

负荷水平约束条件下新能源出力场景生成新方法

许琴¹, 刘一鸣¹, 陈志刚¹, 庞学跃¹, 盛奥², 林济铿²

(1. 中国能源建设集团广东省电力设计研究院有限公司, 广州 510663; 2. 同济大学电子与信息工程学院, 上海 201804)

摘要: 场景序列是场景分析及优化的基础, 其精度直接影响相关的分析及优化计算的有效性。基于此, 提出了一种基于人工智能计及负荷水平约束的新能源出力场景生成新方法, 可明显提升所生成场景的精度, 并克服调度中心没有有效方法的困难。该新方法的基本过程为: 首先采取基于卡尔曼增益信息融合技术进行历史数据扩充, 实现了对应历史样本的有效扩充; 然后, 基于自组织映射网络将历史日负荷曲线进行聚类得到多个日负荷类型簇, 把与每个负荷簇内的日负荷曲线在同一天的新能源出力序列划分为相应的簇, 得到了负荷水平约束下的源荷信息; 最后, 构建基于马尔科夫链(Markov chain, MC)日类型转换关系的状态转移矩阵, 滚动抽样生成未来一段时间的日类型序列, 并结合利用各簇源荷数据训练好的生成对抗网络(generative adversarial network, GAN), 生成对应日类型下的 96 点源荷序列场景。算例证实了所提方法的有效性和先进性。

关键词: 场景系列生成; 人工智能; 数据扩展; 负荷水平约束; 马尔科夫链; 新能源

A New Method for Generating New Energy Output Scenarios with Load Level Constraints

XU Qin¹, LIU Yiming¹, CHEN Zhigang¹, PANG Xueyue¹, SHENG Ao², LIN Jikeng²

(1. China Energy Engineering Group Guangdong Electric Power Design Institute Co., Ltd., Guangzhou 510663, China;

2. School of Electronic and Information Engineering, Tongji University, Shanghai 200084, China)

Abstract: Scenarios sequence generation is the basis of scenario analysis and optimization problem, and its accuracy directly affects the effectiveness of related analysis and optimization calculation. Based on this, a new method of generating new energy output scenarios with load level constraints based on artificial intelligence is proposed, significantly improving the accuracy of the generated scenarios and overcoming the difficulties without effective methods faced by the dispatch center. The process of the new method is as follows: firstly, the historical data expansion strategy based on Kalman gain information fusion technology is adopted, realizing the effective data expansion of corresponding historical samples. Then, based on the self-organizing mapping network, the historical daily load curves are clustered to obtain multiple daily load type clusters, and each of the new energy output sequences on the same day as the daily load curves within each load cluster is corresponding classified into clusters, and the source load clusters with the specific load level constraint are achieved. Finally, a state transfer matrix based on Markov chain (MC) daily type transformation relationship is constructed, and then a sequence of daily types in the near future is generated by rolling sampling. And a generative adversarial network (GAN) already trained on each of source-load cluster data is applied to generate 96-point source-load scenarios sequences for the corresponding daily types. The numerical experiments verify the effectiveness and advancement of the proposed method.

Key words: scenarios sequences generation; artificial intelligence; data expansion; load level constraint; Markov chain; new energy

基金项目: 国家自然科学基金资助项目(51177107); 中国能源建设集团广东省电力设计研究院有限公司科技项目“新能源与负荷随机性特征及时间序列生成方法研究”(EV10271W)。

Foundation item: Supported by the National Natural Science Foundation of China (51177107); the Science and Technology Project of China Energy Engineering Group Guangdong Electric Power Design Institute Co., Ltd., “research on stochastic characteristics and time series generation method of new energy and load”(EV10271W).

0 引言

随着我国新型电力系统建设的推进,以风电光伏为主要代表的可再生新能源装机容量及出力占比在现有的基础上将继续大幅增加^[1-2]。新能源出力的波动性、间隙性将使得新型电力系统运行状态的不确定性特征更加凸显。有效且丰富的新能源出力场景序列蕴含着新能源出力的不确定性特征,是电网进行相关分析、优化调度及规划等的基础^[3-6]。相应地,如何准确地生成符合新能源出力随机特性的有效场景序列是一个备受关注的研究课题。

迄今的新能源出力场景生成方法大致可分为概率模型法、场景聚类法和人工智能法等3种。

1) 概率模型法。采用参数估计法或非参数估计法对新能源出力或灵活负荷历史数据进行概率建模,结合蒙特卡洛或拉丁超立方等采样方法生成源荷时序场景。文献[7]分别利用beta分布和经验分布拟合光照强度分布和负荷分布,经抽样生成场景。文献[8]使用weibull分布和beta分布构建风速和光照强度的概率密度函数,抽样生成风光出力场景。文献[9]提出了多个风电场风速联合分布预测方法,基于此通过抽样生成多风电场出力场景。文献[10-11]分别提出了基于核密度估计方法的横向和纵向点预测误差分布构建方法,进而抽样生成新能源出力场景。概率模型法采用显式概率模型拟合随机过程具有复杂度低和可解释性强的优点;但参数估计法需要新能源出力概率分布先验知识,非参数估计法需要有足够多的历史样本,否则其所构建模型的精度偏低。

2) 场景聚类法。基于相似度测度将具有相似特征的场景聚为一类,每一类的中心组成了典型场景集,基于典型场景集进行随机组合生成未来一段时间的场景。文献[12]采用改进的k-means算法综合考虑手肘法与聚类有效性指标得到具有代表性的典型场景。文献[13]基于Wasserstein距离消减相似度高的场景而得到典型场景。文献[14]采用改进的K-medoids算法对风光荷场景进行聚类而生成代表性的典型场景。文献[15-16]使用卷积自编码网络结合堆栈自编码网络对源荷历史数据进行降维,然后进行聚类得到典型场景。文献[17]利用改进卷积自编码器对数据降维,进而采用聚类结合RESCAE_KMEDOIDS算法生成场景。场景聚类法

在聚类最优性精度要求不高时,具有计算速度快的优点;但其缺点是若聚类精度高且样本数量大时,计算速度过慢。

3) 人工智能法。利用人工智能强大学习能力获得源荷数据所蕴含的分布特性,并基于此生成与之类似的源荷场景。文献[18-19]分别使用基于注意力机制的生成对抗网络和卷积生成网络获得气象因素与风光荷之间的关系生成相应场景。文献[20]将风速、辐照度、负荷聚类为多个数据集,利用各个数据集训练变分自编码器,获得风速、辐照度与风电、光伏出力及负荷需求的关系模型,进而生成风光出力及负荷需求场景。文献[21]利用残差卷积自编码器网络提取风光荷数据的特征,进而生成大量场景并利用k-medoids进行聚类获得典型场景。该方法在历史数据比较丰富时,可以获得精度很高的场景序列;若历史信息数量不足,其所生成的场景精度偏低。

综上所述,目前场景生成方法存在以下问题:1) 概率模型法存在需要不易获得的随机变量的先验分布知识,且要求样本必须非常丰富;2) 聚类方法在聚类精度高且样本大时计算速度过慢;3) 人工智能法需要足够多的历史数据信息,而相当数量的新能源电场是近些年才投入运行,这使得历史数据数量有限;4) 在运行调度中,为了分析及优化的需要,需要特定负荷需求下的新能源出力场景序列,迄今方法均未能满足此要求。

因此,本文提出了一种基于人工智能并计及负荷需求约束的新能源出力场景生成新方法。首先,提出基于卡尔曼增益信息融合技术的历史数据扩充策略,方便地得到了充分多的有效历史源荷数据,并克服了样本不足无法进行人工智能训练的缺点;然后将扩展后的历史日负荷需求曲线聚类为多个簇类型,同时与每个负荷簇内的日负荷曲线在同一天的新能源出力序列划分为相应的簇,得到与日负荷曲线簇相对应的新能源日出力序列簇,即得到了负荷需求约束下的源荷信息。进而,构建基于马尔科夫链日类型转换关系的状态转移矩阵,滚动抽样生成未来一段时间的日类型序列。最后,采用各簇源荷数据分别训练一个生成对抗网络(generative adversarial network, GAN)模型,获得基于各簇源荷数据的统计规律,并结合日类型序列及GAN生成对应日类型下的96点源荷序列,得到了负荷约

束下的源荷场景。

本文方法具有如下创新点。

1) 提出了基于卡尔曼增益信息融合技术的源荷数据扩充方法, 克服了历史数据信息不足无法进行有效的人工智能训练的缺点。

2) 提出了负荷水平约束下基于人工智能的新能源出力序列场景生成方法, 满足了调度中心生成特定负荷水平下的新能源出力场景序列的需求。

1 数据扩充策略及源荷数据聚类

1.1 数据扩充方法

人工智能模型的训练需要大量的数据样本, 而实际样本数据往往有限, 这影响了模型训练的学习效果。近些年新能源大量建成并运行, 但其所累积的历史样本有限, 特别在 2 000 年左右或 20 世纪建成的新能源电场, 因运行经验缺乏和各种主观及客观原因, 在电场建成的初期甚至相当长的一段时间内所累积的数据信息很不完整, 使得其有效历史样本数据也有限。由于上述原因, 当把源荷数据样本用于人工智能模型学习及训练时, 存在着样本数据缺乏或不足而使得人工智能模型训练效果不佳的问题。基于此提出一种基于卡尔曼增益信息融合的新能源出力及负荷需求样本扩充新策略。该策略的基本思想是利用相近的历史记录信息, 从横纵两个维度分别进行随机组合, 然后再利用卡尔曼增益进行横纵信息融合, 而得到可信且足够的扩充样本。具体策略如图 1 所示。

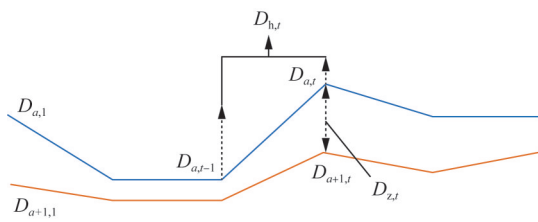


图 1 数据扩充示意图

Fig. 1 Illustration of data expansion

图 1 给出了样本数据示意图, 红色和蓝色线分别代表两天相近的样本数据, $D_{z,t}$ 为由相邻两天样本数据纵向扩充得到的扩充数据; $D_{h,t}$ 为由同一天相邻时刻样本数据横向扩充得到的扩充数据, 计算公式为:

$$D_{z,t} = D_{a,t} + r_{\text{and},z} \times (D_{a+1,t} - D_{a,t}) \quad (1)$$

$$D_{h,t} = D_{a,t-1} + r_{\text{and},h} \times (D_{a,t} - D_{a,t-1}) \quad (2)$$

式中: $D_{a,t-1}$ 和 $D_{a,t}$ 为 1 d 中相邻时刻的样本数据; $D_{a,t}$ 和 $D_{a+1,t}$ 为相邻日同一时刻的样本数据; $t=1, 2, 3, \dots, 96$; $a=1, 2, 3, \dots, 364$; 其中, 当 $t=1$ 时, $D_{h,t}=D_{a,t}$; $r_{\text{and},z}$ 和 $r_{\text{and},h}$ 为 $(0, 1)$ 之间随机数。

在得到横纵向的扩充数据后, 使用基于卡尔曼增益系数的数据融合方法将横向和纵向得到的扩充数据进行融合, 以提升扩充数据的精度, 即:

$$D_{r,t} = D_{h,t} + b_r (D_{z,t} - D_{h,t}) \quad (3)$$

$$b_r = \frac{\sigma^2(D_{h,t})}{\sigma^2(D_{h,t}) + \sigma^2(D_{z,t})} \quad (4)$$

$$\sigma^2(D_{h,t}) = \frac{\sum_{t=1}^{96} (D_{h,t} - E(D_{h,t}))^2}{96} \quad (5)$$

$$\sigma^2(D_{z,t}) = \frac{\sum_{t=1}^{96} (D_{z,t} - E(D_{z,t}))^2}{96} \quad (6)$$

$$E(D_{h,t}) = \frac{1}{96} \sum_{t=1}^{96} D_{h,t} \quad (7)$$

$$E(D_{z,t}) = \frac{1}{96} \sum_{t=1}^{96} D_{z,t} \quad (8)$$

式中: $D_{r,t}$ 为融合后的扩充数据; b_r 为卡尔曼增益系数; $\sigma^2(\cdot)$ 和 $E(\cdot)$ 分别为求取方差和均值。需要指出的是, 当对于负荷进行扩充时, 需要对于对应的新能源出力数据进行类似的扩充, 从而使得每一日负荷数据均有对应的新能源出力数据。

源荷序列作为典型的时间序列, 其在横向相邻时刻及纵向每日同一时刻都有着典型的变换趋势, 相对于只考虑横向扩充数据 D_h 或纵向扩充数据 D_z , 扩充数据 D_r 因同时考虑了纵向日与日之间样本间隔 24 h 的变化关系与横向相邻时刻之间样本间隔 15 min 的变化关系, 使得扩充数据所蕴含的随机规律更符合历史样本本身的统计规律, 算例部分 GAN 的收敛结果以及使用扩充样本训练后生成源荷序列的统计特性结果说明该扩充算法的有效性。

1.2 基于自组织映射网络的源荷数据聚类

在电力电量平衡及运行调度中, 需要测算特定负荷需求下的源荷平衡, 如高负荷水平、低负荷水平等。但进行高负荷水平条件下的源荷平衡时, 传统的场景生成算法大多利用历史所有新能源出力场景进行相应的未来场景生成, 所生成的场景不适合于高负荷水平下的相应计算。本文方法将同一日新能源数据直接归为日负荷聚类中, 相当于统计了特定负荷水平下的所有新能源出力特性, 而排除了其它负荷水平下的新能源出力特性, 基于此所生成的新能源出力场景更加符合特定负荷水平下的新能源

出力特性,相应也更准确。

本文采用自组织映射 (self-organizing map, SOM) 聚类算法将数据扩充之后的日负荷序列聚类为数个极端日和数个典型日。同时,对于每一种日负荷类型,将与其同一日的新能源数据也与负荷归为一类,即只对历史日负荷曲线采用聚类算法聚为多簇,而与日负荷属于同一天的新能源数据直接归为日负荷所在簇,从而得到了与日负荷类型相对应的源荷场景簇,该源荷数据簇即为不同负荷水平约束下的源荷历史数据聚类。

若直接对于每天 96 点 (15 min 分辨率) 日负荷曲线进行聚类,高维样本之间的欧氏距离存在彼此接近倾向,样本点之间的距离关系存在一定程度弱化倾向,从而使得基于欧氏距离聚类结果有效性相对降低^[22-23]。为了提高聚类有效性,本文采取策略对日负荷需求曲线先进行降维,然后再进行聚类,具体是采用日负荷特征指标:日负荷率、日峰谷差率、日最大负荷、日负荷均值等 4 个特征指标代替日负荷序列,从而实现了日负荷曲线的大幅度降维。特征指标的计算公式为:

$$\alpha = \frac{1}{T} \sum_{t=1}^T l(t) / \max \{ l(t) \} \quad (9)$$

$$\beta = \frac{\max \{ l(t) \} - \min \{ l(t) \}}{\max \{ l(t) \}} \quad (10)$$

$$\gamma = \max \{ l(t) \} \quad (11)$$

$$\delta = \frac{1}{T} \sum_{t=1}^T l(t) \quad (12)$$

式中: α 为日负荷率; $\max \{ \}$ 表示最大值; $\min \{ \}$ 表示最小值; $l(t)$ 为 t 时刻负荷; T 为每日的全部时段数量; β 为日峰谷差率; γ 为日最大负荷; δ 为日负荷均值。

在得到每日负荷特征指标后,采用这些指标及 SOM 聚类算法进行历史日负荷曲线的聚类。SOM 是一种无监督的人工神经网络聚类算法,运用竞争学习策略,依靠神经元之间互相竞争逐步优化网络,且使用近邻关系函数来维持输入空间的拓扑结构。相较于如 k-means 等聚类算法, SOM 具有自动寻找最优聚类数,可根据样本相似性自动将其分类,无需事先知道样本标签的优点。计算步骤为:

步骤 1: 对数据标准化处理,依据规则选取竞争层的神经元节点数目并随机初始化权重,令 $t=1$;

步骤 2: 读取一个输入样本 x_i , 计算与竞争层

节点的欧式距离,选取距离最小的节点作为优胜节点;

步骤 3: 依据式 (13) 更新节点的权重,在更新权重的过程中,使用式 (14) 高斯函数作为更新权重来确定优胜节点及其近邻节点的权重更新幅度;使用式 (15) 随迭代次数自动更新的自适应学习率提升学习精度。

$$W_{s+1} = W_s + \eta g(x_i - W_s) \quad (13)$$

$$g(m, n) = e^{-\frac{(c_u - m)^2 + (c_v - n)^2}{2\tau^2}} \quad (14)$$

$$\eta = \frac{\eta_0}{1 + \frac{t}{T/2}} \quad (15)$$

式中: W_s 和 W_{s+1} 分别为更新前和更新后的节点权重; g 为更新权重; (m, n) 为节点在网格中的坐标; (c_u, c_v) 为优胜节点在网格中的坐标; τ 为邻域范围约束变量; η 为自适应学习率; η_0 为初始学习率; t 为当前迭代次数; T 为总迭代次数。

步骤 4: 返回步骤 2, 循环直至读取完所有样本完成一次迭代后令 $t=t+1$; 当 $t=T$ 时结束迭代。

在无监督聚类算法中,模型通过样本之间的距离自动将相似样本分为一类,会将极端场景归类到典型日中,因此在对日负荷进行聚类之前,需要先筛选出极端负荷场景,本文选取日负荷均值大于 $\delta_{\max}$ 以及日负荷均值低于 $\delta_{\min}$ 分别作为高负荷日类型簇和低负荷日类型簇, $\delta_{\max}$ 和 $\delta_{\min}$ 由式 (16) — (17) 确定。余下的数据采用上述聚类方法聚类为不同的典型日。

$$\delta_{\max} = \text{mean}(\delta) + 1.2\text{std}(\delta) \quad (16)$$

$$\delta_{\min} = \text{mean}(\delta) - 1.2\text{std}(\delta) \quad (17)$$

式中: $\text{mean}(\cdot)$ 为求取均值函数; $\text{std}(\cdot)$ 为求取标准差函数; δ 为日负荷均值。

在得到负荷聚类结果之后,对于每一簇中的日负荷曲线,把与其同一日的新能源出力曲线也相应划分为一个簇,就得到了与每一负荷簇相对应的新能源出力簇,由于每一负荷簇代表相应的负荷水平,也就得到了负荷水平约束下的源荷历史数据的聚类。

2 基于马尔科夫链的日类型序列生成

在完成日类型聚类之后,各日类型之间的转换关系可以近似地视为马尔可夫随机变化过程,因此本文使用马氏链描述日类型转换关系,并构建状态

转移矩阵 $\mathbf{P}$ 表示日类型之间的随机变化规律, 基于 $\mathbf{P}$ 通过滚动抽样生成未来一段时间的日类型序列。

$\mathbf{P}$ 是一个 $K \times K$ 的矩阵, 其中, K 为日类型数量, $\mathbf{P}$ 中的元素 p_{kl} 为当前状态 S_c 为日类型 k , 下一日状态 S_{c+1} 转移到日类型 l 的概率, S_c 表示 c 日的状态。

$$p_{kl} = G(S_{c+1} = l | S_c = k) \quad (18)$$

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1K} \\ p_{21} & p_{22} & \cdots & p_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ p_{K1} & p_{K2} & \cdots & p_{KK} \end{bmatrix} \quad (19)$$

式中: G 为从状态 S_c 至状态 S_{c+1} 的转移概率; 元素 p_{kl} 为日类型状态由状态 k 转移到状态 l 的概率 ($k, l = 1, 2, \dots, K$), 其值为:

$$p_{kl} = \frac{n_{kl}}{\sum_{v=1}^K n_{kv}} \quad (20)$$

式中: n_{kl} 为历史数据中由日类型 k 转移到日类型 l 的转移频数; n_{kv} 为历史数据中由日类型 k 转移到日类型 v 的转移频数。

利用状态转移矩阵 $\mathbf{P}$ 并采用 Markov 链蒙特卡洛 (Markov chain Monte Carlo, MCMC) 算法抽样, 即可生成未来一段时间的日类型序列^[24-25]。

3 基于生成对抗网络的场景生成

在得到了负荷水平约束下的源荷历史数据聚类及日类型序列, 本文进而提出基于生成对抗网络的场景生成方法。基本思想为: 利用与每一聚类相对应的每一子集数据分别训练 GAN 模型; 使用与所生成的日类型序列中每一日类型对应的 GAN 模型生成相应日类型下的源荷序列, 从而得到未来一段时间负荷水平约束下的源荷序列场景。

3.1 GAN 网络结构及训练

GAN 由生成器和判别器两部分组成。随机噪声输入生成器 G 中生成数据, 将所生成数据 (称为生成样本) 和真实样本输入到判别器 D 中, 判断数据的真伪。随着训练学习, 生成器逐渐学习生成逼真的样本, 而判别器逐渐提高对生成样本的识别能力。最终, 生成器生成的样本足够逼真, 以至于判别器无法准确区分真实样本和生成样本, 相应训练学习结束。

3.1.1 GAN 网络的结构

图 2 给出了 GAN 的基本结构图, 生成器输入随机噪声生成样本; 判别器的输入为真实样本和生成样本, 输出为样本为真的概率即损失函数, 通过该损失函数更新生成器和判别器网络权重。

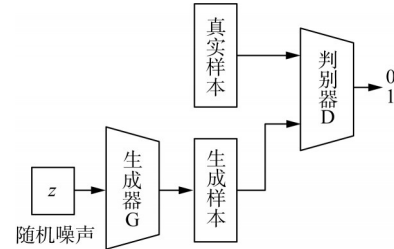


图 2 GAN 结构图

Fig. 2 Structure diagram of GAN

GAN 模型的目标函数如式 (21) 所示, 其中 x 代表真实样本, z 代表随机噪声, $G(\cdot)$ 表示生成器函数, $D(\cdot)$ 表示判别器函数。在训练学习过程中, 首先固定 $D(x)$ 的梯度, 更新 $D(x)$ 的参数使目标函数最大化, 使得 $D(x)$ 趋于 1, $D(G(z))$ 趋于 0, 即希望对于真实样本 x 判别器识别为真, 对于生成数据 $G(z)$ 判别器能识别为假; 然后固定 $D(x)$ 的梯度, 更新 $D(x)$ 的参数使目标函数最小, 使得 $D(G(z))$ 趋于 1, 即希望生成器生成的样本 $G(z)$ 能够让判别器识别为真。

$$\min_{G(z)} \max_{D(x)} V(D(x), G(z)) = E_{x \sim p_{\text{data}}} [\log D(x)] + E_{z \sim p_u} [\log (1 - D(G(z)))] \quad (21)$$

式中: $V(D(x), G(z))$ 为目标函数; x 为真实样本 p_{data} 中的值; z 为服从正态分布 p_u 的随机噪声。

图 3 给出了 GAN 网络结构。以判别器第一层 (Layer1) 为例, 其中 3 和 96 代表卷积之后的数据结构, 32 代表通道数。参数设置参考文献 [25]。

新能源出力和负荷需求曲线各自之间有一定的横向相关性, 即当下时段与相近时段有较强的相关性, 与时间间隔较长的时段相关性较弱。同时源荷序列之间有一定纵向相关性, 针对数据特性结构, 因此在判别器部分使用 3×3 卷积核提取输入信息特征, 对输入数据四周补 0 并使用步长为 1 的卷积核进行卷积处理, 可维持输入输出信息尺寸不变, 以便对输入数据进行多层卷积处理, 提高网络表达能力, 具体结构如图 4 所示。

图 4 中, 黑色实线框所在层为输入层, 其中 w 、

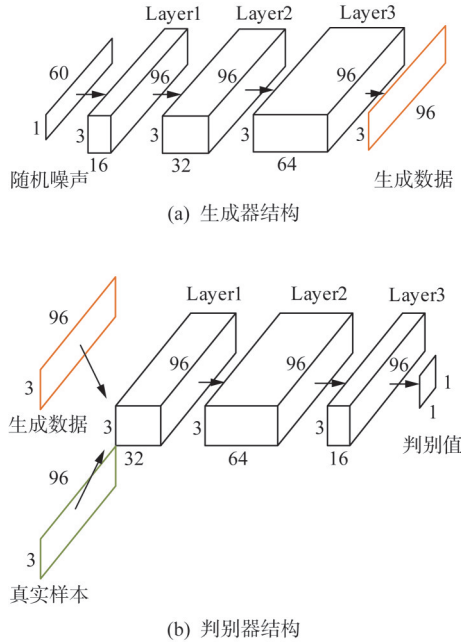


图3 GAN网络结构

Fig. 3 Network structure of GAN

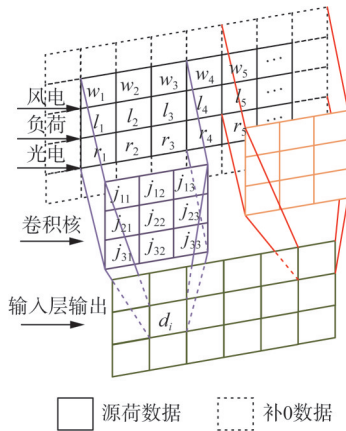


图4 基于卷积核的输入层信息及计算示意图

Fig. 4 Information and computational diagram of the input layer based on convolutional kernels

l 、 r 为拼接后的源荷数据，其四周虚线框为补0数据； j_{ij} ($i, j = 1, 2, 3$)为输入层卷积核权重，使用 3×3 卷积核对输入数据进行卷积计算得到结果，并把结果(即为图4中的 d_i)直接输入到下一层网络中， d_i 的具体计算公式为：

$$d_i = w_1 \times j_{11} + w_2 \times j_{12} + \cdots + l_3 \times j_{23} + \cdots + r_3 \times j_{33} \quad (22)$$

因此，图4清晰地描述了输入层的输入信息，卷积计算过程及输出。生成对抗网络中后续的每一层计算过程均与图4类似。

3.1.2 GAN网络的训练

由于JS(Jensen-Shannon)散度在生成样本和真实样本的分布完全没有重叠时恒为常数，使得基于JS散度的损失函数会出现训练困难以及训练崩溃的问题。Wasserstein距离在两个分布完全没有重叠时也可以有效衡量两个分布之间的距离，因此本文采用Wasserstein距离作为判别器损失函数。具体训练步骤为如下。

步骤1：生成符合正态分布的随机噪声输入到生成器得到生成样本；

步骤2：将生成样本和真实样本输入判别器，判别器输出对真实样本和生成样本的判别值；

步骤3：计算生成器与判别器的损失函数，分别更新生成器与判别器的网络权重，若判别器损失函数未收敛到平稳，返回步骤1；

步骤4：训练结束，使用生成器模型，输入随机噪声可生成符合历史数据统计规律的源荷场景。

3.2 负荷水平约束的源荷场景生成步骤

负荷水平约束的源荷场景生成的步骤如下。

步骤1：输入历史源荷数据，进行标准化处理；

步骤2：若数据不够，采用第1.1节的方法进行数据的扩展；

步骤3：采用1.2节的聚类方法将负荷数据进行聚类，同时也将与每一簇中负荷位于同一日的新能源数据也聚为一簇；

步骤4：依据历史数据每日日类型构造状态转移矩阵，并抽样滚动生成未来一段时间的日类型序列；

步骤5：使用每一簇的源荷数据分别训练一个GAN模型；

步骤6：根据所生成的未来日类型序列，使用日类型对应的GAN模型生成日类型序列中每一日类型相对于的96点源荷场景序列。

上述步骤的流程图如图5所示。

4 算例分析

利用本文算法对广东省某地区一年的风光荷历史数据(间隔15 min做一个记录)进行了计算，生成负荷水平约束的源荷场景。

4.1 聚类有效性验证

图6给出了在剔除极端日后的日负荷聚类后SOM输出层各神经元节点之间的距离图。

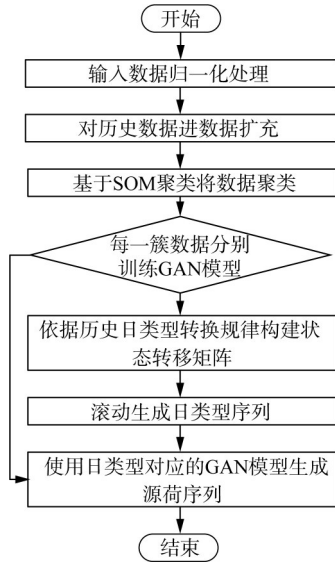


图5 负荷水平约束的源荷场景生成流程图

Fig. 5 Flow chart of source-load scenarios generation with load level constraints

在 SOM 聚类结果图中, 可以视为样本点被归类为到神经元中, 图6中每个六边形为一个神经元节点, 颜色越接近冷色系表示该神经元与附近神经元之间的距离越小, 即左上部分和右下部分冷色系之中的样本点分别视为一簇; 颜色越接近暖色系表示该神经元与附近神经元之间的距离越大, 即以暖

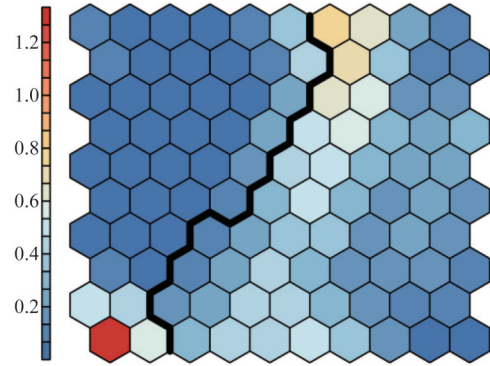


图6 SOM网络聚类结果图

Fig. 6 Results of SOM network clustering

色区域为间隔, 将左上和右下划分开分别为典型日1和典型日2, 再加上事先划分出来的两类极端日, 总体分类结果如图7所示。

在图7中, 高负荷和低负荷聚类结果由式(16)一(17)划分得出, 典型日1和典型日2分别为图5中被划分到暖色区域两侧神经元上的负荷日。可以观察出簇内之间负荷曲线相似, 不同簇之间负荷曲线有明显差异。高负荷簇中负荷水平明显偏高, 且有着较大的峰谷差率, 多对应夏季气温较高的工作日; 低负荷簇中负荷水平明显偏低, 曲线较为平缓, 多对应于节假日; 典型日1簇和典型日2簇则

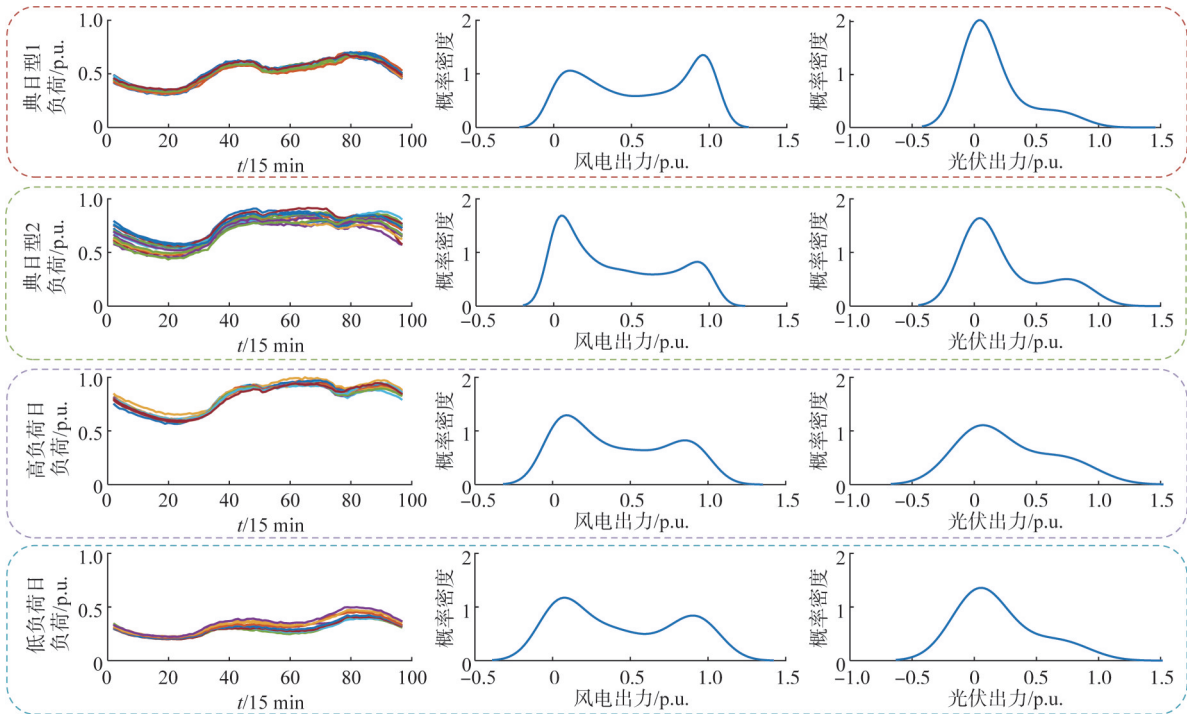


图7 负荷日类型簇划分结果

Fig. 7 Results of cluster division of load day types

对应于典型的 2 峰谷和 3 峰谷正常负荷水平。因此,该结果验证了本文所用聚类方法可以根据日负荷曲线特征,有效地将历史日负荷曲线划分为不同日类型簇,相应得到不同的日类型。

图 7 中也给出了每一负荷簇中相应的源荷出力概率密度函数,可以观察出在典型负荷 1 簇中,风电出力在 1 p. u. 附近的概率密度较高,光伏出力在 0 p. u. 附近概率密度较高,即风电高出力场景较多,光伏低出力场景较多。在典型负荷 2 中,风电与光伏出力在 0 p. u. 附近的概率密度都较高,即风电和光伏的低出力场景都较多。高负荷与低负荷簇中的新能源出力相似但与典型负荷日中的新能源出力特性有一定的差距,如相较于典型日中,光伏出力在 0 p. u. 附近的概率密度偏低,即低出力场景较少。

图 7 的结果是通过对于负荷曲线进行极端日划分及聚类之后得到,每一类负荷曲线代表了相应的负荷水平,基于此的新能源出力簇即为相应负荷水平下的分类及聚类结果。

4.2 数据扩展及 GAN 网络训练收敛性验证

分别采用样本扩充数据与原始样本数据,学习算法为 RMSprop 梯度下降算法,学习率为 0.000 2,批尺寸为 64,对于 GAN 网络分别进行训练。图 8 给出了数据集簇 3(2 012 组)训练时两组训练数据的判别器损失函数变化曲线。可以看出数据扩充后的损失函数收敛速度更快,收敛效果更好,说明了数据扩充的有效性。

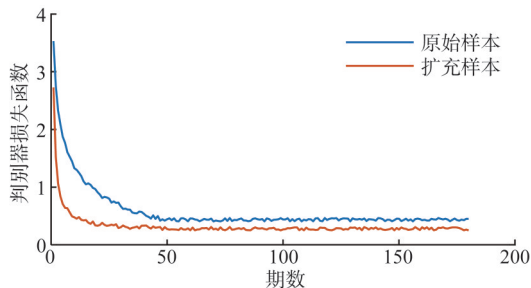


图 8 判别器损失函数变化曲线

Fig. 8 Curves of discriminator loss function

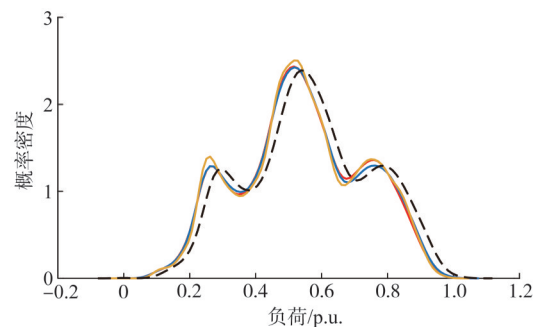
4.3 生成的场景序列有效性验证

分别利用本文所提出基于样本扩充之后训练 GAN 的算法(简称本文算法)、基于文中所述 GAN 模型但使用扩充前数据训练的算法(简称原始样本算法)及对比算法^[21]生成一年的源荷场景进行对比,

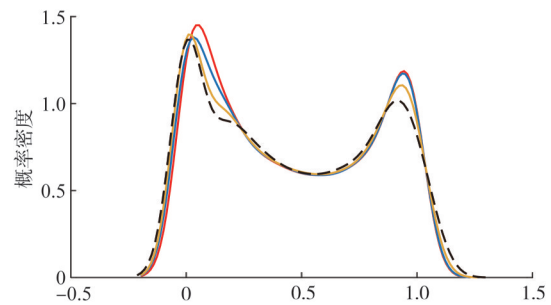
以证明本文算法的有效性与先进性。高质量的生成场景序列应具备如下特征:1)生成场景的分布特性应与历史场景的分布特性相似;2)生成场景之间的相关性应与历史场景之间的相似;3)生成场景的波动特性应与历史场景的波动特性相似。

4.3.1 概率分布特性

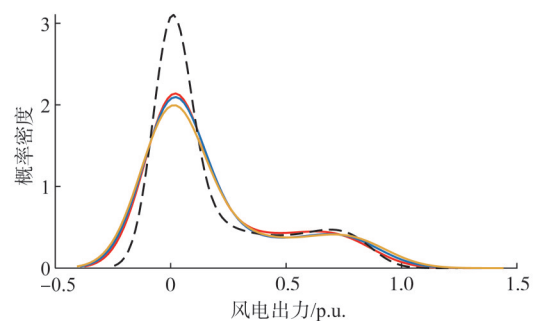
图 9 给出了不同方法所生成一年的风光荷场景序列及历史数据场景的统计特性对比图。可以看出,本文方法所生成的源荷场景统计特性与历史数据的统计特性更为接近,证实了本文算法的有效性。



(a) 负荷概率密度



(b) 风电概率密度



(c) 光伏出力概率密度

— 历史数据 — 使用扩充数据训练
- - - 对比算法 — 使用原始样本训练

图 9 不同方法生成场景的统计特性对比

Fig. 9 Comparison of statistical features of scenarios generated by different methods

为进一步验证本文方法的优越性,使用残差平方和(residual sum of square, RSS)来定量比较生成效果,计算如下。

$$R_{ss} = \sum_{h=1}^H (y_h - y'_h)^2 \quad (23)$$

式中: R_{ss} 为残差平方和; y_h 为生成序列的概率密度函数; y'_h 为历史出力序列的概率密度函数; H 为数据长度。

表3给出了本文方法、对比方法以及使用原始样本算法分别生成源荷序列的概率密度函数的残差对比结果。可以明显看出本文方法相对于其他两个方法的RSS指标均明显较小,具有明显的优势,即利用扩充样本训练GAN生成的数据更接近真实样本的统计规律,说明了数据扩充的有效性。

表3 生成场景与历史场景概率密度函数的RSS对比

Tab. 3 RSS comparison of the probability density functions of scenarios generation and historical data

方法	R_{ss}		
	风	光	荷
本文方法	0.83	1.83	0.95
对比方法	2.36	3.11	5.28
原始样本算法	1.21	2.05	1.27

将生成的源荷场景序列中各种负荷日类型分别挑选出来并分别统计其新能源分布特性。如图10给出了生成的源荷场景中不同负荷水平下新能源出力的分布特性,可以看出,本文方法所生成场景在不同负荷水平条件下的新能源分布特性依然与历史样本相近,即本文方法生成的源荷场景在总体分布特性接近历史样本的同时,考虑了各负荷水平条件下

的新能源分布特性也能够再现历史样本的统计规律。从而为更精准中长期电力电量平衡测算提供参考。

4.3.2 风光荷相关性

风光荷间存在耦合性,在生成源荷场景时生成源荷场景的源荷特性应与历史场景接近。Pearson相关系数是衡量两个连续变量之间线性相关强度和方向的统计量,能准确描述连续变量之间的线性变换趋势,在源荷序列中的具体表现是此增彼减、此减彼增的一种变化趋势。因此使用Pearson相关系数描述生成样本之间的相关性。表4给出了基于不同方法的生成场景源荷序列与历史源荷序列之间的相关系数,可以看出,本文方法所生成源荷系列之间的相关系数与历史源荷序列之间的相关系数,相对于其他方法而言更为接近,从而证实了本文算法生成序列能够很好地再现源荷之间的相关性。

表4 不同方法生成的源荷序列与历史源荷序列之间 Pearson 相关系数

Tab. 4 Pearson correlation coefficients between source-load scenarios series generated by different methods and that of historical data

算法	Pearson 系数		
	风光	风荷	光荷
历史数据	-0.295	-0.192	0.337
本文算法	-0.278	-0.187	0.345
原始样本算法	-0.322	-0.188	0.307
对比算法	-0.239	-0.143	0.298

4.3.3 波动特性

使用纵向每一时刻记录点的标准差 σ_j 描述源荷场景的波动特性。

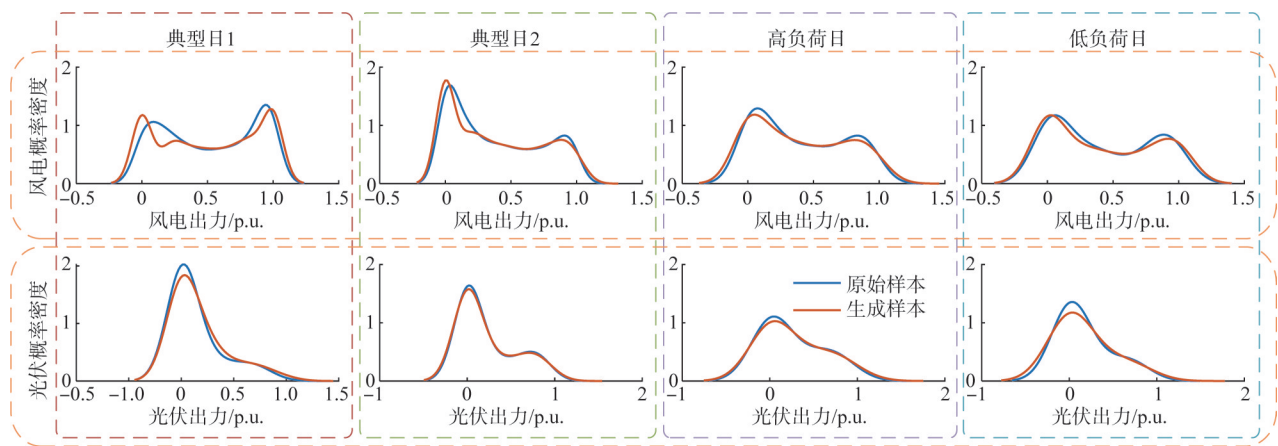


图10 生成场景中各负荷日下的新能源分布特性

Fig. 10 New energy distribution characteristics under each load day in the generation scenario

$$\sigma_j(t) = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_{i,j}(t) - \frac{\sum_{i=1}^N X_{i,j}(t)}{N})^2} \quad (24)$$

式中： $\sigma_j(t)$ 为源荷 t 时段的标准差， $j=1,2,3$ 分别表示风电、光伏、负荷； $t=1,2,\dots,T$ 为时段总数； $X_{i,j}(t)$ 为 i 日 t 时段源荷标么值； N 为场景总天数。

图 11 给出了不同方法生成场景的各个时间点的标准差曲线。

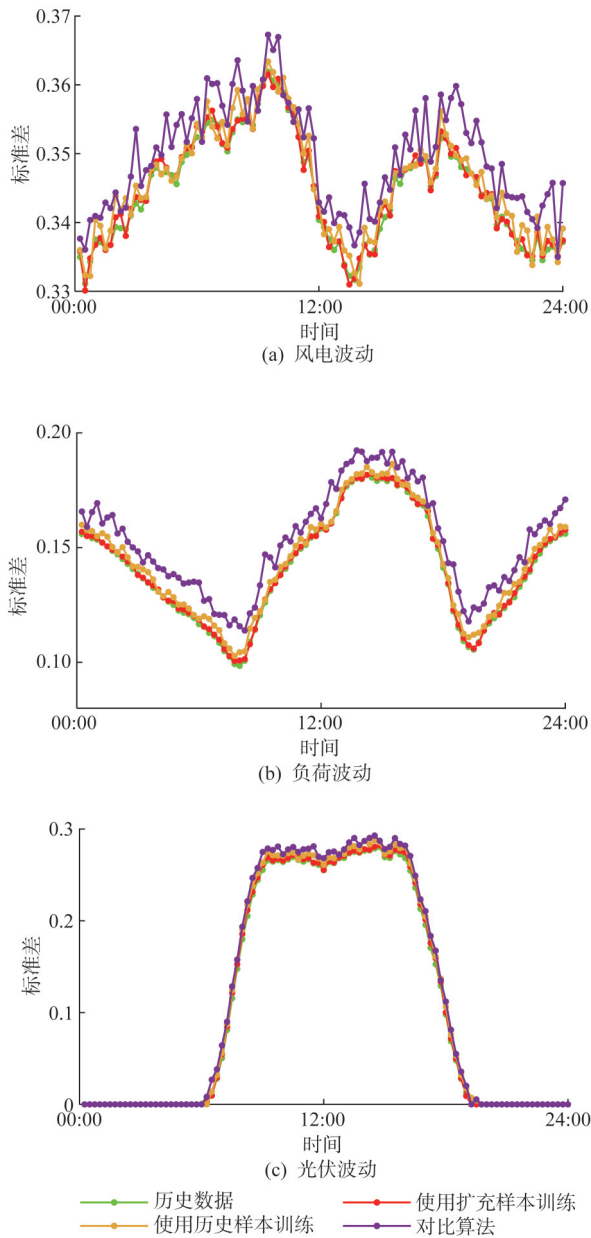


图 11 不同方法生成源荷场景的标准差

Fig. 11 Standard deviation of source-load scenarios generated by different methods and that of historical data

可以看出，本文方法相较于其他方法无论是在贴合度还是在趋势上都与历史数据更加接近。同时，使用样本扩充数据训练网络的标准差曲线相比直接使用历史小样本数据训练的标准差曲线，更加接近于历史样本的标准差，说明文中扩充数据方法的有效性以及本文算法的有效性。

5 结语

本文提出一种基于人工智能并计及负荷水平约束的新能源出力场景生成新方法。该方法具有如下特点：1) 提出基于卡尔曼增益信息融合技术的历史数据扩充策略，以得到充分多的有效历史源荷数据而克服样本不足无法进行人工智能训练的缺点；2) 通过对于负荷曲线进行分类及聚类，并把与每簇中同一天的新能源曲线也划分为一簇或类，由于每一负荷曲线簇代表了相应的负荷水平，相应得到了负荷水平约束下的源荷信息及生成相应的源荷场景。

算例结果表明，与现有方法所生成场景只能在总体特性接近于历史真实样本特性的相比，本文方法所生成场景均能在特定负荷需求下，接近于历史样本的特性，更好地满足了调度中心为了安全校核等，而需要生成特定负荷水平下的源荷场景的需求。限于气象数据的不完整性，文中只使用了源荷数据，在后续的研究中，将进一步考虑不同气象对源荷数据特征的影响。

参考文献

- [1] 卓振宇,张宁,谢小荣,等.高比例可再生能源电力系统关键技术及发展挑战[J].电力系统自动化,2021,45(9):171-191.
ZHUO Zhenyu, ZHANG Ning, XIE Xiaorong, et al. Key technologies and development challenges of power system with high proportion of renewable energy [J]. Automation of Electric Power Systems, 2021, 45 (9): 171 - 191.
- [2] 程杉,张芮嘉,钟仕凌,等.计及需求响应柔性调节的可再生能源优化配置[J].电力系统及其自动化学报,2023,35(2):94-102.
CHENG Shan, ZHANG Ruijia, ZHONG Shiling, et al. Optimal allocation of renewable energy with flexible regulation of demand response [J]. Journal of Electric Power Systems and Automation, 2023, 35 (2): 94 - 102.
- [3] LEE D, BALDICK R. Load and wind power scenario generation through the generalized dynamic factor model [J]. IEEE Transactions on Power Systems, 2017, 32(1):400-410.
- [4] 叶琳浩,申展,许峰,等.考虑光伏不确定性和时序相关性的分布鲁棒光储协同优化配置方法[J].南方电网技术,2023,

- 17(4): 132-143.
- YE Linhao, SHEN Zhan, XU Feng, et al. Distributionally robust collaborative optimal allocation method of photovoltaic and energy storage considering photovoltaic uncertainty and temporal correlation [J]. Southern Power System Technology, 2023, 17(4): 132-143.
- [5] 薛禹胜,雷兴,薛峰,等.关于风电不确定性对电力系统影响的评述[J].中国电机工程学报,2014,34(29):5029-5040.
- XUE Yusheng, LEI Xing, XUE Feng, et al. Review on the influence of wind power uncertainty on power system [J]. Proceedings of the CSEE, 2014, 34(29): 5029-5040.
- [6] 黎静华,孙海顺,文劲宇,等.生成风电功率时间序列场景的双向优化技术[J].中国电机工程学报,2014,34(16):2544-2551.
- LI Jinghua, SUN Haishun, WEN Jinyu, et al. Bidirectional optimization technique for generating wind power time series scenario [J]. Proceedings of the CSEE, 2014, 34(16): 2544-2551.
- [7] 方华亮,李大虎,彭辉,等.基于“互联网+”的分散式太阳能规划方法[J].中国电机工程学报,2017,37(5):1316-1324.
- FANG Hualiang, LI Dahui, PENG Hui, et al. Distributed solar energy planning method based on "Internet+" [J]. Proceedings of the CSEE, 2017, 37(5): 1316-1324.
- [8] 丁苏阳,丁壮壮,林湘宁,等.基于概率图像灰度比对算法的远洋海岛可再生能源机组配置方案优选策略研究[J].中国电机工程学报,2018,38(19):5653-5667.
- DING Suyang, DING Zhuangzhuang, LIN Xiangning, et al. Research on optimal strategy of allocation scheme for offshore renewable energy units based on probabilistic image gray-scale comparison algorithm [J]. Proceedings of the CSEE, 2018, 38(19): 5653-5667.
- [9] 李湃,管晓宏,吴江.基于大气动力模型的多风电场出力场景生成方法[J].中国电机工程学报,2015,35(18):4581-4590.
- LI Pai, GUAN Xiaohong, WU Jiang. Output scenario generation method of multiple wind farms based on atmospheric dynamic model [J]. Proceedings of the CSEE, 2015, 35(18): 4581-4590.
- [10] 张紫菁,张芳,姚文鹏.基于风电出力概率预测模型的源荷储日前概率优化调度[J].电力自动化设备,2022,42(7):190-197.
- ZHANG Zijiang, ZHANG Fang, YAO Wenpeng. Probabilistic optimal scheduling of source load and storage day based on probabilistic forecasting model of wind power output [J]. Electric Power Automation Equipment, 2022, 42(7): 190-197.
- [11] 赵鉴,袁渤巽,于浩,等.基于风光不确定性和相关性的场景生成[J].能源与节能,2021(8):8-12.
- ZHAO Jian, YUAN Boxun, YU Hao, et al. Scene generation based on landscape uncertainty and correlation [J]. Energy and Energy Conservation, 2021(8): 8-12.
- [12] 唐锦,张书怡,吴秋伟,等.基于Copula函数与等概率逆变换的风电出力场景生成方法[J].电力工程技术,2021,40(6):86-94.
- TANG Jin, ZHANG Shuyi, WU Qiuwei, et al. Wind power output scenario generation method based on Copula function and equal probability inverse transformation [J]. Electric Power Engineering Technology, 2021, 40(6): 86-94.
- [13] 王群,董文略,杨莉.基于Wasserstein距离和改进K-medoids聚类的风电/光伏经典场景集生成算法[J].中国电机工程学报,2015,35(11):2654-2661.
- WANG Qun, DONG Wenlue, YANG Li. Wind power/photovoltaic classical scene set generation algorithm based on Wasserstein distance and improved K-medoids clustering [J]. Proceedings of the CSEE, 2015, 35(11): 2654-2661.
- [14] 王华勇,傅强,徐昱,等.基于场景削减的网源协同规划研究[J].自动化仪表,2021,42(5):69-74.
- WANG Huayong, FU Qiang, XU Yu, et al. Research on network source collaborative planning based on scene reduction [J]. Process Automation Instrumentation, 2021, 42(5): 69-74.
- [15] 石亮缘,周任军,张武军,等.采用深度学习和多维模糊C均值聚类的负荷分类方法[J].电力系统及其自动化学报,2019,31(7):43-50.
- SHI Liangyuan, ZHOU Renjun, ZHANG Wujun, et al. Load classification method based on deep learning and multi-dimensional fuzzy C-means clustering [J]. Proceedings of the CSU-EPSA, 2019, 31(7): 43-50.
- [16] 杨晶显,刘俊勇,韩晓言,等.基于深度嵌入聚类的水光荷不确定性源场景生成方法[J].中国电机工程学报,2020,40(22):7296-7306.
- YANG Jingxian, LIU Junyong, HAN Xiaoyan, et al. Source scene generation method of water-light load uncertainty based on deep embedded clustering [J]. Proceedings of the CSEE, 2020, 40(22): 7296-7306.
- [17] 彭雨箏,李晓露,李聪利,等.基于残差卷积自编码的风光荷场景生成方法[J].电力建设,2021,42(8):10-17.
- PENG Yuzheng, LI Xiaolu, LI Congli, et al. Scene-load scene generation method based on residual convolution self-coding [J]. Electric Power Construction, 2021, 42(8): 10-17.
- [18] 李辉,任洲洋,胡博,等.基于时序生成对抗网络的月度风光发电功率场景分析方法[J].中国电机工程学报,2022,42(2):537-548.
- LI Hui, REN Zhouyang, HU Bo, et al. A sequential generative adversarial network based monthly scenario analysis method for wind and photovoltaic power [J]. Proceedings of the CSEE, 2022, 42(2): 537-548.
- [19] 廖文龙,杨德昌,葛磊蛟,等.基于像素卷积生成网络的可再生能源场景生成[J].高压技术,2022,48(4):1320-1331.
- LIAO Wenlong, YANG Dechang, GE Leijiao, et al. Renewable energy scene generation based on pixel convolutional generation network [J]. High Voltage Engineering, 2022, 48(4): 1320-1331.
- [20] 黄南天,王文婷,蔡国伟,等.计及复杂气象耦合特性的模块化去噪变分自编码器多源-荷联合场景生成[J].中国电机工程学报,2019,39(10):2924-2934.
- HUANG Nantian, WANG Wenting, CAI Guowei, et al. Modular denoising variational autoencoder multi-source-load joint scene generation considering complex meteorological coupling characteristics [J]. Proceedings of the CSEE, 2019, 39(10): 2924-2934.
- [21] 彭雨箏,李晓露,李聪利,等.基于残差卷积自编码的风光荷场景生成方法[J].电力建设,2021,42(8):10-17.
- PENG Yuzheng, LI Xiaolu, LI Congli, et al. Scene-load scene generation method based on residual convolution self-coding [J]. Electric Power Construction, 2021, 42(8): 10-17.
- [22] 窦孝祥.考虑风电与负荷不确定性的配电网重构研究[D].吉林:东北电力大学,2017.

- [23] GIRAUD C. Introduction to high-dimensional statistics [M]. Boca Raton: Chapman Hall/CRC Press, 2021.
- [24] 王羽凝. 基于马尔科夫链蒙特卡洛方法的光伏电站可靠性评估[D]. 北京: 华北电力大学, 2017.
- [25] 刘航, 申皓, 杨勇, 等. 基于高阶马尔科夫链的纯电重卡集群负荷预测[J]. 中国电力, 2024, 57(5): 61–69.
LIU Hang, SHEN Hao, YANG Yong, et al. Load forecast of electric trucks aggregation based on higher-order Markov chains [J]. Electric Power, 2024, 57(5): 61–69.
- [26] 董晓翀, 孙英云, 蒲天骄. 基于条件生成对抗网络的可再生能源日前场景生成方法[J]. 中国电机工程学报, 2020, 40(17): 5527–5536.
DONG Xiaochong, SUN Yingyun, PU Tianjiao. Renewable energy day-ahead scenario generation method based on Conditional generative adduction network [J]. Proceedings of the CSEE, 2020, 40(17): 5527–5536.

收稿日期: 2023-07-14; 网络首发日期: 2024-05-22

作者简介:

许琴(1987), 女, 高级工程师, 硕士, 研究方向为电力系统运行分析、电网二次系统设计, xuqin@gedi.com.cn;

刘一鸣(1993), 男, 工程师, 硕士, 研究方向为电力系统运行分析、电网二次系统设计, liuyiming@gedi.com.cn;

林济铿(1967), 男, 通信作者, 教授, 博士, 研究方向为电力系统优化调度、EMS/DEMS、电力市场、电网规划、电网应急、人工智能、电网应急, mejklin@126.com。

(上接第50页 Continued from Page 50)

- CHEN Weinan, HU Zhijian, YUE Jingpeng, et al. Short-term load prediction based on combined model of long short-term memory network and light gradient boosting machine [J]. Automation of Electric Power Systems, 2021, 45(4): 91–97.
- [21] 张未, 余成波, 王士彬, 等. 基于VMD-LSTM-LightGBM的多特征短期电力负荷预测[J]. 南方电网技术, 2023, 17(2): 74–81.
ZHANG Wei, YU Chengbo, WANG Shibin, et al. Multi-featured Short-term power load forecasting based on VMD-LSTM-LightGBM [J]. Southern Power System Technology, 2023, 17(2): 74–81.
- [22] JU Y, SUN G, CHEN Q, et al. A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting[J]. IEEE Access, 2019 (7): 28309–28318.
- [23] MENG A, CHEN Y, YIN H, et al. Crisscross optimization algorithm and its application[J]. Knowledge-Based Systems, 2014(67): 218–229.
- [24] ZHU J, ZENG C, LIU Y, et al. A new double-layer optimization mechanism for multi-area OPF in actual distributed environment[J]. CSEE Journal of Power & Energy Systems (Early Access), 2024(5): 1–9.
- [25] 王雪, 刘林, 刘文迪, 等. 基于纵横交叉算法的新型电力系统惯量延迟优化控制策略[J]. 中国电力, 2024, 57(7): 12–20.
WANG Xue, LIU Lin, LIU Wendi, et al. A novel inertia delay optimization control strategy for new power systems based on crisscross optimization[J]. Electric Power, 2024, 57(7): 12–20.
- [26] DEB K, PRATAP A, AGARWAL S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II[J]. IEEE Transactions on Evolutionary Computation, 2002, 6(2): 182–197.
- [27] BARAN M E, WU F F. Network reconfiguration in distribution systems for loss reduction and load balancing [J]. IEEE Transactions on Power delivery, 1989, 4(2): 1401–1407.
- [28] Electric Reliability Council of Texas. [DB/OL] https://www.ercot.com/gridinfo/load/load_hist, 2023–07.
- [29] COELLO C A C, LECHUGA M S. MOPSO: a proposal for multiple objective particle swarm optimization [C]//Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02, May 12–17 2002, Honolulu, HI, USA. New York: IEEE, 2002: 1051–1056.

收稿日期: 2023-09-19; 网络首发日期: 2024-05-20

作者简介:

杨毅(1983), 男, 高级工程师(教授级), 博士, 研究方向为电气与电子工程、电力系统保护与控制、综合能源系统、柔性交直流输电;

王明深(1990), 男, 工程师, 博士, 研究方向为分布式能源优化调控、车网互动技术、充换电规划运营;

朱继忠(1966), 男, 通信作者, 教授, 博士生导师, 欧洲科学与艺术院院士, 意大利博洛尼亚科学院外籍院士, IEEE Fellow, 博士, 研究方向为电力系统运行、综合智慧能源, zhujz@scut.edu.cn。