

基于生成对抗网络的光伏出力区间预测方法

顾永涛¹, 徐泽禹², 盛庆博³, 张卫⁴, 李炜³, 谭嘉⁵, 李知艺⁵

- (1. 中国石油化工股份有限公司胜利油田分公司 孤东采油厂, 山东 东营 257051;
2. 国家电网有限公司华东分部, 上海 200120;
3. 中国石油化工股份有限公司胜利油田分公司 技术检测中心, 山东 东营 257062;
4. 中国石油化工股份有限公司胜利油田分公司 安全环保质量管理部, 山东 东营 257062;
5. 浙江大学 电气工程学院, 杭州 310007)

摘要: 准确的光伏出力(即发电功率)预测是光伏发电设备规模化并网后,保障电网的安全性和经济性的基础。针对传统点预测方法难以表征预测结果概率可信度的问题,研究了基于条件生成对抗网络的光伏出力区间预测方法。通过建立历史测量数据与预测光伏出力间的映射关系,生成光伏出力可能发生的场景。在此基础上,用 Bootstrap 法和区间估计法计算置信区间的上、下界以确定光伏出力的预测区间范围。算例结果表明,使用该方法得到的短期光伏出力预测区间能准确覆盖实测功率,为电网安全经济运行提供有价值的决策信息。

关键词: 光伏出力预测; 生成对抗网络; 区间预测; 预测场景; 区间估计

DOI: 10.15938/j.jhust.2021.02.013

中图分类号: TM615 **文献标志码:** A **文章编号:** 1007-2683(2021)02-0100-10

Interval Forecast of Photovoltaic Power Output Based on GAN

GU Yong-tao¹, XU Ze-yu², SHENG Qing-bo³, ZHANG Wei⁴, LI Wei³, TAN Jia⁵, LI Zhi-yi⁵

- (1. Gudong Oil Production Plant, Shengli Oilfield, Sinopec, Dongying 257051, China;
2. East China Branch of State Grid, Shanghai 201200, China;
3. Technical Testing Center, Shengli Oilfield, Sinopec, Dongying 257062, China;
4. Safety, Environmental Protection and Quality Management Department, Shengli Oilfield, Sinopec, Dongying 257062, China;
5. College of Electrical Engineering, Zhejiang University, Hangzhou 310007, China)

Abstract: Accurate photovoltaic (PV) power output forecast is the basis of ensuring safe and economic operation of the grid with high ratio of PV power. To solve the problem that point forecast is difficult to express the probability credibility, an interval forecast method of PV power output based on conditional generative adversarial network was studied. The mapping relationship between the historical measurement data and the predicted PV power output was established, then the probable scenarios of PV power output were obtained. And on this basis, the bootstrap method and interval estimation were used to calculate upper and lower bounds of the confidence interval, so as to determine the forecast interval of PV power. Experimental results have shown that the forecast intervals of short-term PV power output obtained by the proposed method can cover the real measured power values, and provide valuable decision information for the safe and economic operation of power grid.

Keywords: photovoltaic forecast; generative adversarial network; interval forecast; forecast scenario; interval estimation

收稿日期: 2020-12-01

基金项目: 中国石油化工股份有限公司科技项目(319020-7); 中央高校基本科研业务费专项资金(2020QNA4013); 国家自然科学基金(52007164)。

作者简介: 顾永涛(1968—),男,硕士,高级工程师;

徐泽禹(1988—),男,硕士,工程师。

通信作者: 李知艺(1989—),男,博士,正高级研究员,博士研究生导师, E-mail: zhiyi@zju.edu.cn.

0 引言

近年来,全球范围内能源转型趋势明显。作为最具前景的可再生能源利用形式之一,光伏发电在电力系统中的渗透率不断提高^[1]。截止到2020年6月,我国光伏发电累计装机容量从2010年26万kW增至2.16亿kW,年均增幅达126%^[2]。目前,全球光伏新增装机容量继续呈不断上涨的趋势,预计到2021年全球光伏累计装机容量将达到700GW^[3]。然而,受太阳光照、云层、湿度等多种气象因素影响,光伏出力(即发电功率)具有很大的间歇性和不确定性。当光伏发电设备规模化并网后,光伏出力的不确定性将严重威胁电网的稳定性和安全性^[4]。因此,准确的光伏出力预测是电网调度部门的重要决策依据,可保证其未雨绸缪,提前调配储能资源和备用机组,以保障电网的安全经济运行水平^[5]。

目前,大多数的光伏出力预测方法是点预测,这类方法得到的预测结果一般是确定性的出力值预测,欠缺对光伏出力不确定性的考虑。因此提供给调度部门关于光伏出力可能的变化信息较少,难以表达预测结果的概率可信度^[6-9]。基于点预测,调度部门只能依据确定性的预测值调配储能资源和备用机组。当实际出力与预测值存在较大差距时,会导致光伏发电消纳不足或备用容量不足,将严重威胁电网的稳定性和经济性^[10]。因此,针对点预测的不足,有部分学者围绕光伏出力开展区间预测的研究。区间预测是研究光伏出力在给定置信度下的预测区间,它由预测结果的上、下界构成。相比于点预测,区间预测能够量化光伏出力的不确定性波动,为光伏并网提供更多有价值的决策信息。依据区间预测的信息,电网调度部门可以灵活调整调度策略,从而可以减少光伏发电损失或减少备用机组不足的窘境,提升光伏并网的经济性和安全性。

随着深度学习技术的发展,越来越多学者将深度学习应用于光伏发电预测,并且验证了深度学习的可行性与优越性。文[1]采用长短时记忆神经网络进行光伏的日前预测,将温度、云层数等气象条件和日前光伏发电值共同作为输入,训练历史数据得到较高精度的光伏预测结果。文[2]首先采用二维小波去噪方法将含噪声的数据预处理,然后基

于k-means算法将预测模型分为4种模式,并应用粒子群优化神经网络算法对每个预测子模型进行训练,最后通过算例验证了模型的泛化能力和预测精度。文[3]使用深度置信网络和长短时记忆网络分别进行训练,对各个深度模型的训练结果动态组合,建立了一种动态组合光伏预测模型。文[4]使用有监督学习和无监督学习并存的训练方式,综合考虑模型的节点、层数和迭代次数的影响确定了最优参数,验证了深度学习模型的优越性。文[5]针对光伏出力预测精度依赖于天气状况的问题,以典型天气日数据作为训练样本,采用孤立森林和长短时记忆神经网络预测相似天气日的光伏发电功率,得到了较高精度的预测结果。

近年来,作为典型深度生成模型的生成对抗网络^[14](generative adversarial nets, GAN)在可再生能源场景生成方面取得了较多成果。GAN的生成器能无监督学习可再生能源出力特征并生成符合数据分布规律的场景集,但现有研究缺少利用生成器输入变量和输出数据间的相关性和未知性关系进行研究。同时,利用GAN出色的生成能力能够生成多样化的场景,从而确定光伏出力可能发生的区间,这对于研究光伏出力区间预测具有较大的价值。

针对上述研究现状,本文首次提出基于条件生成对抗网络的光伏出力区间预测模型。条件生成对抗网络不仅能够学习到光伏的出力特征,可以构造短期的历史测量数据与预测功率值间的映射关系,还能生成光伏出力可能发生的多个场景,从而有效地实现光伏出力的区间预测。本文将历史发电数据作为条件标签,将其与噪声向量拼接输入到生成器。通过生成器和判别器的对抗博弈,生成器可自动学习到光伏发电功率与条件标签的映射关系。算例结果表明,本文所提出的光伏出力区间预测模型不仅能够学习到上述条件映射,而且能够生成较高质量的光伏出力预测区间,为光伏并网的经济安全运行提供有价值的决策信息。

1 条件生成对抗网络(CGAN)

GAN是一种深度生成式模型,由Goodfellow在2014年首次提出。GAN由生成器和判别器两个重要部分组成,生成器将低维空间的噪声向量 z 映射至高维空间,生成尽可能接近真实数据的新样本,判

别器则负责判断输入数据是否为真实数据。通常,生成器和判别器是两个结构对称的卷积神经网络,利用卷积网络较强的特征提取能力可以提升生成的数据质量^[7]。通过两个网络的对抗训练,最终生成器学习到原始数据的分布特征,并且能生成符合真实规律的数据。GAN作为一种生成式模型,在图像生成和缺失数据重建等方面有着广泛应用^[8],但目前却较少被用来作为时间序列预测模型。这是因为生成器虽然可以学习历史数据的分布规律,但由于输入为随机的噪声向量,使得用原始 GAN 难以精确预测未来某个时刻的数据。而条件生成对抗网络^[9] (conditional generative adversarial nets, CGAN) 是在 GAN 基础上进行改进的一种生成式网络,模型的输入中添加了附加条件,可用于指导数据的生成过程,解决了传统 GAN 在有噪声的情况下无法控制生成数据模式的问题。CGAN 的结构如图 1 所示,噪声向量 z 与条件标签 c 合并后输入到生成器 G , 产生生成样本。生成样本和真实样本分别和对应条件标签合并后又输入到判别器 D , 最终输出其输入样本在条件 c 下为真实样本的概率。在博弈训练过程中,生成器试图生成在条件 c 下接近真实数据的新样本,判别器则试图区分在条件 c 下的真实数据或生成数据,两者在这个过程中共同进步。本文把光伏发电历史数据作为训练集,在预测起点把样本划分为真实样本和条件标签,其中预测起点之前的时间序列作为条件,起点之后的数据作为真实样本。与原始 GAN 相比,CGAN 由于在网络的输入中添加了条件,能够学习到历史测量数据与预测功率值间的映射关系,因此能够较好地适应光伏出力预测问题。

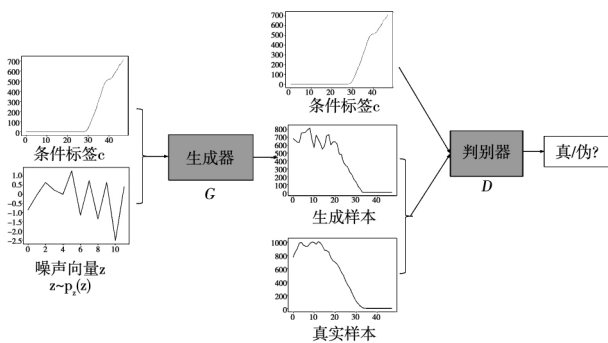


图1 CGAN 基本结构

Fig.1 Basic structure of CGAN

CGAN 中生成器和判别器的损失函数如下:

$$\begin{aligned} L_G &= -E_{z \sim p_z(z)} [D(G(z|c)|c)] \\ L_D &= -E_{x \sim p_r(x)} [D(x|c)] + E_{z \sim p_z(z)} [D(G(z|c)|c)] \end{aligned} \quad (2)$$

式中: E 表示计算期望; $G(z|c)$ 表示在条件 c 下的生成样本; $D(\cdot)$ 表示判别器判定输入样本为真实数据的概率; $z \sim p_z(z)$ 表示从高斯分布 $p_z(z)$ 中采样得噪声向量 z ; $x \sim p_r(x)$ 表示从实际分布采样得一组真实样本 x 。

同时训练 CGAN 的生成器和判别器,总的训练过程目标函数为:

$$\min_G \max_D V(G, D) = E_{x \sim p_r(x)} [D(x|c)] - E_{z \sim p_z(z)} [D(G(z|c)|c)] \quad (3)$$

构造该目标函数的原则是: 当训练生成器时,生成器试图生成既符合条件 c 的映射又符合真实数据分布的数据,从而欺骗判别器,此时 CGAN 的训练目标是 minimized 上述目标函数。当训练判别器时,它试图区分在条件 c 下的真实数据或生成数据,因此训练判别器时需要最大化上述目标函数。CGAN 训练过程本质上是一个极小极大值双方博弈,在博弈结束后,生成器即可学习到真实数据与条件 c 的映射关系,从而能在给定条件 c 下生成服从真实规律的数据。

为避免训练过程中梯度消失等问题,采用基于 Wasserstein 距离的损失函数来代替传统的 JS 散度,其定义如式(4)所示。Wasserstein 距离可用来衡量两个分布的相似性,其值越小,表明两个分布相似性越高。

$$W(p_r(x), p_g(x')) = \inf_{\gamma \in \Pi(p_r(x), p_g(x'))} E_{(x', x) \sim \gamma} [\|x' - x\|] \quad (4)$$

式中: x 和 x' 分别表示真实样本和生成样本; $\Pi(p_r(x), p_g(x'))$ 是以实际分布 $p_r(x)$ 和生成分布 $p_g(x')$ 为边缘分布的联合概率分布 γ 的集合; $W(p_r(x), p_g(x'))$ 表示 $\gamma(x', x)$ 期望的下确界。

一般地,直接求解 Wasserstein 距离比较困难,所以采用如下 Kantorovich-Rubinstein 对偶形式:

$$W(p_r(x), p_g(x')) = \frac{1}{K} \sup_{\|D\|_L \leq K} E_{x \sim p_r(x)} [D(x|c)] - E_{z \sim p_z(z)} [D(G(z|c)|c)] \quad (5)$$

式中: $\|D\|_L \leq K$ 表明判别器需要满足 K -Lipschitz 连续条件,即约束判别器参数的梯度在一定范围内。

2 基于 CGAN 的光伏出力区间预测

2.1 光伏出力预测建模

为指导生成器预测光伏出力的过程,在给定历史数据 $c = \{x_0, \dots, x_t\}$ 下,使用 CGAN 建立 $x' = \{x_{t+1}, \dots, x_{t+h}\}$ 的光伏出力预测模型,其中 h 为预测的时间步长。这样一来,模型构建了历史数据与预测功率值间潜在的映射关系,然后通过生成器出色的生成能力得到多个有效的光伏出力预测场景,在给定置信度下即可得到预测区间的上、下界。

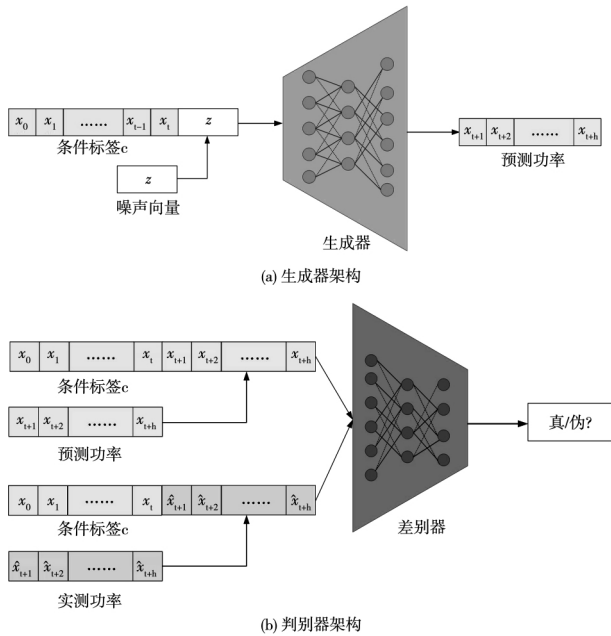


图2 生成器和判别器架构

Fig. 2 The structure of the generator and the discriminator

图2展示了生成器和判别器的架构及其输入输出变量,其中历史数据作为条件标签分别提供给生成器和判别器。从均值为0、标准差为1的高斯分布中采样得噪声向量 z , 将其与条件 $c = \{x_0, \dots, x_t\}$ 横向拼接后输入到生成器,生成器试图输出与历史数据映射的预测功率 $\{x_{t+1}, \dots, x_{t+h}\}$ 。对于判别器,其输入既来自生成器输出的预测功率,也来自实测功率 $x = \{\hat{x}_{t+1}, \dots, \hat{x}_{t+h}\}$ 。两者分别与 $c = \{x_0, \dots, x_t\}$ 横向拼接输入到判别器,判别器试图识别出在条件 c 下无效的预测值,同时输出其输入样本为真实值的概率。将预测功率判定为真实值的概率越小,则表明判别器对生成样本的识别能力越强。

因此,基于条件 $c = \{x_0, \dots, x_t\}$ 的 CGAN 损失函数如下:

$$L_G = -E_{z \sim p_z(z)} [D(G(z | \{x_0, \dots, x_t\}) | \{x_0, \dots, x_t\})] \quad (6)$$

$$L_D = -E_{x \sim p_r(x)} [D(\{\hat{x}_{t+1}, \dots, \hat{x}_{t+h}\} | \{x_0, \dots, x_t\})] + E_{x' \sim p_g(x')} [D(\{x_{t+1}, \dots, x_{t+h}\} | \{x_0, \dots, x_t\})] \quad (7)$$

式中: E 表示计算期望; $z \sim p_z(z)$ 表示从高斯分布 $p_z(z)$ 中采样得噪声向量 z ; $G(z | \{x_0, \dots, x_t\})$ 表示在条件 $c = \{x_0, \dots, x_t\}$ 下的预测功率; $D(\cdot)$ 表示判别器判定输入样本为实测功率的概率; $x \sim p_r(x)$ 表示从实际数据分布中采样得实测功率 $x = \{\hat{x}_{t+1}, \dots, \hat{x}_{t+h}\}$; $x' \sim p_g(x')$ 表示从生成数据分布中采样得预测功率 $x' = \{x_{t+1}, \dots, x_{t+h}\}$ 。生成器和判别器不断进行博弈对抗,在训练结束后生成器学习到历史数据 $\{x_0, \dots, x_t\}$ 与实测功率 $\{\hat{x}_{t+1}, \dots, \hat{x}_{t+h}\}$ 间的映射关系。在给定随机噪声和 $\{x_0, \dots, x_t\}$ 的输入下,生成器可生成多个有效的光伏出力预测场景,为区间预测提供关键输入信息。

2.2 预测区间生成方法

光伏出力区间预测是在给定置信度下计算出光伏出力的上界和下界,从而构成一个可能的区间范围。相比于点预测,区间预测能够量化光伏出力的不确定性波动,为光伏并网提供更多有价值的决策信息,因而区间预测正逐步受到研究学者的重视。目前常用的非参数区间预测方法有核密度估计^[20]、分位数回归^[21]等,但这些方法存在的问题是计算复杂、需要大量的样本。本文采用了 Bootstrap 法^[22]来计算光伏出力预测区间,该方法无需预先假设分布,而且在数据波动明显时对于区间预测有着较高的准确率。基于 Bootstrap 法的预测步骤如图3所示。

假设对每个测试样本,生成器生成 M 个预测场景。基于 Bootstrap 进行重采样,对这 M 个预测场景,按有放回地抽取 M 次。重复 N 次,即可得到 N 个容量为 M 的 Bootstrap 样本: $(P_1^k, P_2^k, \dots, P_M^k)$, $k = 1, 2, \dots, N$ 。此时,计算每个样本的算数均值为

$$P^k = \frac{1}{M} \sum_{i=1}^M P_i^k \quad (8)$$

然后,在 $1 - \alpha$ 置信度下对 $(P^1, P^2, \dots, P^N)$ 进行区间估计,置信区间的上、下界分别为

$$\left. \begin{aligned} P_{U\alpha} &= \bar{P} + \frac{\sigma}{\sqrt{N}} z_{\alpha/2} \\ P_{L\alpha} &= \bar{P} - \frac{\sigma}{\sqrt{N}} z_{\alpha/2} \end{aligned} \right\} \quad (9)$$

式中: $\bar{P}$ 和 σ 分别表示 $(P^1, P^2, \dots, P^N)$ 样本中光伏

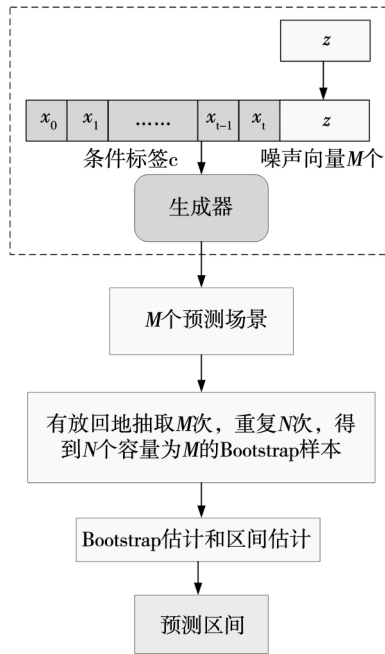


图3 基于 Bootstrap 的区间预测流程图

Fig. 3 Flow chart of bootstrap-based interval forecast

出力的均值和标准差; $z_{\alpha/2}$ 是在 α 显著水平下的置信因子, 可通过查 z 值表获得。一般地, 取置信度 $1 - \alpha$ 为 0.95 时(即在 0.05 显著水平下)的置信因子为 1.96。通过引入 Bootstrap 法在有限的预测场景下进行自采样, 不仅能反映光伏出力的概率分布、提高预测精度, 还能显著减少计算时间。

2.3 预测区间评价指标

为了评估所提方法的预测性能, 引入区间预测的评价指标来对预测结果进行分析。目前区间预测的评价指标主要有预测区间覆盖率^[23]、预测区间平均宽度^[24]、覆盖宽度^[25]等, 通常选取前两个指标进行综合评价。

1) 预测区间覆盖率

覆盖率(coverage rate, CR) 用来评价预测区间对实测值的覆盖情况, 定义为实测值 x_i 落在预测区间内的概率, 即:

$$CR = \frac{1}{n} \sum_{i=1}^n \varepsilon_i \quad (10)$$

式中: ε_i 为第 i 个测量点的布尔量, 当 x_i 落在预测区间内, ε_i 取 1, 否则取 0; n 为采样点的总数。

2) 预测区间平均宽度

预测区间平均宽度(predict interval width, PIW) 是光伏出力不确定性变化的表征, 定义为预测区间上、下界的平均差, 即:

$$PIW = \frac{1}{n} \sum_{i=1}^n (U_i - L_i) \quad (11)$$

式中: U_i 和 L_i 分别是预测区间在第 i 个采样点的最大值和最小值; n 为采样点总数。

覆盖率和预测区间平均宽度分别从两个角度反映预测性能, 覆盖率越高说明预测结果越准确, 预测区间宽度越小说明预测结果越有效、越具有参考价值。理想的区间预测应有较高的覆盖率和较小的区间宽度, 但往往随着覆盖率的增加, 预测区间宽度也随之增大。导致预测结果的不确定性增加, 降低了预测结果对于系统决策的参考价值。

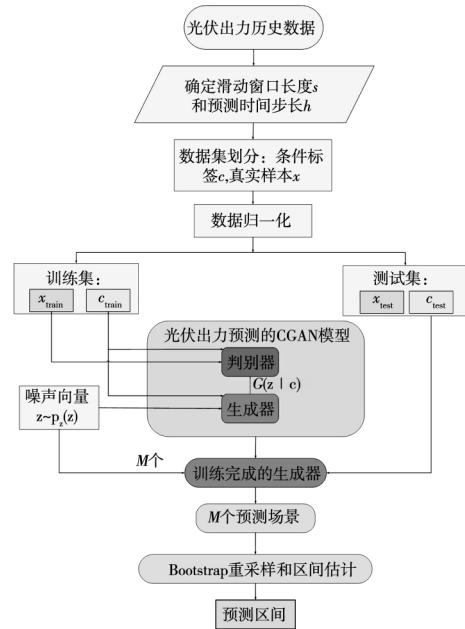


图4 光伏出力区间预测模型框架

Fig. 4 Framework of PV power output interval forecast model

2.4 光伏出力区间预测模型框架

光伏出力区间预测的整体算法流程框图如图 4 所示。首先对光伏出力实测数据预处理, 确定预测的时间步长 h , 以长度为 s 的滑动窗口划分数据集。每组数据中历史数据部分作为条件标签 c , 而待预测部分作为真实样本 x 。将数据归一化并对数据集按 9:1 的比例划分训练集和测试集后, 基于训练集对用于光伏出力预测的 CGAN 模型进行训练。通过生成器和判别器的对抗博弈, 生成器能够在给定历史数据下生成接近实测功率的光伏出力预测场景。另外, 为构成含更多不确定性信息的预测区间, 需要生成 M 个预测场景。采用 Bootstrap 法对这些场景进行重采样, 用区间估计的方法, 在给定置信度下计算这些生成场景的置信区间, 最终得到区间预测的结果。

3 实际系统算例分析

3.1 数据获取及处理

选用2012年12月1日至2020年9月30日期间比利时全国范围内的光伏出力实测数据^[24]作为算例样本,数据分辨率为15 min,原始数据集共有271680个采样点数据。虽然该数据集已经过相关专家的修正处理,但在数据清洗过程中我们仍发现数据中存在9个nan值,利用这些nan值前4个序列数据的平均值替代nan。考虑到生成器将低维空间映射至高维的独特优势,预测的时间步长不宜取得太小,而过大的步长又会降低预测精度。因此为实现良好的预测效果,算例将12 h作为预测的时间步长。同时用长度为48个点(窗口长度为12 h)的滑动窗口对数据集进行切片,共得到5659组数据,每组数据均包括条件标签 c (48个点)和光伏出力实测值 x (48个点)。为方便判别器卷积层参数的设计,本文将光伏出力实测值维度变换为(3, 4, 4)。对切片后的数据集按9:1的比例划分训练集和测试集,所有数据都归一化为 $[0, 1]$ 。

3.2 CGAN 模型参数设置

在Python编程环境中搭建基于PyTorch深度学习框架的CGAN模型进行算例验证。其中,生成器和判别器均包含多层神经网络结构。生成器的输入由特征量为12的一维噪声向量与条件标签 c 横向拼接组成,其中噪声向量服从均值为0、标准差为1的高斯分布。通过一层全连接层和三层卷积层,生成器将输入数据映射至光伏出力实测数据 x 的相同维度(3, 4, 4)。除输出层以外的每一层后均添加了批归一化层,而且采用LeakyReLU作为激活函数,以避免训练过程中梯度消失的问题。判别器的结构与生成器基本对称,由三层卷积层和一层全连接层组成。同样地,除输出层以外的每一层后均添加了批归一化层,隐藏层采用LeakyReLU作为激活函数。但由于采用了基于Wasserstein距离的损失函数,因此输出层不添加任何激活函数^[27]。

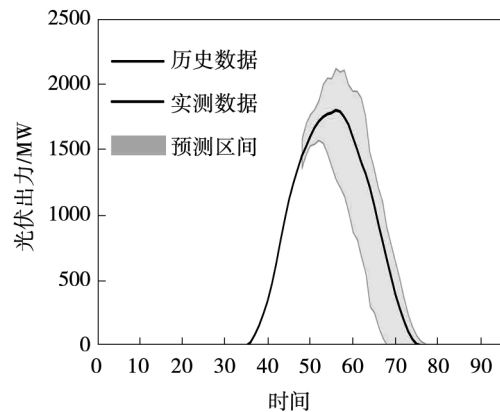
此外,采用每更新判别器4次,更新1次生成器的交替训练方式,以提高判别器的判别能力,进而促进生成器的收敛^[28]。训练过程中,批处理数设为64,采用RMSProp优化算法更新网络模型权重参数,并将学习率设置为 1×10^{-4} 。

3.3 结果分析

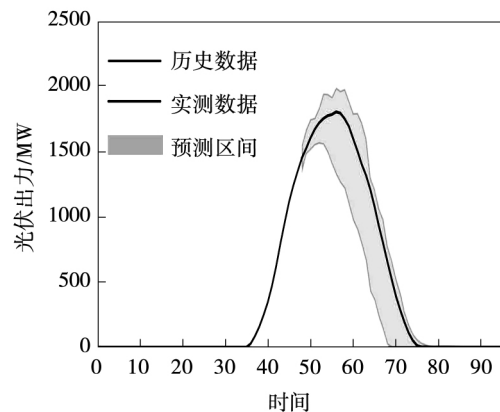
模型训练完成后,从均值为0、标准差为1的高斯分布中随机采样一组噪声向量 z 。将其与测试集数据中的条件标签 c 合并后输入到生成器,得到在条件 c 下的光伏出力预测结果。重复上述过程 M

次,即对每个测试样本采样 M 个噪声向量,并分别输入生成器后得到 M 个光伏出力预测场景。采用Bootstrap法(这里 N 取50)和区间估计法,计算这些生成场景在不同置信度下的置信区间。

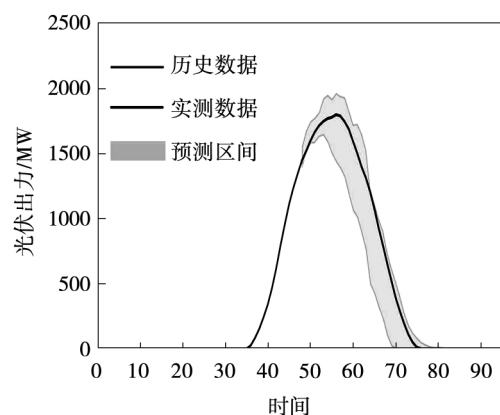
图5给出了在 $M=100$ 的情况下,随机选取的某测试样本分别在95%、90%和85%置信度下的预测区间。图中绿线表示已知的历史光伏出力情况,红线表示光伏出力实测值,蓝色区域为得到的预测区间。本文所有的算例结果图,横坐标均为时间。



(a) 置信度为95%时的预测区间



(b) 置信度为90%时的预测区间



(c) 置信度为85%时的预测区间

图5 不同置信度下的光伏出力预测区间($M=100$ 时)

Fig. 5 PV power output forecast intervals under different confidences when $M=100$

分辨率为 15 min。从图上可以看出,在不同置信度下,预测区间都能有效覆盖实测值,说明模型对于光伏出力的预测精度较高。

为探究预测场景的个数 M 对于预测区间的影响,分别在 $M = 10$ 、 $M = 100$ 和 $M = 1\,000$ 三种情况下进行实验,计算得到 95%、90% 和 85% 置信度下的预测区间,分别如图 5、图 6、图 7 所示。

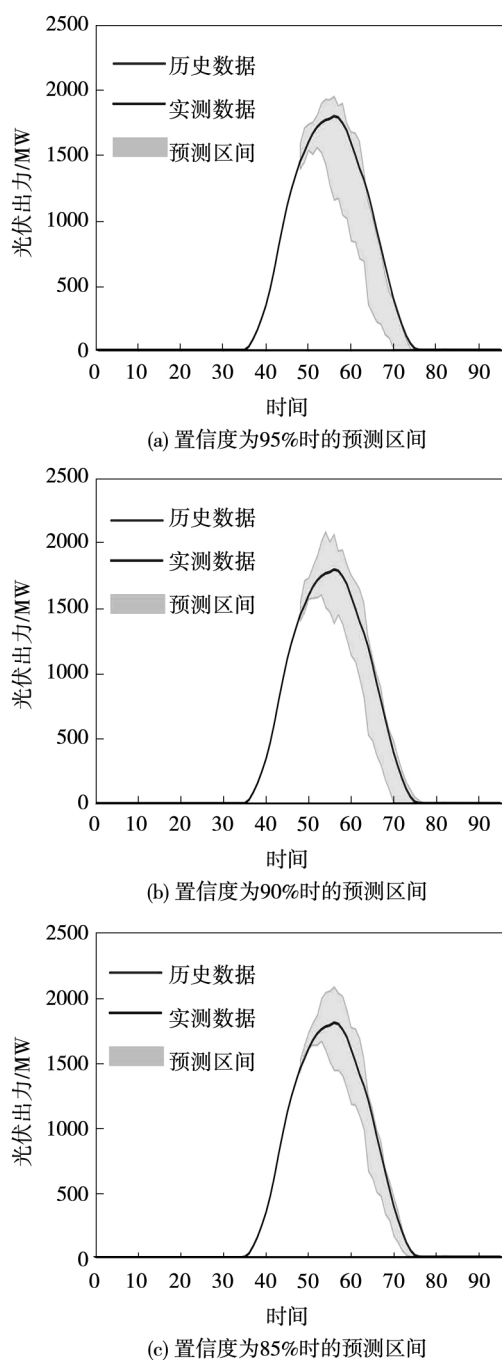


图6 不同置信度下的光伏出力预测区间($M = 10$ 时)

Fig. 6 PV power output forecast intervals under different confidences when $M = 10$

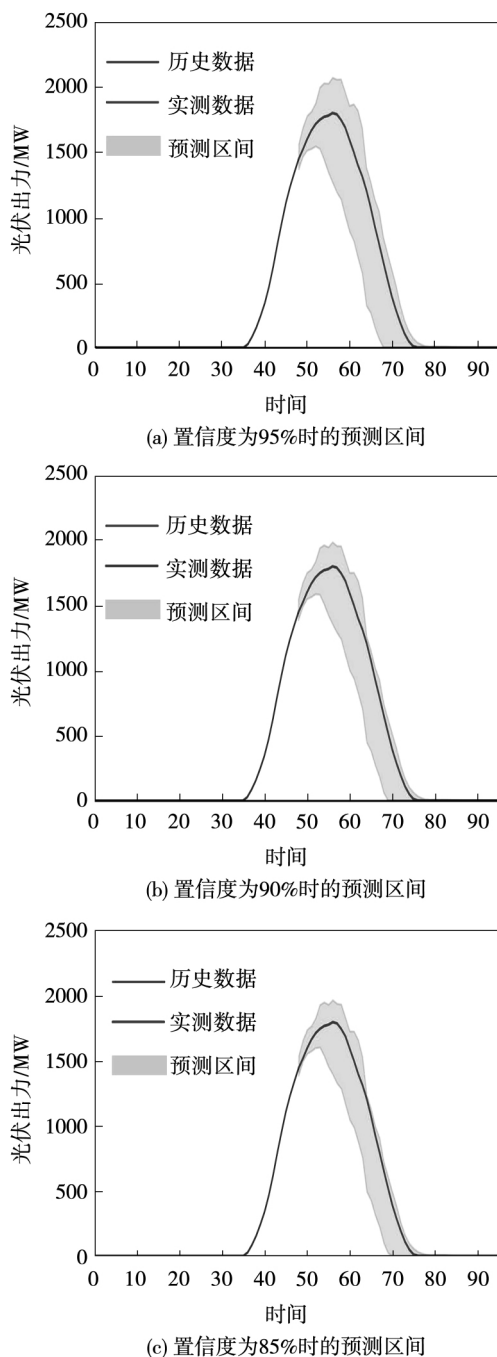


图7 不同置信度下的光伏出力预测区间($M = 1\,000$ 时)

Fig. 7 PV power output forecast intervals under different confidences when $M = 1\,000$

从图 6 可以看出,当 $M = 10$ 时,预测区间无法对全部实测值进行有效覆盖,即覆盖率较低。而且相比于图 5 的结果,在同一置信度下,图 6 的预测区间宽度更小。这是由于当 M 是一个很小的值时,预测场景样本不够丰富,导致所构建的预测区间窄。其中包含的光伏出力不确定性信息太少,因而对实测值的覆盖率低。当 $M = 100$ 或 $M = 1\,000$ 时,如

图5和图7所示,在同一置信度下,对同一测试样本预测得到的区间几乎相同。这说明当 M 增加到一定值时,预测区间不会再有较大改变,此时预测场景的个数 M 对于预测区间几乎无影响。这为 M 的选取提供了指导: M 不能选的太小,否则预测区间无法有效覆盖实测值; M 也不必选的过大,以减少计算时间。

表1给出了当 $M=10$ 、 $M=100$ 和 $M=1\,000$ 时,在95%、90%和85%三种置信度下,针对所有测试样本计算的区间预测性能指标。从表中数据可以看出,在同一置信度下, $M=100$ 与 $M=1\,000$ 时的覆盖率和预测区间平均宽度几乎相同。这进一步佐证了当 M 达到一定值时,区间预测性能几乎不会再改变。并且当 M 一定时,随着置信度的增大,覆盖率逐渐变大,说明预测精度在不断提高。但同时预测区间平均宽度也在增大,较大的区间宽度增加了预测结果的不确定性,使得预测结果对于系统调度的指导性降低,其具有的参考价值 and 有效性减小。但总的来说,预测区间平均宽度远小于光伏发电的总监测容量(2 915.88 MW—4 037.14 MW),因此在这三种置信度下,预测结果的准确性和有效性都得到了保证。

表1 模型的区间预测评价指标

Tab. 1 Interval prediction evaluation indicators of the model

场景 个数	置信度/ (%)	评价指标	
		覆盖率/(%)	区间宽度/MW
$M=10$	95%	95.83	346.63
	90%	93.75	331.77
	85%	93.64	254.11
$M=100$	95%	97.91	385.73
	90%	94.26	310.49
	85%	93.75	280.50
$M=1\,000$	95%	97.91	370.84
	90%	95.83	313.81
	85%	93.75	273.11

为进一步验证所提方法对于光伏出力连续预测的性能,选取连续3天的某测试样本进行90%置信度下的区间预测。这里 M 取100,预测结果如图8所示。图中,实线表示光伏出力实测值,阴影区域表

示使用所提方法得到的预测区间。图中,预测区间对实测值实现了大范围的覆盖,表明预测精度较高。并且通过连续输入3天的历史测量数据,得到的预测区间上、下界变化趋势均与实测值一致,说明模型的生成器学习到了光伏出力的变化规律以及历史测量数据与预测功率值间的映射关系。通过这种映射关系,只需给定光伏出力历史测量数据就可以生成有效的预测场景,从而实现光伏出力区间的准确预测。

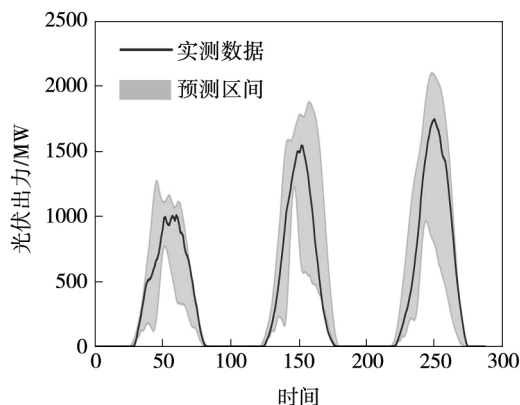


图8 90%置信度下连续3天的区间预测($M=100$ 时)

Fig. 8 Interval forecast for 3 consecutive days with 90% confidence when $M=100$

综上所述,算例结果表明所提方法在三种置信度下得到的预测区间均对实测值有着较高的覆盖率,且区间上、下界符合实际光伏出力的波动趋势,区间平均宽度远小于光伏发电的总监测容量,在可接受范围内。因此从设置预测时间步长为12 h的条件来看,模型在短期光伏出力预测中具有良好的准确性和适用性。

6 结 论

本文应用CGAN实现了短期光伏出力区间预测。该方法不涉及任何显式的建模,依靠纯数据驱动的方式描述光伏出力可能的波动范围,为电网储能资源和备用机组的配置提供决策支持。将历史发电数据作为网络的条件标签,与噪声向量拼接输入到生成器中。通过生成器和判别器的对抗博弈,CGAN能自动学习到光伏的出力特征,构造短期的历史测量数据与预测功率值间的映射关系。然后通过改变与条件标签拼接的噪声向量,输入到生成器中可生成多个有效场景。在此基础上,采用 Boot-

strap 和区间估计得到不同置信度下的预测区间。预测结果表明,模型在短期光伏出力预测中具有较高的准确性和有效性,而且可用于光伏出力连续预测。在后续研究中,可围绕预测时间步长的选取对模型预测性能的影响展开。

参考文献:

- [1] 张鹏,罗正华,唐成达,等. 一种基于改进萤火虫算法的光伏 MPPT 控制方法 [J]. 哈尔滨理工大学学报, 2020, 25(3): 53.
ZHANG Peng, LUO Zhenghua, TANG Chengda, et al. Photovoltaic MPPT Control Method Based on Improved Firefly Algorithm [J]. Journal of Harbin University of Science and Technology, 2020, 25(3): 53.
- [2] 国家能源局. 2020 年上半年光伏发电并网运行情况 [DB/OI]. (2020-07-31) [2020-12-01]. <https://www.nea.gov.cn/>.
National Energy Administration. Grid-connected Operation of PV Power Generation in The First Half of 2020. [DB/OI]. (2020-07-31) [2020-12-01]. <https://www.nea.gov.cn/>.
- [3] Solar Power Europe. Global market outlook 2017-2021: solar boom continues [EB/OI]. (2020-10-08) [2020-12-01]. <http://www.solarpowereurope.org/global-market-outlook-2017-2021-solar-boom-continues/>.
- [4] 刘伟,彭冬,卜广全,等. 光伏发电接入智能配电网后的系统问题综述 [J]. 电网技术, 2009, 33(19): 1.
LIU Wei, PENG Dong, BU Guangquan, et al. A Survey on System Problems in Smart Distribution Network With Grid-Connected PV Generation [J]. Power System Technology, 2009, 33(19): 1.
- [5] 于佳弘. 基于深度学习的太阳能光伏发电短期预测方法 [D]. 杭州: 浙江大学, 2018.
- [6] PEDRO H T C, COIMBRA C F M. Assessment of Forecasting Techniques For Solar Power Production With No Exogenous Inputs [J]. Solar Energy, 2012, 86(7): 2017.
- [7] 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.
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.
- [8] YANG D, JIRUTITIJAROEN P, WALSH W M. Hourly Solar Irradiance Time Series Forecasting Using Cloud Cover Index [J]. Solar Energy, 2012, 86(12): 3531.
- [9] ZENG Jianwu, QIAO Wei. Short-term Solar Power Prediction Using a Support Vector Machine [J]. Renewable Energy, 2013, 52(2): 118.
- [10] 黎敏,林湘宁,张哲原,等. 超短期光伏出力区间预测算法及其应用 [J]. 电力系统自动化, 2019, 43(3): 10.
LI Min, LIN Xiangning, ZHANG Zheyuan, et al. Interval Forecast Algorithm and Application of Short-Term PV Power output [J]. Automation of Electric Power Systems, 2019, 43(3): 10.
- [11] 李崎勇,班斓. 基于长短记忆神经网络的短期光伏发电预测技术研究 [J]. 华北电力大学学报(自然科学版), 2020, 47(4): 46.
LI Jiyong, BAN Lan. Research on Short-Term PV Power Generation Forecasting Based on LSTM [J]. Journal of North China Electric Power University, 2020, 47(4): 46.
- [12] 孙新程,万玥,丁宏,等. 基于小波去噪-KPCA 神经网络的光伏功率预测方法 [J]. 电力工程技术, 2020, 39(2): 110.
SUN Xincheng, WAN Yue, DING Hong, et al. Forecasting Method of Photovoltaic Output Power Based on Wavelet Denoising/KPCA/PSO [J]. Power Engineering Technology, 2020, 39(2): 110.
- [13] 熊图,赵宏伟,蔡智洋,等. 动态组合深度学习模型在短期负荷及光伏功率预测中的应用 [J]. 可再生能源, 2020, 38(4): 458.
XIONG Tu, ZHAO Hongwei, CAI Zhiyang, et al. Load and PV Output Forecasting Based on Dynamic Ensemble Deep Learning Model [J]. Renewable Energy Sources, 2020, 38(4): 458.
- [14] 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.
SHI Jiaqi, ZHANG Jianhua. Ultra Short-Term Photovoltaic Refined Forecasting Model Based on Deep Learning [J]. Electric Power Construction, 2017, 38(6): 28.
- [15] 张倩,马愿,李国丽,等. 频域分解和深度学习算法在短期负荷及光伏功率预测中的应用 [J]. 中国电机工程学报, 2019, 39(8): 2221.
ZHANG Qian, MA Yuan, LI Guoli, et al. Applications of Frequency Domain Decomposition and Deep Learning Algorithms in Short-term Load and Photovoltaic Power Forecasting [J]. Proceedings of the CSEE, 2019, 39(8): 2221.
- [16] 王坤峰,左旺孟,谭营,等. 生成式对抗网络: 从生成数据到创造智能 [J]. 自动化学报, 2018, 44(5): 769.
WANG Kunfeng, ZUO Wangmeng, TAN Ying, et al. Generative Adversarial Networks: From Generating Data to Creating Intelligence [J]. Acta Automatica Sinica, 2018, 44(5): 769.
- [17] 唐贤伦,杜一铭,刘雨微,等. 基于条件深度卷积生成对抗网络的图像识别方法 [J]. 自动化学报, 2018, 44(5): 855.
TANG Xianlun, DU Yiming, LIU Yuwei, et al. Image Recognition With Conditional Deep Convolutional Generative Adversarial Networks [J]. Acta Automatica Sinica, 2018, 44(5): 855.
- [18] 王守相,陈海文,潘志新,等. 采用改进生成式对抗网络的电力系统量测缺失数据重建方法 [J]. 中国电机工程学报, 2019, 39(1): 56.
WANG Shouxiang, CHEN Haiwen, PAN Zhixin, et al. A Reconstruction Method for Missing Data in Power System Measurement Using an Improved GAN [J]. Proceedings of the CSEE, 2019, 39(1): 56.
- [19] 刘云鹏,许自强,和家慧,等. 基于条件式 Wasserstein 生成对抗网络的电力变压器故障样本增强技术 [J]. 电网技术, 2020,

- 44(4): 1505.
- LIU Yunpeng, XU Ziqiang, HE Jiahui, et al. Data Augmentation Method for Power Transformer Fault Diagnosis Based on Conditional Wasserstein Generative Adversarial Network [J]. Power System Technology, 2020, 44(4): 1505.
- [20] 刘晓楠, 周介圭, 贾宏杰, 等. 基于非参数核密度估计与数值天气预报的风速预测修正方法 [J]. 电力自动化设备, 2017, 37(10): 15.
- LIU Xiaonan, ZHOU Jiegui, Jia Hongjie, et al. Correction Method of Wind Speed Prediction Based on Non-parametric Kernel Density Estimation and Numerical Weather Prediction [J]. Electric Automation Equipment, 2017, 37(10): 15.
- [21] LIU Y, YAN J, HAN S, et al. Uncertainty Analysis of Wind Power Prediction Based on Quantile Regression [C] //Power and Energy Engineering Conference, 2012: 1.
- [22] JOHNSON R W. An Introduction to the Bootstrap [J]. Teaching Statistics, 2001, 23(2): 49.
- [23] 杨锡运, 关文渊, 刘玉奇, 等. 基于粒子群优化的核极限学习机模型的风电功率区间预测方法 [J]. 中国电机工程学报, 2015, 35(S1): 146.
- YANG Xiyun, GUAN Wenyuan, LIU Yuqi, et al. Prediction Intervals Forecasts of Wind Power based on PSO-KELM [J]. Proceedings of the CSEE, 2015, 35(S1): 146.
- [24] NI Qiang, ZHUANG Shengxian, SHENG Hanming, et al. An Ensemble Prediction Intervals Approach for Short-Term PV Power Forecasting [J]. Solar Energy, 2017, 155: 1072.
- [25] QUAN Hao, SRINIVASAN D, KHOSRAVI A. Short-Term Load and Wind Power Forecasting Using Neural Network-Based Prediction Intervals [J]. IEEE Transactions on Neural Networks and Learning Systems, 2014, 25(2): 303.
- [26] Elia Group. Belgium Solar Power Generation data [DB/OL]. (2020-09-30) [2020-12-01]. <https://www.elia.be/>.
- [27] 陈文兵, 管正雄, 陈允杰. 基于条件生成式对抗网络的数据增强方法 [J]. 计算机应用, 2018, 38(11): 3305.
- CEHN Wenbing, GUAN Zhengxiong, CEHN Yunjie. Data Augmentation Method Based on Conditional Generative Adversarial Net Model [J]. Computer Application, 2018, 38(11): 3305.
- [28] 谭本东, 杨军, 赖秋频, 等. 基于改进 CGAN 的电力系统暂态稳定评估样本增强方法 [J]. 电力系统自动化, 2019, 43(1): 149.
- TAN Bendong, YANG Jun, LAI Qiupin, et al. Data Augment Method for Power System Transient Stability Assessment Based on Improved CGAN [J]. Automation of Electric Power Systems, 2019, 43(1): 149.

(编辑: 温泽宇)