步骤直至运算结束，将各回归树的预测值与权重系数的乘积之和相加后得到最终结果，算法原理如图1所示，GBDT 算法的详细内容见文献[4]。

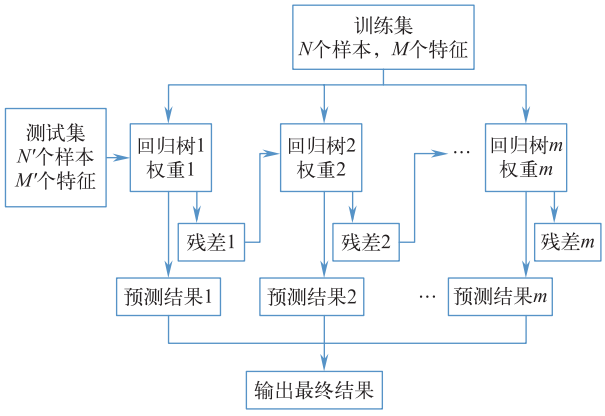


图 1　梯度提升决策树算法原理

Fig. 1 General architecture of gradient boosting decision tree algorithm

GBDT模型是多项式回归模型，采用加法结构，最终输出结果为

$$\begin{aligned} F_m &= \beta_0 + \beta_1 f_1(x_0,\ x_1,\ \cdots,\ x_i) + \\ &\cdots + \beta_m f_m(x_0,\ x_1,\ \cdots,\ x_i) \\ &= F_{m-1} + \beta_m f_m(x_0,\ x_1,\ \cdots,\ x_i) \end{aligned} \tag{1}$$

式中，β_1，…，β_m为各回归树的权重系数；β_0为初始值；$f_1(\cdot)$，…，$f_m(\cdot)$为各回归树的预测值，权重系数β_m的求解方法为

$$\begin{aligned} \beta_m = \arg\min \sum_{i=1}^{m-1} E\big[y,\ F_{m-2}(x_0, x_1, \cdots, x_i) + \\ \beta_{m-1} f_{m-1}(x_0,\ x_1,\ \cdots,\ x_i)\big] \end{aligned} \tag{2}$$

模型的损失函数采用均方差函数，可表示为

$$E = \frac{1}{2}(\hat{y} - y)^2 \tag{3}$$

式中，$\hat{y}$为预测值；y为真实值。

损失函数的最速下降方向为负梯度，用$-g_m$表示，表达式为

$$-g_m = -\frac{\partial E(y_i, f_m)}{\partial f_m} \tag{4}$$

模型中第m轮的残差实际上是第$m-1$轮的梯度，表达式为

$$-g_{m-1} = y_i - f_m \tag{5}$$

文中模型以历史负荷数据、地方经济发展数据和气候特征数据作为训练集，用式（6）表示。将训练集纳入算法模型中进行优化训练，调参优化后得到的最终结果作为输出值。该模型能够拟合非线性可分的多组数据，灵活处理不同特征因素之间的复杂关系，但需要根据历史经验提前设置变量参数，参数选择不当时可能造成过拟合现象[5]。

$$D = \{(x_{n1},\ x_{n2},\ \cdots,\ x_{ni}),\ y_n\} \tag{6}$$

2　样本特征提取

2.1　特征因素

城市经济发展迅速，用电需求量快速增长，国民生产总值逐年提高。本次负荷预测研究综合考虑地区经济发展速度、居民生活水平、用电量和温度等因素的影响。文中模型提取的经济数据如表1所示，自然环境数据为年最高温、年最低温和年平均温，共27个影响因素作为输入特征，将以上特征值用特征向量可表示为（x_1，x_2，…，x_{27}）$\in R^{27N}$。

表 1　模型输入的经济数据

Tab. 1 Economic data entered by the model

经济数据	经济数据
第一产业增加值	人口自然增长率
第二产业增加值	人均 GDP
第三产业增加值	年最大负荷
第一产业就业人员	年总用电量
第二产业就业人员	人均用电量
第三产业就业人员	城镇居民人均可支配收入
地方财政总收入	城镇居民人均消费支出
金融机构存款余额	农村居民人均可支配收入
金融机构贷款余额	农村居民人均消费支出
户籍人口	居民消费价格指数 CPI
常住人口	社会消费品零售总额
固定资产投资	城镇失业率

2.2　特征值提取

（1）经济因素特征的提取

从政府统计和供电公司相关网站上提取表1中的数据，将其分别存入特征向量的第1～24列。这些指标从宏观上反映社会、经济和人口的

发展情况，发展加快时用电需求量也会相应迅速增长。

（2）温度特征的提取

从天气网网站上提取了区域的年最高温、年最低温和年平均温，将其对应数据分别存入特征向量的第25～27列。夏季和冬季由于温度过高、过低的影响，用电量明显高于春、秋两个季节。日最大用电负荷和日最高用电量与气温有着明显的联系。

3 案例分析

3.1 数据集描述

本文以邓州市为例，从邓州市政府统计网和《邓州市电力专项规划》项目中提取2010—2020年表1中所需的经济数据，从天气网提取邓州市2010—2020年的年最高温、年最低温和年平均温数据，将以上数据存入特征向量的对应位置作为模型的数据集。基于机器学习预测3:1的原则，将2010—2015年划分为训练集，2016、2017年划分为测试集进行训练，2018—2020年的数据集去除了年最大负荷、年总用电量，为待预测年。

3.2 模型设置

GBDT算法的核心思想是沿着原来模型损失函数负梯度下降最快的方向来建立新模型，其基函数为CART树，提升树模型可以表示为CART树的加法模型，即

$$F_m(x_1\cdots x_{27},\ a_m)=F_{m-1}(x_1\cdots x_{27},\ a_{m-1})+\beta_m f_m(x_1\cdots x_{27},\ a_m) \tag{7}$$

式中，β_m是第m颗回归树的学习率，也是其预测结果的权重系数，调节范围在0～0.1之间；a_m是模型参数，如最大迭代次数、树的最大深度等，文中最大迭代次数取100次，树的最大深度为10。子采样Subsample取值为1.0，其他参数优先使用默认值或根据需要更改。

构建好数据集后，在Python环境下录入提取的数据，导入Scikit-learn、Numpy、Scipy和Matplotlib等科学计算包，并通过GBDT模型对离散的样本点进行拟合，最后输出2018—2020年的预测值，通过对比预测值与真实值，得出预测数据的精确性。

3.3 实验结果

将样本数据和特征向量放入模型中进行训练，利用训练后的模型对邓州市2018—2020年的年最大负荷、年总用电量进行预测，历年真实数据和三种预测方法的预测数据构成年最大负荷、年总用电量曲线，如图2、图3所示。

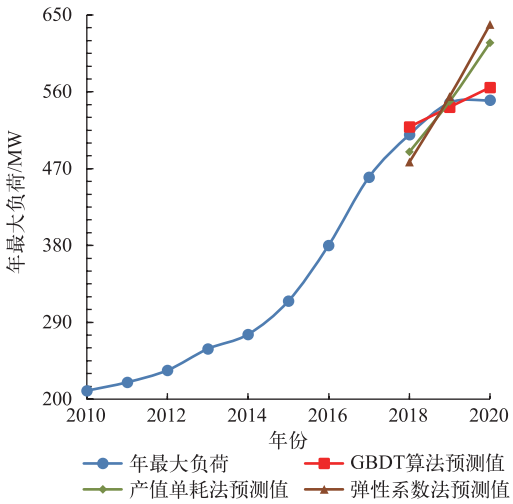


图2 邓州市年最大负荷曲线图

Fig. 2 The annual maximum load curve of Dengzhou City

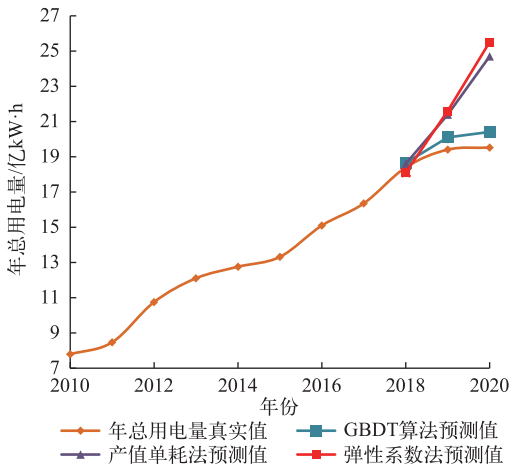


图3 邓州市年总用电量曲线图

Fig. 3 Curve of annual total electricity consumption in Dengzhou

2018—2020年邓州市年总用电量真实值分别为18.4亿kW·h、19.4亿kW·h和19.5亿kW·h，

以真实值为基准，可得到GBDT算法以及传统电力规划方法中产值单耗法、弹性系数法的预测结果相对真实值的相对误差，如表2、表3所示。

表 2 年最大负荷预测误差

Tab. 2 Yearly maximum load forecast error（%）

预测方法	年最大负荷相对预测误差		
	2018 年	2019 年	2020 年
GBDT 算法	1.76	−0.91	2.73
产值单耗法	−4.02	0.31	12.27
弹性系数法	−6.37	1.35	16.13

表 3 年总用电量预测误差

Tab. 3 Forecast error of total electricity consumption in the year

（%）

预测方法	年总用电量相对预测误差		
	2018 年	2019 年	2020 年
GBDT 算法	1.30	3.61	4.56
产值单耗法	1.09	10.31	26.60
弹性系数法	−1.63	11.34	30.71

由表2、表3可以看出，GBDT算法模型预测的年最大负荷相对误差都在3%以内，年总用电量相对误差都在5%以内，整体预测准确度较产值单耗法和弹性系数法高出许多。产值单耗法虽然在某一点的预测值更接近真实值，但最大预测误差较大，整体预测效果不佳。

4 结束语

本文采用了一种基于改进梯度提升决策树算法的负荷预测方法，综合考虑了地方经济发展和温度对负荷的影响，提高了中期负荷预测的准确率。以邓州市为例，验证本文所提方法的准确性，实验结果表明，与传统电力规划负荷预测方法相比，该方法的预测结果效果更好，具有更高的应用价值。EA

参考文献

[1] 康重庆，夏清，刘梅．电力系统负荷预测[M]．2版．北京：中国电力出版社，2016．

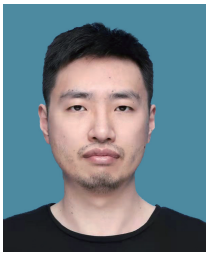
[2] 罗澍忻，麻敏华，蒋林，等．考虑多时间尺度数据的中长期负荷预测方法[J]．中国电机工程学报，2020，40(增刊1)：11-19．

[3] 田杨阳，张小斐，耿俊成，等． 基于改进聚类和 LSTM 的居民小区中长期负荷预测[J]．河南电力，2020(增刊2)，2020：58-63．

[4] FRIEDMAN J H．Stochastic gradient boosting[J]．Computational Statistics and Data Analysis，2002，38(4)：367-378．

[5] 徐永瑞，左丰恺，朱新山，等．改进GBDT算法的负荷预测研究[J]．电力系统及其自动化学报，2021，33(8)：94-101．

作者简介



谢中元

谢中元（1984—），男，工程师，研究方向为电网规划等。

(收稿日期：2021-06-15)