

基于天气分类的BIPV系统发电量预测方法

戴 姝¹, 赖洪亮¹, 李 远²

(1. 佛山电力设计院有限公司, 广东 佛山 528200;

2. 华南理工大学电力学院, 广东 广州 510000)

摘要: 建筑集成光伏(Building Integrated Photovoltaic, BIPV)系统发电量受天气条件影响呈现出典型的波动性和间歇性, 并网时会影响电网系统的安全稳定运行。提出一种基于天气分类和改进K-均值聚类的BIPV系统发电量预测方法。首先提取温度因子、湿度因子和晴朗因子三维特征构成特征向量, 对33种标准天气进行量化表征, 然后提出一种局部密度改进的K-均值聚类算法进行自适应聚类, 将天气类型划分为3个子类, 最后对每个子类分别建立BP神经网络模型进行发电量预测。采用某市工业园区屋顶BIPV系统实际数据开展实验, 结果表明该方法能够获得较高的预测精度, 具有一定的推广应用价值。

关键词: BP神经网络; 建筑集成光伏; 预测模型; 自适应聚类; 特征提取

中图分类号: TP183 文献标识码: A 文章编号: 1003-7241(2025)03-0038-05

Forecasting Method of Power Generation in BIPV System Based on Weather Classification

DAI Shu¹, LAI Hong-liang¹, LI Yuan²

(1. Foshan Electric Power Design Institute Co., Ltd., Foshan 528200 China;

2. School of Electric Power, South China University of Technology, Guangzhou 510000 China)

Abstract: The power generation of building integrated photovoltaic (BIPV) system shows typical volatility and intermittence due to weather conditions, which affects the safe and stable operation of the grid system when connected to the grid. A forecasting method of BIPV system power generation based on weather classification and improved K-means clustering is proposed. First, the three-dimensional features of temperature factor, humidity factor and sunny factor are extracted to form a feature vector to quantitatively characterize 33 standard weather conditions. Then, a local density improved K-means clustering algorithm is proposed for adaptive clustering, and weather types are divided into three subcategories. Finally, the BP neural network model is established for each subcategory to forecast the generation capacity. The actual data of BIPV system on the roof of an industrial park in a city are used to carry out the experiment. The results show that the method can achieve high prediction accuracy and has certain popularization and application value.

Keywords: BP neural network; building integrated photovoltaic; forecast model; adaptive clustering; feature extraction

0 引言

一直以来,光伏发电系统面临着光伏组件成本高、光电转换效率低、相对于传统化石能源发电系统投资成本大等一系列问题,严重限制了光伏发电系统的大规模普及应用^[1-2]。建筑集成光伏(Building Integrated Photovoltaic, BIPV)以降低光伏发电系统的投资建设成本为目标,将建筑材料中的新技术新进展引入光伏发电系统,促使产生的电能就地利用,从而大大减少传输点损耗,提升经济效益,已逐渐成为城市光伏大规模应用的重要发展方向^[3-5]。我国“十四五”住房和城乡建设科技发展规划已经明确将BIPV作为推动建筑从能耗型,节能型向产能型

转变,打造绿色建筑解决方案。

然而,光伏发电系统的发电量与太阳辐照强度正相关,同时受天气状况影响,呈现出典型的间歇性和波动性特点。研究表明,当光伏发电量达到整个电力系统发电量的10%~15%时,其间歇性和波动性会影响电网的安全稳定运行,因此对BIPV系统发电量进行高精度预测,能够帮助电力调度部门制定合理的发电计划,从而提升经济效益,保证电网系统的稳定运行^[6-9]。

目前国内外学者主要从三个方向对BIPV发电量预测进行研究:利用RETScreen, PVsyst等光伏设计软件的模拟预测法;利用太阳辐照数据,光伏地理位置和太阳能-电能转换效率的物理预测法;基于历史发电数据和气象数据的统计预测法。其中模拟预测法通常假设光伏板

*基金项目: 广东省自然科学基金(2021A1515012054)

收稿日期: 2023-09-28

为平行阵列分布,在这种情况下能够获得较高的预测精度,但是对于BIPV系统,光伏组件与建筑物的结合方式复杂多变,此时模拟预测法获得的分析结果较为片面,不满足预测精度要求^[10];物理预测法的核心在于将太阳能到电能转换过程中涉及的诸如太阳辐照强度,气候影响和太阳能-电能转换效率等各个环节建立精确的数学模型。如果能够获得上述过程的精确数学模型,则物理预测法能够获得非常高的预测精度,然而实际生产生活中,对于不同BIPV系统而言上述精确的数学模型难以构建,通常只能采用通用或经验模型,此时物理预测法的预测精度会出现较大偏差,限制了该类方法在实际生产生活中的应用^[11];统计预测法是近年来随着机器学习技术的发展而兴起的一种智能化预测方法,通过机器学习理论对历史发电数据和气象数据进行分析并建立卷积神经网络、最小二乘支持向量机等模型,将气候信息作为模型输入从而获得发电量预测结果,由于统计预测法不需要额外的太阳辐射测量装置且具有较强的非平稳、非线性函数拟合能力,一经提出便引起了广泛关注,已经成为当前的主流方法^[12-15]。

由于BIPV系统发电量与太阳辐照度、温度和湿度等气候条件强相关,本文提出一种基于天气分类的BIPV系统发电量预测方法。首先提取温度因子、湿度因子和晴朗因子三维特征构成特征向量并对国标中规定的33种天气类型进行量化表征,然后提出一种改进的K-均值聚类算法对特征向量进行自适应聚类,最后对每个聚类分别建立BP神经网络模型进行发电量预测,利用某市工业园区屋顶BIPV系统实际数据开展实验。

1 基于改进K-均值的天气分类

对于一个确定的BIPV系统,其光伏阵列固有属性(例如安装朝向和阵列面积等)是确定的,可以认为这些信息隐含在历史发电量数据中,在这种情况下,BIPV系统发电量主要受太阳辐照度影响,而太阳辐照度又与具体的天气条件密切相关,不同天气条件下的发电量会存在较大差异,因此本文首先提出一种基于改进K-均值的天气分类方法,对GB/T22164-2008定义的33种专业天气类型进行自适应聚类,将其聚集为少数几类天气类型以降低后续预测模型的复杂度并提升预测精度。

1.1 特征提取

为了进行天气类型聚类,首先需要采用特征提取的方式对天气类型进行量化表征,本文定义温度因子、湿度因子和晴朗因子3种特征对不同天气状况进行描述,其具体定义为:

特征1:温度因子

温度因子 T 定义为一天中最高气温 T_{\max} 和最低气温

T_{\min} 之间的比值,即: $T=T_{\max}/T_{\min}$;

特征2:湿度因子

湿度因子 H 定义为一天中最大湿度 H_{\max} 和最小湿度 H_{\min} 之间的比值,即: $H=H_{\max}/H_{\min}$;

特征3:晴朗因子

晴朗因子 E 定义为地外辐照度理论值 E_{out} 与地表辐照度实测值 E_{in} 之间的比值,即 $E=E_{\text{out}}/E_{\text{in}}$ 。其中

$$E=1367 \times k^2 (\sin\delta \cdot \sin\varphi + \cos\delta \cdot \cos\varphi \cdot \cos\tau) \quad (1)$$

式中: k^2 表示日地距离修正系数; φ 表示太阳纬度, δ 表示太阳赤纬角, τ 表示太阳时角。对于每一天的天气数据分别计算上述三维特征并构成特征向量 $F_n=[T_n, H_n, E_n]$, $n=1,2,\dots,N$, N 为数据集中的样本总数,此时不同天气对应的文本信息被转换为量化数据的形式,便于后续处理分析。

1.2 改进的K-均值聚类

将天气信息量化为特征向量后,需要对其进行进一步分析以便将具有相似特征的天气划分为同一个子集,从而降低后续预测模型的复杂度并提升预测精度。K-均值聚类是数据分析领域中应用最为广泛的一种无监督聚类方法,根据数据之间欧氏距离的大小将其划分为不同的子类,具有原理简单,容易实现等优点^[16]。然而,传统K-均值算法在使用时需要预先设置聚类个数 K ,对于本文涉及的天气聚类需求,显然无法直接使用K-均值,因此本文提出一种基于样本局部密度的改进K-均值聚类方法,聚类过程中能够自动确定最优聚类个数 K ,从而提升聚类性能。

首先介绍空间密度,邻域半径,局部密度和样本贡献率的概念:

(1) 样本 F_n 的空间密度 $d(F_n)$ 定义为:

$$d(F_n) = \frac{\sum_{n=1}^N \|F_n - F_n\|_{l_2}}{\sum_{i=1, i \neq n}^N \|F_n - F_i\|_{l_2}} \quad (2)$$

式中, $\|\cdot\|_{l_2}$ 表示求变量的 l_2 范数。

(2) 样本 F_n 的邻域半径 $r(F_n)$ 定义为:

$$r(F_n) = \frac{1}{N} \sum_{n=1}^N \exp(-d(F_n)) \quad (3)$$

(3) 样本 F_n 的局部密度 $\rho(F_n)$ 定义为:

$$\rho(F_n) = d(F_n) \times r(F_n) \quad (4)$$

(4) 样本 F_n 的贡献率 $\eta(F_n)$ 定义为:

$$\eta(F_n) = \frac{\rho(F_n)}{\sum_{n=1}^N \rho(F_n)} \quad (5)$$

有了上述概念,则所提方法的具体实现步骤为:

步骤1:对于每个特征向量样本 F_n ,根据式(2)~式(4)计算得到对应的空间密度 $\text{dist}(F_n)$,邻域半径 $r(F_n)$ 和局

部密度 $\rho(F_n)$ ，并根据式(5)计算得到每个样本的局部密度贡献率；

步骤2: 将样本局部密度贡献率按从大到小排列，选取累积贡献率超过90%的样本个数为最优聚类数目 K ；

步骤3: 将局部密度最大的 K 个样本设置为初始聚类中心；

步骤4: 将样本空间中所有样本划分至与其距离最近的聚类中；

步骤5: 按照式(6)计算得到新的聚类中心：

$$F_k^* = \frac{1}{n_k} \sum_{n=1}^{n_k} F_n \quad (6)$$

式中， n_k 为第 k 个子类中的样本数。

步骤6: 重复步骤4~步骤6，直到连续两次聚类结果一致，则算法终止。

2 预测模型

2.1 BP神经网络

正如前文所述，BIPV 系统发电量呈现出典型的波动性和间歇性，传统时间序列建模方法由于无法描述数据中的非平稳信息，通常难以获得理想的预测性能，神经网络模型通过对人脑思维过程中神经元的活动状态进行模拟，能够实现对复杂非线性模型的自适应分析，因此适合用于对 BIPV 系统发电量进行预测，BP神经网络是其中应用最广泛、理论最成熟的一种模型^[17]，因此，本文采用BP神经网络对改进K-均值聚类得到的每个子类分别建立模型进行预测。

图1给出了BP神经网络典型的三层网络架构，从左到右依次为输入层神经元、隐含层神经元和输出层神经元。作为一种误差逆向传播算法，BP神经网络通过信号的正向传播过程和误差的反向传播过程的不断迭代，使输出结果与预期结果之间的误差逐渐减小直至满足模型精度要求。

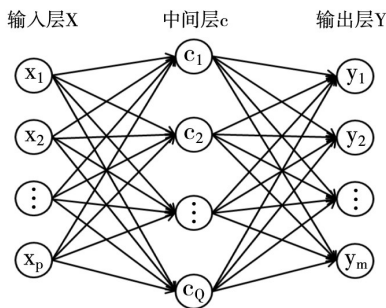


图1 三层BP网络结构

对于具有 n 个输入神经元的BP神经网络模型 $x_i, i=1, \dots, n$ 其输入信号通过正向传播经隐含层到达输出层的过程可以表示为：

$$y_k = \sum_{j=1}^l \omega_{jk} \cdot f\left(\sum_{i=1}^n w_{ij} x_i - \theta_j\right) \quad (7)$$

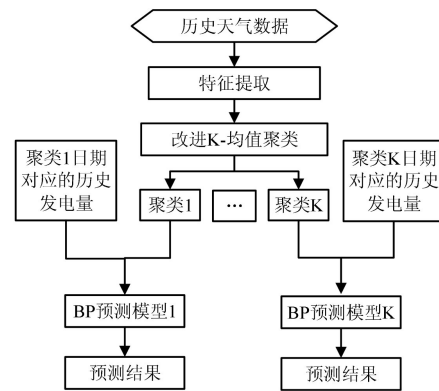
式中： $w_{ij}, i=1, \dots, n$ 和 $\omega_{ij}, j=1, \dots, l, k=1, \dots, m$ 分别为输入层到隐含层之间以及隐含层到输出层之间的连接权值， θ 为模型阈值， $f(\cdot)$ 为 Sigmoid 激活函数，具体形式为：

$$f(u) = \frac{1}{1 + \exp(-u)} \quad (8)$$

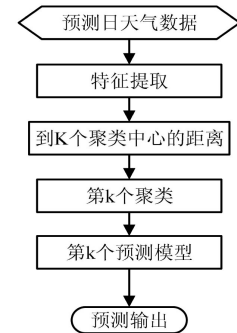
计算当前输出值 y_k 与期望输出值 y_k^* 之间的误差 δ ，并采用梯度下降法按照误差减小的方向对权值和阈值进行优化，从而完成BP神经网络的逆向传播。

$$\delta = \sum_{k=1}^m (y_k^* - y_k)^2 \quad (9)$$

2.2 算法流程总结



(a) 训练阶段流程图



(b) 测试阶段流程图

图2 所提方法预测流程图

利用所提基于改进K-均值聚类方法对BIPV系统发电量进行预测的流程如图2所示，其中图2(a)为训练阶段过程，图2(b)为测试阶段过程。根据图2(a)可以总结得到训练阶段的具体实现步骤为：

步骤1: 对历史发电量数据对应日期的天气数据提取三维特征向量，从而将天气数据集转变为特征向量集合；

步骤2: 利用所提改进的K-均值类算法对特征向量集合进行聚类分析，将其自动划分为 K 个子类；

步骤3: 利用每个子类的天气数据和对应日期发电量数据作为BP神经网络模型的输入，分别得到 K 个最优BP神经网络预测模型；

步骤4:分别利用BP神经网络模型对训练数据进行预测,得到预测结果,验证模型准确性。

根据图2(b)可以总结得到测试阶段的具体实现步骤为:

步骤1:对于待预测日期,从天气预报中获得其天气信息,进而提取三维特征向量;

步骤2:分别计算待测日期特征向量到K个聚类对应聚类中心的距离,并将其划分至距离最小的聚类中;

步骤3:利用步骤2得到的聚类对应的BP神经网络模型对进行预测,得到最终的预测结果。

3 试验与结果分析

为了验证本文提出的基于改进K-均值聚类的BIPV系统发电量预测性能,采用某市工业园区屋顶光伏电站2013年~2018年发电量数据开展验证实验,利用网络爬虫从当地气象站网站抓取全年的天气信息,通过对天气信息进行分析可知,该年度天气包含了全部33种天气类型。试验过程中,按月份随机从数据集中抽取30%的数据作为测试样本集,剩余70%的数据作为训练样本集,并保证训练和测试样本集中都包含33种天气类型。

表1 局部密度最大的8个样本对应的贡献率和累积贡献率

样本标号	局部密度贡献率/%	局部密度累积贡献率/%
1	54.42	54.42
2	21.58	76.00
3	17.33	93.33
4	2.16	95.49
5	1.21	96.50
6	1.20	97.70
7	0.30	98.00
8	0.10	98.10

根据图2所示流程,首先对训练样本集提取特征构成特征向量结合,然后利用所提改进K-均值聚类算法对特征向量集合进行聚类,表1给出了聚类过程中前8个局部密度最大的样本对应的样本贡献率和累积贡献率,可以看出前3个样本的累积贡献率已经超过93%,因此最优聚类数目K=3,图3给出了聚类结果,表2给出了3种聚类中包含的天气类型,可以看出聚类后的三种天气类型存在

较为明显的差异,其中第1类天气以晴天为主,天气比较晴朗,第2类天气主要为正常的雨雪天,第3类天气属于恶劣天气范畴,这种天气分类方式与实际生活中的主观分类方法较为一致,验证了所提改进K-均值聚类的性能。

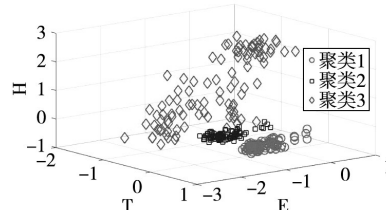


图3 聚类结果

表2 天气聚类与实际天气对照表^[18]

天气类型	GB/T22164-2008天气类型
1	晴,多云转晴,晴转多云,多云,雾,阴,阴转多云,多云转阴
2	小雨,小雪,中雨,中雪,阵雨,阵雪,冻雨,小到中雨,小到中雪,雷阵雨,雨夹雪,雷阵雨伴有冰雹
3	大雨,大雪,暴雨,暴雪,中到大雨,中到大雪,大暴雨,特大暴雨,大到暴雨,大到暴雪,暴雨到大暴雨,大暴雨特大暴雨,沙尘暴

完成天气聚类后,按照图2所示流程开展预测性能验证实验,图4给出了三种天气类型的预测结果,其中图4(a)为第一类天气类型中的晴天,图4(b)为第二类天气类型中的小到中雨,图4(c)为第三类天气类型中的大雪。图4中,“○—”表示BIPV系统实测发电量,“*—”表示所提方法得到的预测结果,同时为了对比,采用文献[19]所示马尔科夫链方法得到的预测结果,用“□—”表示。从图4(a)所示结果可以看出,对于晴天数据,所提方法对每个时刻的预测结果均接近于实际发电量,对所有时刻预测结果进行统计平均,得到的平均预测误差为5.6%,而马尔科夫链方法得到的平均预测误差为7.2%,并且对于任意时刻,其预测精度均优于所提方法;从图4(b)所示结果可以看出,对于测试用小到中雨天气,所提方法的预测结果与真实发电量最为接近,对于任意时刻,所提方法的预测精度均优于马尔科夫链方法,所提方法的平均预测误差为6.7%,文马尔科夫链方法的平均预测误差为8.9%,从图4(c)所示结果可以看出,对于测试用大雪天气,所提方法的预测结果与真实发电量最为接近,对于任意时刻,所提方法的预测精度均优于马尔科夫链方法,所提方法的平均预测

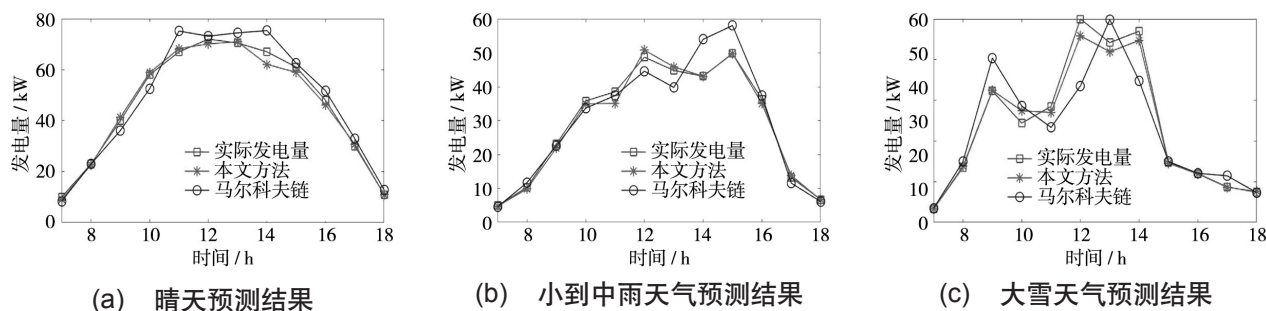


图4 不同天气类型发电量预测结果

误差为9.5%，马尔科夫链方法的平均预测误差为13.7%。同时对比图4(a)，图4(b)和图4(c)可知，随着天气状况的恶化，所提方法和马尔科夫链方法的预测性能均出现了不同程度的下降，但是相对而言所提方法性能下降不明显，表现出了一定的稳健性。

4 结束语

针对BIPV系统发电量预测难题，本文提出一种基于天气分类和改进K-均值聚类算法的BIPV系统发电量预测方法，首先通过特征提取和改进K-均值聚类将33种天气类型自动划分为3种天气子集，然后对每个子集分别建立BP神经网络模型进行预测，采用某地区实际光伏发电系统历史发电量数据开展验证试验，结果表明，本文方法能够获得更高的预测精度，相对于马尔科夫链模型在三种天气类型下预测精度分别提升1.6%，2.2%和4.2%，并且对于不同天气类型表现出了更高的适应能力，具有较高的应用前景。

参考文献：

- [1] 张海涛,李文娟,李雪峰,等.基于变分模态分解和时间注意力机制TCN网络的光伏发电功率预测[J].电测与仪表,2024,61(12):156-163.
- [2] 王步云,张成刚,郭斌.含风电和光伏的电力系统多目标经济调度[J].自动化技术与应用,2020,39(3):108-112.
- [3] 刘邦银,蔡涛,段善旭.建筑继承光伏的发展概述[J].高科技与产业化,2009(3):104-106.
- [4] 中华人民共和国国务院.国务院关于促进光伏产业健康发展的若干意见[EB/OL].北京:国务院,(2013-07-15)[2013-07-19].http://www.gov.cn/jxqgk/pub/govpublic/mrlm/201307/t20130715_66188.html.
- [5] 徐静,陈正洪,唐俊,等.建筑光伏并网发电系统的发电量预测初探[J].电力系统保护与控制,2012,40(18):82-87.
- [6] 易旷怡.太阳能光伏建筑一体化协同设计研究[D].北京:北京交通大学,2013.

(上接第20页)

合能源系统优化调度[J].电力建设,2021,42(10):28-39.

- [10] 罗首权,丁孝华,韩韬,等.基于目标级联分析法的区域源网荷储系统日前运行优化[J].电工电能新技术,2021,40(7):11-19.
- [11] 余涛,冯斌,韦冬妮,等.含分布式能源的主动配电网“源-网-荷-储”协调经济调度[J].水利水电技术(中英文),2021,52(6):215-222.
- [12] 孙惠,翟海保,吴鑫.源网荷储多元协调控制系统的研究及应用[J].电工技术学报,2021,36(15):3264-3271.
- [13] 张忠会,雷大勇,李俊,等.基于自适应 ϵ -支配多目标粒子群算法的含SOP的主动配电网源-网-荷-储双层协同规划模型[J].电网技术,2022,46(6):2199-2212.
- [14] 张高航,李凤婷.计及源荷储综合灵活性的电力系统日前优化调度[J].电力自动化设备,2020,40(12):159-167.

[7] 王嘉媛,尚可政,王式功,等.甘肃河西太阳能光伏组件最佳倾角及辐照量研究[J].中国科技论文,2015,10(15):1757-1764.

- [8] NORDIN N.Prediction of AC power output in grid-connected photovoltaic system using Artificial Neural Network with multivariable inputs 2016 IEEE Conference on Systems[C].Process and Control(ICSPC),2016:192-195.
- [9] 龙承潮.建筑用光伏发电量精准预测方法与光伏组串优化设计研究[D].北京:北京建筑大学,2015.
- [10] 江娥.武汉市光伏组件安装的最佳倾角分析[D].武汉:武汉工程大学,2015.
- [11] MERAL M E,DINCER F.A review of the factors affecting operation and efficiency of photovoltaic based electricity generation systems[J].Renewable and Sustainable Energy Reviews,2011,15(5):2176-2184.
- [12] 史佳琪,张建华.基于深度学习的超短期光伏精细化预测模型研究[J].电力建设,2017,38(6):28-35.
- [13] 叶林,陈政,赵永宁,等.基于遗传算法-模糊径向神经网络的光伏发电功率预测模型[J].电力系统自动化,2015,39(16):16-22.
- [14] 于群,朴在林,胡博.基于EEMD和BP神经网络的短期光伏功率预测模型[J].电网与清洁能源,2016,32(7):132-137.
- [15] 代倩,段善旭,蔡涛,等.基于天气类型聚类识别的光伏系统短期无辐照度发电预测模型研究[J].中国电机工程学报,2011,31(34):28-35.
- [16] 管霖,赵琦,周保荣,等.基于多尺度聚类分析的光伏功率特性建模及预测应用[J].电力系统自动化,2018,42(15):24-30.
- [17] 丁明,王磊,毕锐.基于改进BP神经网络的光伏发电系统输出功率短期预测模型[J].电力系统保护与控制,2012,40(11):93-99.
- [18] 蒋小波,徐小艳,刘乐平,等.天气分类和BP神经网络的光伏短期出力预测[J].电源技术,2020,12(20):2232-2239.
- [19] 海涛,闻科伟,周玲,等.基于气象相似度与马尔可夫链的光伏发电预测方法[J].广西大学学报,2015,40(6):1452-1460.

作者简介:戴姝(1992-),女,本科,工程师,研究方向:电力工程建筑设计。

[15] 杨谦,刘继春,蒋万泉.光伏不同渗透率下考虑源网荷储深度互动的电力系统调峰策略[J].电力建设,2021,42(9):74-84.

- [16] 帅挽澜,朱自伟,李雪萌,等.考虑风电消纳的综合能源系统“源-网-荷-储”协同优化运行[J].电力系统保护与控制,2021,49(19):18-26.

作者简介:黄健(1976-),男,高级工程师,研究方向:电力系统规划和运行。