DOI: 10.19783/j.cnki.pspc.200234

基于梯度提升树计及非线性的电力负荷影响因素分析

庞传军[1,2]，余建明[1,2]，张 波[1,2]，刘 艳[1,2]

（1. 南瑞集团有限公司（国网电力科学研究院有限公司），江苏 南京 211106；
2. 北京科东电力控制系统有限责任公司，北京 100192）

摘要：分析负荷影响因素对电力负荷的影响对于电网调度人员了解负荷特性，提高负荷预测准确度具有重要的意义。针对传统相关性分析方法不能考虑复杂非线性影响的问题，采用先训练负荷预测模型，再分析相关性的思路，提出基于负荷预测模型的相关性分析方法，发现两者之间的非线性相关关系。首先，利用梯度提升树(Gradient Boosting Decision Tree, GBDT)的非线性建模和特征提取能力训练负荷预测模型。然后，基于预测模型提出采用重要性衡量影响因素对负荷的非线性影响，识别重要影响因素。最后，利用负荷对影响因素的偏依赖量计算各类影响因素变化对负荷变化趋势的非线性影响。采用实际的负荷数据进行验证，并与皮尔逊相关系数法进行对比。实验结果表明该方法能够有效识别影响负荷的重要因素，并能够发现各类因素和负荷之间的非线性关系。

关键词：负荷影响因素；电力负荷；非线性相关；梯度提升树；偏依赖量

Nonlinear correlation analysis of influence factors of a power load based on a gradient boosting decision tree

PANG Chuanjun[1,2], YU Jianming[1,2], ZHANG Bo[1,2], LIU Yan[1,2]

(1. NARI Group Corporation (State Grid Electric Power Research Institute), Nanjing 211106, China;
2. Beijing KeDong Electric Power Control System Co., Ltd., Beijing 100192, China)

Abstract: Analyzing the influence of factors on a power load is important for power grid dispatchers to understand load characteristics and improve the accuracy of load forecasting. The traditional correlation analysis method cannot consider complex nonlinear effects, so a new method, which can find the nonlinear correlation, is proposed to analyze the correlation of factors and power load. First, this paper uses the learning ability of the gradient boosting decision tree to train a power load forecasting model. Then, based on the forecasting model, it proposes the importance of measuring the nonlinear correlation between factors and power load. Finally, the partial dependence of power load on the factors is used to calculate the nonlinear effects of influencing factor changes on the trend of load changes. The real power load is used for verification and compared using the Pearson correlation coefficient. The experimental results show that the method can effectively identify the important factors which affect power load, and can find the nonlinear relationship between various factors and the power load.

This work is supported by Science and Technology Project of State Grid Corporation of China (No. 5700-202055368A-0-0-00).

Key words: load influencing factors; power load; nonlinear correlation; gradient boosting tree; partial dependence

0 引言

电力系统负荷受经济因素、天气、季节、节假日等多种因素的影响。并且随着风电、光伏等分布式清洁能源大规模接入电网[1-2]、电力市场的开展[3]和电动汽车的广泛应用[4-6]，影响电力系统负荷的因素愈加复杂。目前，针对电力系统负荷的研究较多的集中在负荷预测领域，不断通过改进预测算法提升电力负荷预测的准确程度[7-9]。然而预测模型训练之前的负荷特性分析和负荷影响因素分析也是影响负荷预测准确度的重要瓶颈之一[10]。不同影响因素

基金项目：国家电网公司科技项目资助(5700-202055368A-0-0-00)

引起的日负荷曲线形状的改变也截然不同。分析影响因素对电力负荷的影响，根据不同因素的影响有针对性地采取合适的预测方法，对于提高电力负荷预测准确度，确保电网安全、经济、高效运行有着重要的意义[11]。还可以评估需求响应潜力、提升“源-网-荷”互动水平和清洁能源消纳能力。在电力市场运营过程中也可以让电力市场参与者了解影响负荷的主要因素，指导其市场行为。

目前，电力负荷影响因素分析的研究成果与负荷预测相比相对较少。一般都是将影响因素分析作为训练负荷预测模型之前的特征选择的过程。主要有两种方法：基于灰色关联度的方法和皮尔逊相关系数法。文献[12]提出一种基于改进灰色关联度的电力负荷影响因素量化分析模型，构建了宏观与微观相结合、层次化的负荷影响因素指标体系，系统地反映新形势下经济、政策等对负荷的影响。文献[13]引入灰色关联度理论对影响郑州电网负荷特性变化的因素进行量化分析，找出引起负荷特性变化的主要因素。文献[14-17]采用皮尔森相关系数计算气温等因素和负荷之间的相关性性，然后再选取相关性强的因素训练负荷预测模型。文献[18]采用皮尔逊相关系数分析方法分析经济相关因素对负荷的影响。文献[19]在构建饱和负荷预测模型之前采用主成分分析和小波分析法，通过主成分及其基频分量的提取，从影响电力负荷的众多因素中筛选出主要因素。文献[20]构建了电力负荷回归方程，用变量相关因子来研究非气象因素对夏季降温负荷的影响；文献[21]运用计量经济学中的变量相关性检验来获得电力需求量与其影响因素之间的相关性特征。

传统的相关性分析方法只能衡量影响因素与负荷之间的线性关系，但是由于影响因素与用电负荷之间的关系具有高度复杂性、高度非线性等特点。此类方法不能真实反映影响因素对负荷的非线性影响。另外，传统方法只能通过一个数值反映因素对负荷的影响程度，不能反映影响因素的变化趋势对负荷的影响。机器学习技术具有强大的非线性建模能力和特征提取能力，能够学习各类影响因素与负荷之间的非线性关系。目前，较多的机器学习技术应用在电力负荷预测领域。其中，梯度提升树(Gradient Boosting Decision Tree, GBDT)算法是一种有监督集成学习方法。通过由决策树构成的弱预测模型不断迭代，以最小化前一轮的预测误差为目标训练强预测模型。该方法具备良好的非线性特征提取能力和泛化能力，已经被应用于电力负荷预测[22]、风电及光伏预测[23]、故障检测[24-25]、电压稳定性评估[26]等领域。因此，本文采用先训练基于GBDT的负荷预测模型，再分析相关性的思路，利用GBDT预测模型学习到的非线性关系，识别影响负荷的重要因素。

首先，提出计及非线性的各类因素对负荷影响大小的评估方法。负荷预测模型训练完成后，通过随机重新排列训练数据集的某一影响因素形成新的数据集。利用预测模型计算随机排列之后的预测损失衡量每个影响因素的重要性。其次，提出负荷对某一影响因素的偏依赖函数，利用偏依赖函数衡量该影响因素变化对负荷变化趋势的影响。采用实际历史负荷数据训练基于GBDT的负荷预测模型。然后，计算负荷影响因素重要性和负荷对影响因素的偏依赖量。实验结果表明，该方法可以分析各类影响因素对电力负荷的非线性影响。

1 基于GBDT负荷预测模型

1.1 基于梯度提升算法的负荷预测

电力系统负荷预测是典型的有监督学习问题。是基于历史负荷数据 y 和历史影响因素数据 x，建立预测模型 $\hat{F}(x)$ 的过程。模型的建立过程是采用历史数据训练，最小化损失 $L\left(y,\hat{F}(x)\right)$ 的期望值的过程[27]。如式(1)所示。

$$\hat{F}=\arg\min_{F}E_{x,y}\left[L\left(y,\hat{F}(x)\right)\right] \tag{1}$$

式中：$\hat{F}$ 为训练完成的负荷预测模型；$L\left(y,\hat{F}(x)\right)$ 为训练过程中采用的损失函数；x 为负荷历史影响因素；y 为历史负荷。

梯度提升算法是一种集成学习算法[28]。其核心思想是对于历史负荷数据组成的训练数据集，训练若干个弱预测模型，通过一定的结合策略形成一个强预测模型。即 $\hat{F}(x)$ 是由若干个子模型 $h_i(x)\in \mathrm{H}$ 以加权和的形式构成的。如式(2)所示。

$$\hat{F}(x)=\sum_{i=1}^{m}\gamma_i h_i(x) \tag{2}$$

式中：$h_i(x)$ 为弱预测模型；γ_i 为每个弱预测模型的权重；m 为弱预测模型的数量。

根据机器学习中的经验风险最小化原理。$\hat{F}(x)$ 可以最小化在训练集上的平均损失求得。一般采用迭代算法不断优化模型。

开始时，模型由一个常数函数组成，然后采用贪心算法不断迭代，生成最后的模型，如式(3)和式(4)。

$$F_0 = \arg\min_{\gamma}\sum_{i=1}^{n} L(y_i,\gamma) \tag{3}$$

$$F_m(x) = F_{m-1}(x) + \arg\min_{h\in \mathrm{H}}\sum_{i=1}^{n} L\left(y_i, F_{m-1}(x_i) + h(x_i)\right) \tag{4}$$

式中：F_0 为初始预测模型；n 为训练样本数量；L 为损失函数；F_m 为第 m 次迭代训练后的预测模型。

电力负荷是连续的变量，每一个弱预测模型都是连续的可微函数，即 H(与公式中不一致？答：与公式中一致)是一个定义在 R 上的可微函数的集合。在迭代过程中可以采用梯度下降法进行模型的更新。采用损失函数的负梯度寻找最优解，如式(5)和式(6)。

$$F_m(x) = F_{m-1}(x) - \gamma_m \sum_{i=1}^{n} \Delta F_{m-1} L\left(y_i, F_{m-1}(x_i)\right) \tag{5}$$

$$\gamma_m = \arg\min_{\gamma}\sum_{i=1}^{n} L\left(\begin{array}{l} y_i, F_{m-1}(x_i) - \\ \gamma \dfrac{\partial L\left(y_i, F_{m-1}(x_i)\right)}{\partial F_{m-1}(x_i)} \end{array}\right) \tag{6}$$

式中：F_m 为第 m 次迭代训练后的预测模型；ΔF_{m-1} 为第 $m-1$ 次迭代后损失函数的梯度；γ_m 为第 m 次迭代后弱预测模型的权重；n 为训练样本数量。

1.2 基于梯度提升决策树的负荷预测

梯度提升算法中的弱预测模型通常由决策树构成[29]。每个决策树采用决策树算法进行拟合。算法输入包括：由影响因素历史数据 x 和历史负荷 y 组成的训练集 $D=\{x_i,y_i\}_{i=1}^{n}$，损失函数 $L(y,F(x))$，迭代次数 M，学习率 η。输出为训练完成的负荷预测模型 $F_M(x)$。GBDT 训练过程如下：

1) 根据式(3)初始化模型。

2) 对于 $m=1,\cdots,M$，执行第 3)至第 6)。

3) 对于每个样本 $i=1,\cdots,n$，计算梯度。

$$r_{im} = -\left[\frac{\partial L\left(y_i, F_{m-1}(x_i)\right)}{\partial F_{m-1}(x_i)}\right]_{F(x)=F_{m-1}(x)} \tag{7}$$

4) 利用 $\{x_i,r_{im}\}, i=1,\cdots,n$，拟合一棵回归决策树，得到第 m 棵决策树 γ_m。其对应的叶子节点区域为 $R_{mj}, j=1,\cdots,J$，其中 J 为第 m 棵回归树叶子节点数目。

5) 对应每一个叶子节点区域 $j=1,\cdots,J$，根据式(8)计算最佳拟合。

$$\gamma_{mj} = \arg\min_{\gamma}\sum_{x_i\in R_{mj}} L\left(y_i, F_{m-1}(x_i)+\gamma\right) \tag{8}$$

6) 根据式(9)更新模型。

$$F_m(x) = F_{m-1}(x) + \eta\sum_{j=1}^{J}\gamma_{mj}, I\left(x\in R_{mj}\right) \tag{9}$$

式中：I 为指示函数，当 x 属于叶子区域 R_{mj} 时为 1，否则为 0；η 为学习率。

7) 输出最终模型。

$$F_M(x) = F_0(x) + \sum_{m=1}^{M}\sum_{j=1}^{J}\gamma_{mj}, I\left(x\in R_{mj}\right) \tag{10}$$

2 影响因素与负荷相关性评估方法

2.1 负荷影响因素重要性评估

基于历史数据 D，完成 GBDT 的训练后得到训练模型 $F_M(x)$，随机改变历史负荷影响因素数中某一因素的数据排列，保持其余影响因素不变。计算损失 $L\left(y,F_M(x)\right)$ 的变化量，采用损失变化量衡量该因素对负荷的影响重要程度。某一因素 f_j 对负荷影响重要程度定义如式(11)。

$$f_j = \frac{1}{K}\sum_{k=1}^{K} s_{k,j} - s \tag{11}$$

式中：K 为随机打乱影响因素 f_j 的次数；s 为打乱之前模型的初始损失；$s_{k,j}$ 为第 k 次随机打乱影响因素后模型的损失。计算重要性的输入包括：完成训练的模型 $F_M(x)$，训练集 $D=\{x_i,y_i\}_{i=1}^{n}$，随机打乱的负荷影响因素的次数 K。输出为每个影响因素的重要性。计算过程如下：

1) 根据损失函数 $L\left(y,F_M(x)\right)$ 计算初始损失 s。

2) 对于每个影响因素 $1,\cdots,J$，执行第 3)至第 5)。

3) 随机打乱属性 j，其他属性保持不变。生成新的训练集 $\tilde{D}_{k,j}$。

4) 采用新的训练集 $\tilde{D}_{k,j}$ 计算损失 $s_{k,j}$。

5) 根据式(11)计算影响因素 j 的重要性。

6) 输出所有影响因素 $1,\cdots,J$ 重要性。

2.2 负荷对影响因素的偏依赖量

负荷影响因素重要性计算只是反映了因素对负荷影响的相对重要性，并不能量化分析负荷影响因素的变化趋势对负荷的影响。基于预测模型，将关注的影响因素在该因素的取值范围内取不同的值，其他因素不变，这时模型在所有训练集上的平均预测负荷就是所关注负荷影响因素的函数，该函数称为模型预测负荷对影响因素的偏依赖函数。如式(12)所示。

$$\hat{f}_{X_S(X_S)} = \frac{1}{n}\sum_{i=1}^{n} F_M\left(X_S, X_C^i\right) \tag{12}$$

式中：n 为样本数量；X_S 为关注的负荷影响因素；X_C 剩余不关注的影响因素；X_S 和 X_C 共同组成模型输入 X 。偏依赖函数定义了所关注的影响因素对模型预测结果的边际影响。

2.3 负荷影响评估流程

基于 GBDT 负荷预测模型、负荷影响因素重要性和影响因素偏依赖，对电力负荷影响因素进行分析的流程如图 1 所示。

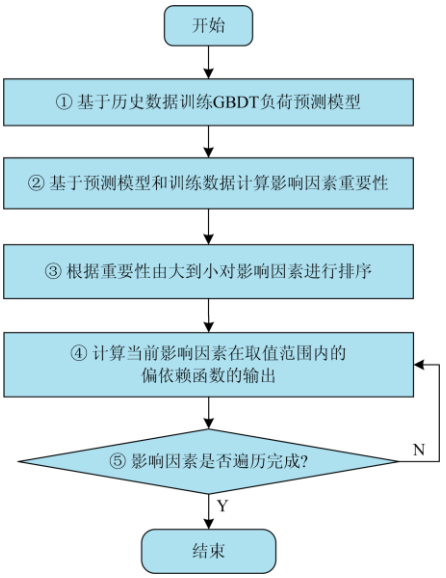


图 1 负荷影响因素分析流程

Fig. 1 Analysis process of correlation between power load and influencing factors

3 算例分析

3.1 实验数据

采用本文方法对某省其中一个地区的负荷影响因素进行分析。以该地区 2002 年至 2018 年每日每小时历史负荷及温度数据进行验证。选取的负荷影响因素如表 1 所示。

表 1 实验所选取的负荷影响因素

Table 1 Power load influencing factors in this experiment

标识	负荷影响因素	说明
Temperature	气温	摄氏度
Year	年	对应年份
Month	月份	对应月份(1~12)
day_of_month	每月第几天	每月第几天
day_of_week	周几	当前周几(0~6)
hour	时刻	当前时刻(0~23)
is_business	是否工作日	是否工作日(0 否；1 是)
day_of_year	全年第几天	当年第几天(1~365)

其中2018年5月份负荷与其影响因素的相关关系如图 2 所示。

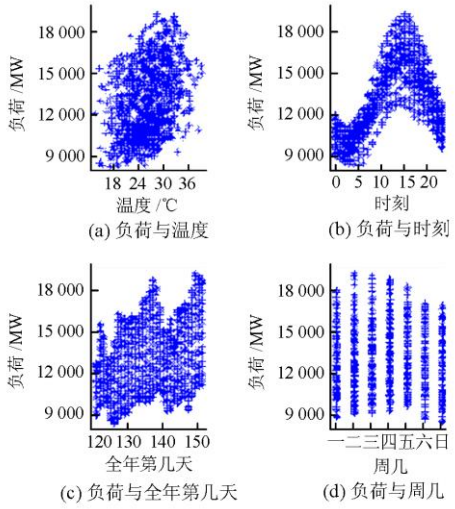


图 2 2018 年 5 月负荷与影响因素之间的相关关系

Fig. 2 Correlation between power load and influencing factors in May 2018

由图 2 可见，负荷与其影响因素之间并没有明显的线性相关关系。因此，采用皮尔逊相关系数无法正确衡量负荷与影响因素之间的相关关系。

3.2 负荷预测模型训练

基于历史数据训练 GBDT 负荷预测模型，需要设置损失函数、学习率、决策树最大深度和决策数个数等重要参数。损失函数采用均方误差，如式(13)。

$$L\left(y,\hat{F}\left(x\right)\right)=\frac{1}{n}\sum_{i=1}^{n}\left(y-\hat{F}\left(x\right)\right)^{2} \tag{13}$$

式中：n 为样本数量；y 为实际负荷；$\hat{F}\left(x\right)$ 为训练完成的负荷预测模型。

参数设置采用网格搜索(Grid Search)方法[30-31]寻找最优参数组合，如表 2 所示。

表 2 预测模型训练所需参数的最优设置

Table 2 Optimal setting of parameters required for prediction model training

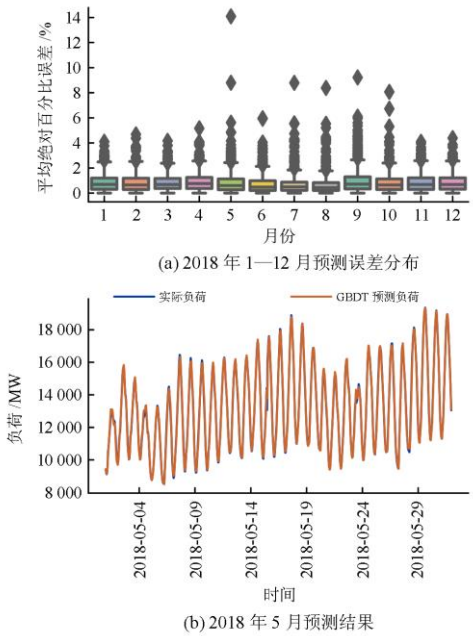
参数名称	参数说明	参数最优设置
学习率 η	是对每一棵新加上去的树设置一个收缩权重，用于防止过拟合	0.1
决策树最大深度	用于表征单棵树的复杂程度	4
决策树个数 M	对应于 1.2 节训练流程中的参数 M，每迭代一次增加一棵决策树	100

训练完成后采用平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)衡量预测性能，MAPE 定义如式(14)。

$$e_{\mathrm{MAPE}}=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i-y_i}{y_i}\right|\times 100\% \tag{14}$$

式中：n为训练样本数量；y_i为实际负荷；$\hat{y}_i$为预测负荷。

将 2002 年—2017 年历史负荷作为训练数据对模型进行训练。2018 年历史负荷作为测试数据测试预测模型性能，预测结果如图 3 所示。



(a) 2018 年 1—12 月预测误差分布

(b) 2018 年 5 月预测结果

图 3　2018 年 5 月负荷预测结果

Fig. 3 Power load forecast results in May 2018

图 3(a)为 2018 年每个月预测误差分布；图 3(b)为 2018 年 5 月份预测结果。由图 3 可知，模型每个月的 MAPE 大部分在 1%~2.5%之间，取得了较好的预测效果。

3.3　影响因素重要性

首先采用皮尔逊相关系数法[32]计算负荷与各影响因素之间的相关关系；然后，基于训练完成的 GBDT 负荷预测模型，利用 2.1 节方法计算影响因素对负荷的重要性。分别从大到小排序后结果如图 4 所示。

由图 4 可知，采用重要性衡量该地区负荷因素对负荷的影响，并按照大小进行排序，其结果与皮尔逊相关系数法顺序有所不同。当采用重要性计算对负荷影响时，影响最大的是时刻；当采用皮尔逊相关系数计算时，影响最大的是气温。因为采用皮尔逊相关系数只是体现了因素与负荷的线性关系，并没有体现两者的非线性关系。时间对负荷的影响在不同时间段相关性并不同，有的时间段正相关，有的时间段负相关，采用相关系数就会导致正负相互抵消，不能正确反映时间对负荷的影响程度。采用影响因素重要性可以考虑影响因素对负荷的非线性影响，正确反映影响因素对负荷影响的大小。

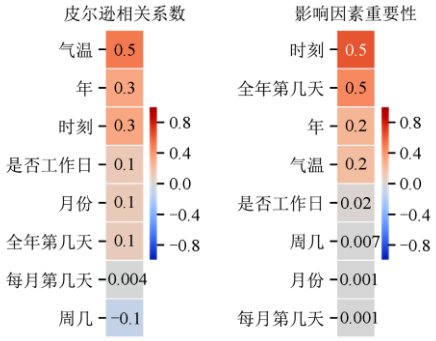


图 4　采用相关系数和重要性分别计算各类影响因素对负荷的影响

Fig. 4 Correlation coefficient and importance are used to calculate the influence of various influencing factors on the power load

3.4　影响因素对负荷的影响

基于负荷预测模型，按照从大到小的顺序逐一计算负荷对每个影响因素的偏依赖量，结果如图 5 所示。由于月份和每月第几天两个影响因素对负荷的重要性较小，忽略了负荷对这两个影响因素偏依赖量的计算结果。由图 5(a)可知，负荷与时刻在 00:00—13:00 是非线性的正相关关系，在 13:00—23:00 是非线性的负相关关系。如果采用皮尔逊相关系数，正负相关将会抵消。因此，采用皮尔逊相

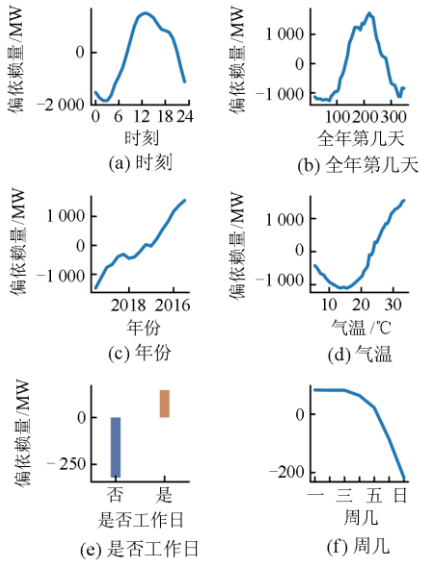


图 5　负荷对影响因素的偏依赖量计算结果

Fig. 5 Result of partial dependence of load on influencing factors

关系数进行相关性分析时，时刻与负荷相关性并没有排在第一位。进一步说明采用影响因素重要性能够反映影响因素与负荷的非线性相关关系。其他影响因素分析如表 3 所示。

表 3 图 5 中各因素对电力负荷影响分析

Table 3 Analysis of influence of various factors in Fig. 5 on power load

图序号	负荷偏依赖量分析	原因分析
(a)	负荷与时刻在 00:00—13:00 是非线性正相关关系，在 13:00—23:00 是非线性负相关关系	与人们的作息规律有关
(b)	第 220 天之前，负荷逐渐上升，之后负荷是逐渐下降	季节导致气温变化从而影响负荷
(c)	年份对负荷的影响一直是非线性正相关	经济发展水平逐年提高促进用电负荷逐年升高
(d)	气温在 15 ℃，负荷随气温升高逐渐降低；之后随气温升高逐渐升高	气温 15 ℃之后空调负荷逐渐增大
(e)	工作日对负荷的影响正相关，非工作日负相关	非工作日工业负荷或办公园区减少
(f)	周一至周五，负荷变化不大；周六、周日对负荷负相关	周六、周日工业负荷或办公园区负荷减少

采用偏依赖量可以同时衡量两个因素对负荷的影响，分别计算负荷对(时刻，温度)、(全年第几天，温度)的偏依赖量，计算结果如图 6 和图 7 所示。

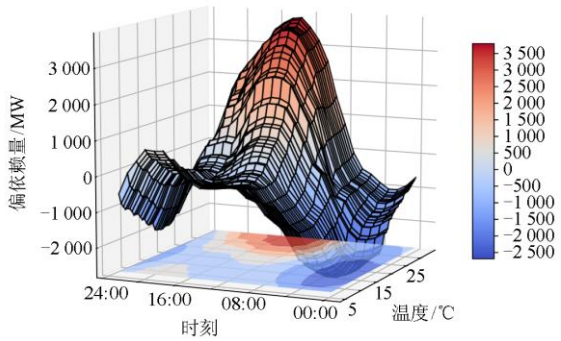


图 6 时刻和温度对负荷的影响

Fig. 6 Time and temperature effects on power load

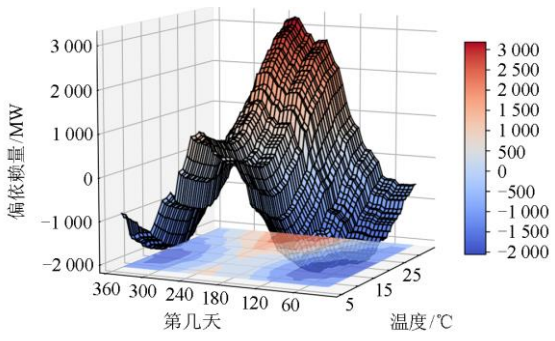


图 7 全年第几天和温度对负荷的影响

Fig. 7 Day-of-year and temperature effects on power load

由图 6 可知，在白天工作时间，负荷对气温较敏感。在 09:00—13:00，如果气温升高会导致用电负荷的大幅度升高。

由图 7 可知，在夏天，负荷对气温较敏感。即每年的第 170 天—第 200 天，气温升高会导致负荷大幅度升高。

4 结论

针对现有相关性分析方法的不足，在训练基于梯度提升树的负荷预测模型的基础上，提出利用影响因素重要性和偏依赖量衡量影响因素对负荷的非线性影响。采用实际的电力负荷数据进行验证，并详细分析了时刻、气温、当前日期、是否工作日等因素对负荷的非线性影响。实验结果表明：① 采用影响因素重要性衡量各因素对负荷影响的大小，可以考虑影响因素与电力负荷的非线性关系，识别影响负荷的重要因素；② 采用偏依赖量可以衡量影响因素变化对负荷趋势的非线性影响影响。

采用本文方法衡量各种因素对负荷的影响，对于电网调度运行人员，可以根据不同负荷采取不同的预测方法提升预测准确度；对于电力市场人员，可以了解不同因素对负荷的影响，指导其参与电力市场的行为。

风电、光伏等清洁能源出力与电力负荷相似，都受多种因素的复杂影响，具备随机性和波动性。采用该方法识别影响清洁能源出力的重要因素，发现各类影响因素对清洁能源出力的影响也是后续研究方向。

参考文献

[1] 韩轩, 邱晓燕, 沙熠, 等. 基于变置信度机会约束规划的风电并网优化调度[J]. 高压电器, 2018, 54(10): 179-186.
HAN Xuan, QIU Xiaoyan, SHA Yi, et al. Optimal dispatching under grid connection of wind farm based on variable confidence chance constrain model[J]. High Voltage Apparatus, 2018, 54(10): 179-186.

[2] 赖昌伟, 黎静华, 陈博, 等. 光伏发电出力预测技术研究综述[J]. 电工技术学报, 2019, 34(6): 1201-1217.
LAI Changwei, LI Jinghua, CHEN Bo, et al. Review of photovoltaic power output prediction technology[J]. Transactions of China Electrotechnical Society, 2019, 34(6): 1201-1217.

[3] 丁业豪, 麦琪. 电力市场用电量需求分析预测模型研究[J]. 电测与仪表, 2017, 54(14): 14-23.
DING Yehao, MAI Qi. Research on power market electricity demand analysis and forecasting model[J].

Electrical Measurement & Instrumentation, 2017, 54(14): 14-23.

[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] 苏小林, 张艳娟, 武中, 等. 规模化电动汽车充电负荷的预测及其对电网的影响[J]. 现代电力, 2018, 35(1): 45-54.
SU Xiaolin, ZHANG Yanjuan, WU Zhong, et al. Forecasting the charging load of large-scale electric vehicle and its impact on the power grid[J]. Modern Electric Power, 2018, 35(1): 45-54.

[6] 李振兴, 孟晓星, 田斌, 等. 电动汽车接入充放电对配电网继电保护的影响[J]. 高压电器, 2018, 54(4): 188-194.
LI Zhenxing, MENG Xiaoxing, TIAN Bin, et al. Influence of electric vehicle charging and discharging on distribution network relay protection[J]. High Voltage Apparatus, 2018, 54(4): 188-194.

[7] 吴润泽, 包正睿, 王文韬, 等. Hadoop 架构下基于模式匹配的短期电力负荷预测方法[J]. 电工技术学报, 2018, 33(7): 1542-1551.
WU Runze, BAO Zhengrui, WANG Wentao, et al. Short-term power load forecasting method based on pattern matching in Hadoop framework[J]. Transactions of China Electrotechnical Society, 2018, 33(7): 1542-1551.

[8] 刘建华, 李锦程, 杨龙月, 等. 基于 EMD-SLSTM 的家庭短期负荷预测[J]. 电力系统保护与控制, 2019, 47(6): 40-47.
LIU Jianhua, LI Jincheng, YANG Longyue, et al. Short-term household load forecasting based on EMD-SLSTM[J]. Power System Protection and Control, 2019, 47(6): 40-47.

[9] HUANG Li, YANG Yongbiao, GAO Honglei, 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.

[10] MUSTAPHA M, MUSTAFA M W, KHALID S N, et al. Classification of electricity load forecasting based on the factors influencing the load consumption and methods used: an-overview[C] // 2015 IEEE Conference on Energy Conversion (CENCON), October 19-20, 2015, Johor Bahru, Malaysia: 442-447.

[11] 廖峰, 徐聪颖, 姚建刚, 等. 常德地区负荷特性及其影响因素分析[J]. 电网技术, 2012, 36(7): 117-125.
LIAO Feng, XU Congying, YAO Jiangang, et al. Load characteristics and its influencing factors of Chang-de Area[J]. Power System Technology, 2012, 36(7): 117-125.

[12] 王雁凌, 吴梦凯, 周子青, 等. 基于改进灰色关联度的电力负荷影响因素量化分析模型[J]. 电网技术, 2017, 41(6): 1772-1778.
WANG Yanling, WU Mengkai, ZHOU Ziqing, et al. Quantitative analysis model of power load influencing factors based on improved grey relational degree[J]. Power System Technology, 2017, 41(6): 1772-1778.

[13] 康田园, 尹淑萍, 王现法, 等. 大型城市电网负荷特性及其影响因素分析[J]. 电测与仪表, 2016, 53(6): 51-56.
KANG Tianyuan, YIN Shuping, WANG Xianfa, et al. Load characteristics of large city power grid and analysis on its influencing factors[J]. Electrical Measurement & Instrumentation, 2016, 53(6): 51-56.

[14] 马瑞, 周谢, 彭舟, 等. 考虑气温因素的负荷特性统计指标关联特征数据挖掘[J]. 中国电机工程学报, 2015, 35(1): 43-51.
MA Rui, ZHOU Xie, PENG Zhou, et al. Data mining on correlation feature of load characteristics statistical indexes considering temperature[J]. Proceedings of the CSEE, 2015, 35(1): 43-51.

[15] 程志友, 余国晓, 丁柏宏. 采用改进温湿度变量策略的夏季短期负荷预测方法[J]. 电力系统保护与控制, 2020, 48(1): 48-54.
CHENG Zhiyou, YU Guoxiao, DING Baihong. Summer short-term load forecasting method based on improved temperature and humidity variable strategy[J]. Power System Protection and Control, 2020, 48(1): 48-54.

[16] 张志, 杜延菱, 崔慧军, 等. 考虑关联因素的智能化中长期电力负荷预测方法[J]. 电力系统保护与控制, 2019, 47(2): 24-30.
ZHANG Zhi, DU Yanling, CUI Huijun, et al. Intelligent mid-long electricity load forecast method considering associated factors[J]. Power System Protection and Control, 2019, 47(2): 24-30.

[17] 席雅雯, 吴俊勇, 石琛, 等. 融合历史数据和实时影响因素的精细化负荷预测[J]. 电力系统保护与控制, 2019, 47(1): 80-87.
XI Yawen, WU Junyong, SHI Chen, et al. A refined load forecasting based on historical data and real-time influencing factors[J]. Power System Protection and Control, 2019, 47(1): 80-87.

[18] CHEN Yitui. The factors affecting electricity consumption and the consumption characteristics in the residential

sector-a case example of Taiwan[J]. Sustainability, 2017, 9(8): 1-16.

[19] 李亦言, 韩冬, 严正, 等. 复杂城市化因素下的饱和负荷预测模型[J]. 电网技术, 2016, 40(9): 2824-2831.
LI Yiyan, HAN Dong, YAN Zheng, et al. Saturated load forecasting model under complex urbanization characteristics[J]. Power System Technology, 2016, 40(9): 2824-2831.

[20] 刘继东, 韩学山, 楚成博, 等. 考虑非气象因素的电网夏季降温负荷研究[J]. 电力自动化设备, 2013, 33(2): 40-46.
LIU Jidong, HAN Xueshan, CHU Chengbo, et al. Cooling load of summer grid considering non-meteorological factors[J]. Electric Power Automation Equipment, 2013, 33(2): 40-46.

[21] 谭忠富, 张金良, 吴良器, 等. 中长期负荷预测的计量经济学与系统动力学组合模型[J]. 电网技术, 2011, 35(1): 186-190.
TAN Zhongfu, ZHANG Jinliang, WU Liangqi, et al. A model integrating econometric approach with system dynamics for long-term load forecasting[J]. Power System Technology, 2011, 35(1): 186-190.

[22] MAYRINK V, HIPPERT H S. A hybrid method using exponential smoothing and gradient boosting for electrical short-term load forecasting[C] // 2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI), November 2-4, 2016, Cartagena, Colombia: 1-6.

[23] TORRES-BARR A, ALONSO A, DORRONSORO J R. Regression tree ensembles for wind energy and solar radiation prediction[J]. Neuro Computing, 2019, 326-327: 151-160.

[24] 黄新波, 王享, 田毅, 等. 于 PSO-ELM 融合动态加权 AdaBoost 的变压器故障诊断方法[J]. 高压电器, 2020, 56(5): 39-46.
HUANG Xinbo, WANG Xiang, TIAN Yi, et al. Transformer fault diagnosis algorithm based on PSO-ELM fusion dynamically weighted AdaBoost[J]. High Voltage Apparatus, 2020, 56(5): 39-46.

[25] 贾科, 宣振文, 林瑶琦, 等. 基于 Adaboost 算法的并网光伏发电系统的孤岛检测法[J]. 电工技术学报, 2018, 33(5): 1106-1113.
JIA Ke, XUAN Zhenwen, LIN Yaoqi, et al. An islanding detection method for grid-connected photovoltaic power system based on Adaboost algorithm[J]. Transactions of China Electrotechnical Society, 2018, 33(5): 1106-1113.

[26] 肖繁, 王涛, 饶渝泽, 等. 基于梯度提升决策树静态电压稳定裕度评估[J]. 电测与仪表, 2020, 57(20): 39-45.
XIAO Fan, WANG Tao, RAO Yuze, et al. Static voltage stability margin prediction based on gradient boosting regression tree[J]. Electrical Measurement & Instrumentation, 2020, 57(20): 39-45.

[27] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012.

[28] FRIEDMAN J H. Greedy function approximation: a gradient boosting machine[J]. The Annals of Statistics, 2001, 29(5): 1189-1232.

[29] FRIEDMAN J H, MEULMAN J J. Multiple additive regression trees with application in epidemiology[J]. Statistics in Medicine, 2003, 22(9): 1365-1381.

[30] CHICCO D. Ten quick tips for machine learning in computational biology[J]. Biodata Mining, 2017, 10(1): 1-35.

[31] BERGSTRA J, BENGIO Y. Random search for hyper-parameter optimization[J]. Journal of Machine Learning Research, 2012, 13: 281-305.

[32] RODGERS J L, NICEWANDER W A. Thirteen ways to look at the correlation coefficient[J]. The American Statistician, 1988, 42(1): 59-66.

收稿日期：2020-03-08

作者简介：

庞传军(1984—)，男，通信作者，硕士，高级工程师，研究方向为电力系统及自动化，人工智能技术在电力系统中的应用；E-mail: pangchuanjun@sgepri.sgcc.com.cn

余建明(1979—)，男，博士，高级工程师，主要研究方向为电力系统自动化、人工智能技术在电力系统中的应用；E-mail: yujianming@sgepri.sgcc.com.cn

张　波(1978—)，男，硕士，高级工程师，主要研究方向为电力系统自动化。E-mail: zhangbo7@sgepri.sgcc.com.cn

(编辑　姜新丽)