

基于高斯过程回归与不确定性耦合关系的电力系统规划典型场景提取技术

李峰¹, 高效海², 郑鹏飞¹, 刘帅¹, 高洁¹

(1. 国网山东省电力公司威海供电公司, 山东 威海 264200; 2. 国网山东省电力公司经济技术研究院, 山东 济南 250021)

摘要:针对典型场景生成方法存在未能全面考虑风、光、负荷等不确定因素间的耦合关系、传统聚类方法在高维数据集上表现较差、提取的典型场景不能很好体现原数据特征等问题。首先在改进核函数的基础上, 结合贝叶斯公式和多元高斯分布, 利用高斯过程回归(GPR)对电力系统中的多种不确定因素的耦合关系进行建模, 生成模拟运行数据; 其次, 采用时序分段典型场景提取方法, 划分总调度区间为若干子区间并分别进行中心点聚类, 得出子区间带权典型场景并用笛卡尔积连接生成全调度区间典型场景集; 然后, 应用基于陆地移动距离(EMD)的方法, 进行典型场景提取效果评价; 最后, 通过算例验证了提取的典型场景能更好保留原始基础场景集合的概率分布特性, 充分体现原始数据集中不确定因素之间的耦合关系。结果说明所述方法提取的典型场景能更好体现原数据特征。

关键词:典型场景; 高斯过程回归; EMD距离; 中心点聚类

DOI:10.19781/j.issn.1673-9140.2022.01.008 中图分类号:TM73 文章编号:1673-9140(2022)01-0064-10

Extraction of typical scenarios for power system planning based on Gaussian process regression and uncertainty coupling relationship

LI Feng¹, GAO Xiaohai², ZHENG Pengfei¹, LIU Shuai¹, GAO Jie¹

(1. State Grid Shandong Power Supply Company, Weihai 264200, China;

2. State Grid Shandong Power Supply Company, Economic Research Institute, Jinan 250021, China)

Abstract: The typical scene generation method has problems such as failing to fully consider the coupling relationship between uncertain factors like wind, light and load, traditional clustering methods perform poorly on high-dimensional data sets, and the extracted typical scenes cannot well reflect the characteristics of the original data. To solve these problems, combining with Bayesian formula and multivariate Gaussian distribution, this paper firstly uses Gaussian process regression (GPR) to model the coupling relationship of various uncertain factors in the power system and generate the simulation operation data on the basis of improving the kernel function. Secondly, using the time series segmented typical scene extraction method, the total scheduling interval is divided into several sub-intervals. The center points are clustered respectively. Sub-interval weighted typical scenes are obtained and connected by Cartesian product to generate typical scene set of the full scheduling interval. Then, a method based on land movement distance (EMD) is applied to evaluate the extraction effect of typical scenes. Finally, it is verified that the extracted typical

收稿日期:2020-09-14; 修回日期:2020-12-21

基金项目:国家电网山东省电力公司科技项目(52061318006P)

通信作者:李峰(1972-),男,本科,高级工程师,主要从事电力系统自动化、电网规划研究;E-mail:wh_lifeng@163.com

scenes can better retain the probability distribution characteristics of the original basic scene set and fully reflect the coupling relationship between uncertain factors in the original data set. The results show that the typical scenes extracted by the method can better reflect the characteristics of the original data.

Key words: typical scene; Gaussian process regression; EMD distance; center point clustering

迫于化石能源枯竭和环境保护的压力,世界各国对于可再生新能源的开发正在展开^[1]。全球能源转型处于关键阶段,发展多能互补综合能源电力系统是实现能源转型的重要途径^[2]。随着新能源对电力系统的影响逐渐增大,其内在的不确定性与负荷的不确定性也对系统规划带来了新的挑战,如何进行多能源电力系统协调规划成为热点研究问题。

多能源电力系统协调规划中,风电概率模型具有随机性与波动性,很多情况下无法在电力系统规划与运行问题中直接使用,因此需要先将概率模型离散化,离散化后的结果即为场景。通过随机变量的概率模型抽样得到场景集合的过程即为场景生成。场景法常被用于应对如风电、光伏等随机量的接入给电力系统带来的不确定性问题,能够将带有不确定性的概率分布问题量化为具有确定性的几个典型场景的组合,继而用确定性问题方法进行处理,大大降低了求解难度,减小了计算的工作量。

现已有多种基于场景法的实际应用模型。文献[3]提出了一种求解含风电无功优化模型的最优场景法,可以在不使用蒙特卡洛模拟的情况下处理模型中风电的不确定性;文献[4]采用分层聚类算法对风电出力样本进行聚类分析,然后采用类间样本离差平方和作为聚类数的判定依据划分样本,得到聚类分组数未知的情况下地区风电出力典型场景的选取方法;文献[5]基于改进的 K -means 聚类算法对风电、光伏、潮汐、负荷进行同步聚类,在不同的时间尺度下生成电力系统典型运行场景;文献[6]以风电场出力、负荷水平为标准,对网络状态进行场景聚类,综合所有场景调度信息,在传统运行规划制定方法的基础上,提出风电接入下断面极限传输功率运行规划的场景聚类提取方法。

场景法运用的前提是根据历史数据模拟生成大量的运行场景。文献[7]首先得到符合风电输出功率的连续正态概率分布,然后通过蒙特卡洛法对概率分布进行抽取,得到用于生成典型场景的大量样

本;文献[8]考虑不确定因素间的相关性,对风电、光伏、负荷 3 个随机变量进行相关性建模;文献[9]提出基于数据挖掘的贝叶斯网络模型,描述了风—光—负荷之间的非线性相关性;文献[10-12]考虑了风光等分布式电源以及负荷的自然增长时序数学模型,通过时间序列方法模拟出包含自然增长信息的电力系统场景样本;文献[13]利用拉丁超立方方法对连续正态分布进行抽样,通过更少的抽样次数实现对输入分布的重建;文献[14]通过马尔科夫链方法减弱了随时间延长而造成的误差累积效应。

在拥有大量样本集的基础上,场景法需要进一步通过聚类方法形成带权典型场景集。文献[15]通过双层自适应 K -means 方法,在构建局部典型曲线的基础上进行二次聚类,从而构建全局模型;文献[16]考虑了曲线形态,基于云模型提取出了电力负荷的典型模式;文献[17]分析比较了多种聚类有效性评价指标,并在此基础上评价了多种聚类方法和多种降维方法;文献[18]从数据中提取了 7 类降维特征指标取代历史数据作为新的输入,利用熵权法对指标进行客观赋权,并用模糊 C -means 法进行聚类分析;文献[19]建立了一种自适应的 K 均值算法,利用临界测试在不满足条件时进行聚类中心的动态分离,在可能的最小聚类数目情况下,将相关性高的聚类中心进行合并。

上述文献虽然部分考虑了风、光、负荷等电力系统中的不确定因素,但其典型场景生成时并未全面考虑相互耦合关系,导致生成的场景中各不确定因素仍较为独立。另外, C -means 聚类、 K -means 聚类、层次聚类等传统聚类方法在高维数据集上的表现较差,提取的典型场景不能很好地体现原数据的特征。因此,本文提出一种基于高斯过程回归的场景生成技术,通过联合高斯分布考虑不同不确定因素之间的相互耦合关系,进而进行场景生成;同时,将总调度区间划分为多个子调度区间,对每个子调度区间单独进行典型场景提取并结合成横跨总调度

区间的典型场景,从而实现了高维问题向低维问题的转化;最后,利用 Earth Mover's Distance 衡量所提取的典型场景对原场景的表征程度,并在此基础上对算法超参数进行优化,进行典型场景提取效果评价。

1 基于高斯过程回归的场景生成技术

1.1 高斯过程回归

高斯过程又称为正态随机过程,是一种基于贝叶斯理论及统计学习理论的具有概率意义的核学习机。在已知一系列点及对应值的前提下,设各个点对应的取值符合高维多元高斯分布,在此基础上求取点到取值的映射空间中不同映射的概率即为高斯过程^[20]。

映射的概率后验分布可以通过贝叶斯公式求得,即

$$p(f|D) = \frac{p(f)p(D|f)}{p(D)} \quad (1)$$

式中 f 为从输入空间到取值空间的映射; D 为输入数据点的集合; $p(D|f)$ 为对数据点的最大似然估计; $p(f)$ 为映射概率的先验分布。

$$f = (f(x_1), f(x_2), \dots, f(x_N)) \quad (2)$$

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\} = (X, y) \quad (3)$$

式(2)、(3)中 $x_i (i = 1, 2, \dots, N)$ 为输入空间中的 n 维输入量; $y_i (i = 1, 2, \dots, N)$ 为高斯过程回归用于短期负荷预测建模时,训练集合 D 中的训练样本点。

对于先验分布来说,由于数据点特征未知,另外考虑到对称性,设映射 f 的先验分布符合标准多元高斯分布^[21],即

$$f \sim G_P(\bullet | 0, \mathbf{K}) = N(0, \mathbf{K} + \sigma^2 \mathbf{I}) \quad (4)$$

其中表征了测量值与实际值的误差为

$$\epsilon_i = y_i - f(x_i), \epsilon_i \sim N(\bullet | 0, \sigma) \quad (5)$$

式(4)、(5)中 \mathbf{K} 为相关度矩阵,衡量了不同数据点的耦合程度。 \mathbf{K} 中元素通过核函数来计算。核函数是 GPR 中极其重要的一环,不同应用场景应选择特定的核函数与核函数中的超参数^[22]。电力系统适用的核函数将在下一节中讨论。

GPR 的计算目标是实现测试点对应的预测值

取得最大的概率值^[23]。通过联合概率分布边缘化可以消去概率分布中特定的变量。通过对映射积分进行边缘化,即

$$p(y_{\text{new}} | x_{\text{new}}, D) = \int p(y_{\text{new}} | x_{\text{new}}, f, D) p(f | D) df \quad (6)$$

式中 $p(y_{\text{new}} | x_{\text{new}}, f, D)$ 为在已知映射以及数据点的前提下,求取得目标值的概率; $p(f | D)$ 可通过式(1)计算。代入后的具体推导过程可参照文献[24]。预测值概率分布为

$$\begin{aligned} p(y_{\text{new}} | x_{\text{new}}, D) &= N(\mu_*, \sigma_*^2) \\ \mu_* &= \mathbf{K}_{*,\cdot} (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{y} \\ \sigma_*^2 &= \mathbf{K}_{*,*} - \mathbf{K}_{*,\cdot} (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{K}_{\cdot,*} + \sigma^2 \end{aligned} \quad (7)$$

式中 y_{new} 为测试点预测值; x_{new} 为测试点值; μ_* 为测试点概率分布均值; σ_*^2 为测试点概率分布方差。训练点和测试点组成的数据点集合对应的相关度矩阵为

$$\mathbf{K}_{\text{whole}} = \begin{bmatrix} \mathbf{K} & \mathbf{K}_{\cdot,*} \\ \mathbf{K}_{*,\cdot} & \mathbf{K}_{*,*} \end{bmatrix} \quad (8)$$

与式(5)相同,式(8)的相关度矩阵中的元素将通过式(9)中的核函数来计算。

GPR 适用于高维数据的处理。电力系统应用中常常涉及到一个场景内包含多个时序时间点的情况,且每个时序点都是一个维度。因此,随着场景时间点数目的增多,电力系统场景维度常常超过 20。此外,各个点对应的高斯分布之间通过相关性系数进行耦合,测试点回归结果也包含了不同训练点的影响。在此基础上,可以利用高斯过程回归对电力系统中的多种不确定因素的耦合关系进行建模,使得生成的场景数据体现多种不确定因素的影响。另外,GPR 可以根据不同的任务采用不同的核函数将数据映射到特定的高维空间,增加了算法的灵活性。综上所述,GPR 适合应用于电力系统中的数据处理和场景问题。

1.2 核函数的确定

本文根据风、负荷不确定性的影响,应用基于改进核函数的 GPR 进行场景生成。选用的 GPR 核函数为

$$k(x_n, x_m) =$$

$$\lambda \left(\theta_0 \exp \left(-\frac{\theta_1}{2} (x_n - x_m)^2 \right) + \theta_2 + \theta_3 x_n^T x_m \right) \quad (9)$$

式中 θ_0 为常用的径向基项核函数项系数; θ_2 为常数项; θ_3 则对应了输入变量的一个线性函数的参数模型的系数; λ 表征了风、负荷 3 项不确定因素之间的耦合关系。 λ 在不同情况下有不同取值, 即

$$\lambda = \begin{cases} c_0, & x_m \in X_n \\ \frac{1-c_0}{2}, & x_m \notin X_n \end{cases} \quad (10)$$

式中 X_n 为 x_n 所在的数据点集合, 包括风电数据点集以及负荷数据点集。 如果 x_m 与 x_n 同处同一数据点集, 说明 x_m 与 x_n 为同一项不确定因素的 2 个样本数据, x_m 的取值对 x_n 对应预测值影响较大, 在相关度矩阵中表征为二者相关度取值较大; 若 x_m 与 x_n 处于不同数据点集, 说明二者为不同项不确定因素的 2 个样本数据, x_m 的取值对 x_n 对应预测值影响较小, 在相关度矩阵中表征为二者相关度取值较小。 因此, c_0 应在 0~1 之间取得一个较大值, 表征同类数据点之间的强相关性。

2 时序分段典型场景提取方法

电力系统中, 根据大量的风、负荷等历史数据所提取出的具有代表性的典型场景, 可以用于反映周期内风电、负荷的变化特征。 经典场景集是可以代表原始场景的一组场景, 每一个场景由覆盖总调度区间的所有时间点构成。 聚类算法根据不同场景间的距离进行计算。 然而, 如果选用总调度区间作为一个场景进行聚类, 则会出现维度过高的问题, 且场景距离不能很好地反映场景之间的差异程度, 聚类效果较差。

因此, 本文提出一种对总调度区间进行拆分的时序分段典型场景提取分法, 首先将总调度区间分成若干个子区间, 即拆分为部分间数据特征差异较大、部分内数据特征较为统一的几个部分; 然后对各个部分分别进行聚类分析, 得出各个部分的带权典型场景集合; 最后将各个部分采用笛卡尔积进行连接, 生成总调度区间长度的带权经典场景集。

2.1 时间维度的划分

设时间的总长度为 T , 选取 $k-1$ 个分位点, 则总调度区间被分为 k 个子调度区间。 根据每个子调度区间内数据变化的趋势应该相似的原则, 将时间轴上连续的确定性场景等距离切分成若干时间尺度相同的场景。 因此选取二阶导数作为子调度区间的划分标准, 即

$$t_k = \{t \mid |f''(t)| > \zeta, t \in T\} \quad (11)$$

需要选取合适的 ζ , 使得总调度区间被划分成合理数量的几个区间。 一方面, 划分数量不能太少, 否则可能导致各个区间内数据变化趋势的一致性较差, 无法聚类出较好的结果; 另一方面, 划分数量也不能太多, 否则即使区间内数据变化趋势趋于一致, 但由于区间数量过多, 使得经过笛卡尔积连接之后的全调度区间典型场景数目过大, 失去提取典型场景的意义。 一般来说, 划分的典型场景数目为 3~10, 能使场景具有一定带表性且有利于后续的分析。

根据各子调度区间内数据变化趋势相似的原则, 对全调度区间进行分段时, 若所分得子场景无法等长, 则尽量保留各个子场景的最典型部分, 在此基础上对每个子区间进行部分舍弃, 使得最终划分出的子调度区间等长。 图 1 为场景的子调度区间划分。

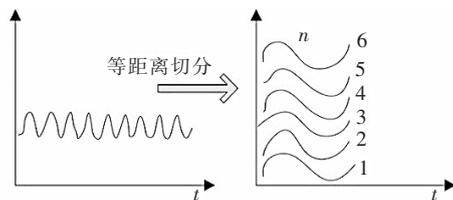


图 1 场景划分

Figure 1 Schematic diagram of deterministic scene set reduction

2.2 聚类算法说明

聚类算法是一种无监督学习算法, 是一种不需要人的指导, 机器自行从数据中找出规律, 将具有相似性的事物聚在一起的算法^[25]。 聚类时, 样本往往处于欧式空间中, 每一类的中心点由算法学习得出, 但通常需要提前指定中心点数量。 本文使用中心点聚类方法, 该算法克服了传统 K -means 算法容易被离群点影响的缺点, 所迭代出的中心点受异常点的影响较小^[26]。 中心点聚类的一般步骤。

1) 确定中心点数量 K 。

2)在数据点集合中随机选择 K 个点作为初始类中心。

3)计算点集合其余所有点到 K 个中心点的距离,并且归入最近类中心所代表的类。

4)对于每个类簇,计算类中所有点到其余所有点的距离之和,选取最小者为新的类中心。

5)重复步骤 3、4,直到满足迭代停止条件或者聚类结果不发生改变,即迭代收敛为止。

2.3 笛卡尔积连接

对各个子调度区间分别应用中心点聚类进行典型场景提取,得到不同子调度区间的几组典型场景集。对不同子调度区间的典型场景集应用笛卡尔积连接,形成全调度区间典型场景集。笛卡尔积连接又叫做交叉连接,具体指 2 个集合中的每个元素都与另一集合的所有元素相组合,如图 2 所示。全调度区间典型场景提取流程如图 3 所示。

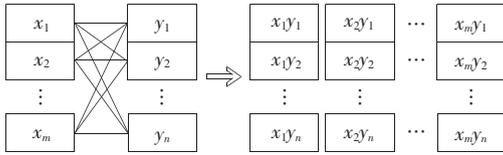


图 2 子调度区间通过笛卡尔积连接

Figure 2 Sub-scheduling intervals are connected by Cartesian product

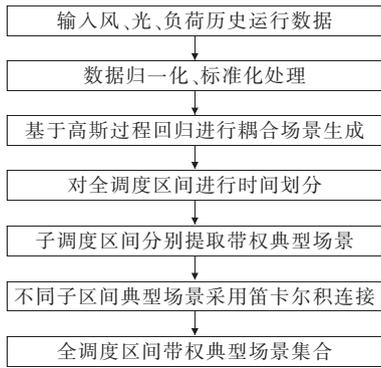


图 3 基于高斯过程回归的典型场景提取方法

Figure 3 Typical scene extraction method based on Gaussian process regression

设每个调度区间提取出的典型场景数目为 N_1, N_2, \dots, N_k , 则经过笛卡尔积连接的全调度区间典型场景数为 $\sum_{i=1}^k N_i$ 。可以看出,全调度区间典型场景总数目 N 与每个子调度区间的典型场景数目成正比,与子调度区间数目 k 呈指数关系。因此要控制典型场景数量,最主要的是选取合适的子调度区间数目 k 。

3 典型场景提取评价及优化

在提取典型场景后,需要评价典型场景对原场景集表征效果,以衡量典型场景提取算法的优劣。在此基础上可以进一步对典型场景提取算法进行超参数优化。本文选用基于 Earth Mover's Distance (EMD)的方法进行典型场景提取效果评价。

3.1 EMD 距离介绍

EMD 距离是 2 个概率分布间距离的一种度量方式^[27],可以实现信息与信息之间多对多的优良匹配,因此很适用于信息特征单元聚类问题。对于 2 个给定的概率分布,EMD 将二者分别视为分散在空间的泥土和一组坑洞。EMD 衡量的就是用泥土将坑洞填满所需的最小工作量。设 P_{b1}, P_{b2} 为 2 个概率分布, P_{b1}, P_{b2} 的离散化概率分布点集分别为

$$\begin{cases} P = \{(p_1, \omega_{p_1}), (p_2, \omega_{p_2}), \dots, (p_m, \omega_{p_m})\} \\ Q = \{(q_1, \omega_{q_1}), (q_2, \omega_{q_2}), \dots, (q_m, \omega_{q_m})\} \end{cases} \quad (12)$$

式中 $p_i (i = 1, 2, \dots, m)$ 为 P_{b1} 的离散化特征数据点; $\omega_{p_i} (i = 1, 2, \dots, m)$ 为对应的权重,权重由该特征数据点所代表的的数据点个数决定; $q_i (i = 1, 2, \dots, n)$ 为 P_{b2} 的离散化特征数据点; $\omega_{q_i} (i = 1, 2, \dots, n)$ 为对应的权重。则有 $D = [d_{i,j}]$, $d_{i,j}$ 为从特征数据点 i 到特征数据点 j 的代数距离,即

$$d_{i,j} = p_i - q_j \quad (13)$$

现需要求解得到流矩阵 $F = [f_{i,j}]$, 其中 $f_{i,j}$ 为第 i 个土堆到第 j 个坑洞的运输数量,即从特征数据点 i 到特征数据点 j 的流。流矩阵的求解是一个以整体费用最小化为目标的约束线性规划,即

$$\begin{cases} \min & \sum_{i=1}^m \sum_{j=1}^n f_{i,j} d_{i,j} \\ \text{s. t.} & f_{i,j} \geq 0, 1 \leq i \leq m, 1 \leq j \leq n \\ & \sum_{j=1}^n f_{i,j} \leq \omega_{p_i}, 1 \leq i \leq m \\ & \sum_{i=1}^m f_{i,j} \leq \omega_{q_j}, 1 \leq j \leq n \\ & \sum_{i=1}^m \sum_{j=1}^n f_{i,j} = \min \left\{ \sum_{i=1}^m \omega_{p_i}, \sum_{j=1}^n \omega_{q_j} \right\} \end{cases} \quad (14)$$

上述线性问题实际上是求取最佳目标函数,在

满足一系列约束的条件下,使得整体成本最小。求解得到流矩阵后,得 P_{b1} 、 P_{b2} 之间的 EMD 距离为

$$E_{MD}(P_{b1}, P_{b2}) = \sum_{i=1}^m \sum_{j=1}^n f_{i,j} d_{i,j} / \sum_{i=1}^m \sum_{j=1}^n f_{i,j} \quad (15)$$

将典型场景集合各点概率密度函数离散化得到的概率分布 P_i 、原场景集合各点密度函数离散化得到的概率分布 P'_i 代入公式中的 P_{b1} 、 P_{b2} 概率分布,即可进行同一时间点下典型场景集合与原场景集合之间的 EMD 距离计算。

3.2 典型场景表征效果评价方法

对于长度为 T 的全调度区间中的每个时间点 T_i ,基于提取的带权典型场景集合,可以计算出当前时间点的出力概率密度函数 $P(T_i)$ 以及原始数据场景集合每个时间点的出力概率密度函数 $P'(T_i)$ 。然后,通过 EMD 距离的计算,可以衡量同一时间点下典型场景集合概率分布与原场景集合概率分布之间的距离 l_i 。最后对所有时间点的 EMD 距离求代数平均得到全调度区间的平均 EMD 距离 l_m 。此时 l_m 即为反映典型场景集对原场景集表征效果的评价指标。

基于平均 EMD 距离与数据集中验证集,不符合收敛条件的平均 EMD 距离将进行超参数优化。超参数优化对机器学习算法的性能起着至关重要的作用,实际应用中,通过不断调整参数值来提高算法的实践性能。在贝叶斯方法中,超参数(hyperparameters)指控制模型参数分布的参数,选取最优的超参数可以使得 GPR 的结果获得较高的精度。

在本文基于 GPR 进行典型场景提取的模型中,场景个数对提取结果和子场景的划分均产生影响,因此视其为待优化参数。鉴于 GPR 超参数优化问题中,超参数设置不具备先验知识的情况下,本文采用文献[28]的超参数优化算法—改进粒子群优化算法(ADPSO-GPR),对 GPR 中超参数进行自适应优化。计算全调度区间的 EMD 距离后,当不符合式(12)中的收敛条件时,该算法对典型场景进行缩减,调整子调度区间的数量和典型场景数目。反复迭代直至全调度区间平均 EMD 距离符合收敛条件,使最终输出的典型场景集最具代表性,达到最优效果。基于 EMD 距离的典型场景表征效果评价流程如图 4 所示。

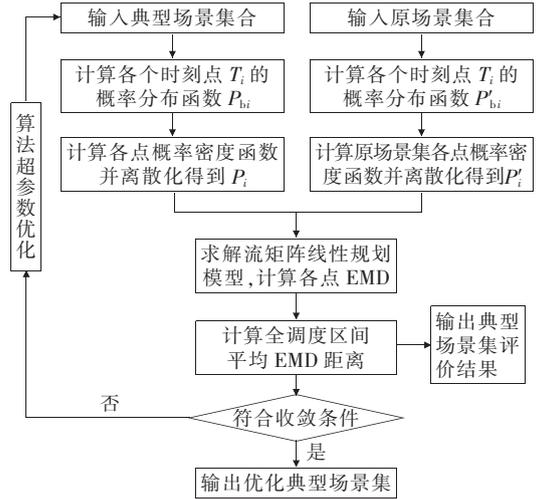


图 4 基于 EMD 距离的典型场景集评价及优化方法
Figure 4 Evaluation and optimization method of typical scene set based on EMD distance

3.3 风电、负荷 Pearson 相关性系数

统计学中,考察 2 个变量之间相关程度的系数为相关系数。对于 2 个变量 $X = \{x_1, x_2, \dots, x_n\}$ 和 $Y = \{y_1, y_2, \dots, y_n\}$, Pearson 相关系数为

$$\rho = \frac{L_{XY}}{\sqrt{L_{XX}L_{YY}}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (16)$$

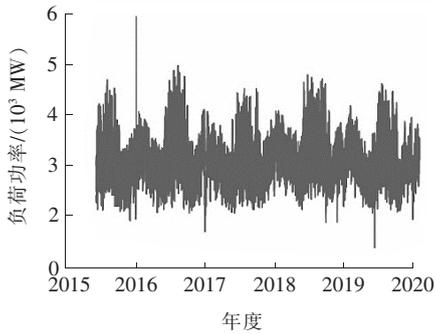
计算出的相关系数大小主要含义如下:

- 1) 相关系数为 0, 表示 2 变量间无关联;
- 2) 2 变量为正相关, 相关系数为 0.00~1.00;
- 3) 2 变量为负相关, 相关系数为 -1.00~0.00。

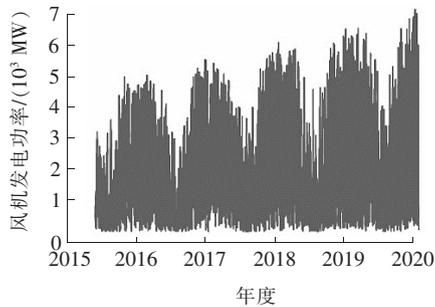
由此可知,相关系数的绝对值越大,相关性越强。相关系数越接近于 0,相关性越弱。计算风电、负荷 Pearson 相关性系数时,将风电、负荷相关数据分别代入变量 $X = \{x_1, x_2, \dots, x_n\}$ 和 $Y = \{y_1, y_2, \dots, y_n\}$,即可通过式(16)计算得到风电、负荷 Pearson 相关性系数。根据计算出的相关系数大小主要含义,即可判断风电出力与负荷 2 变量间的耦合关系。

4 算例分析

为验证本文方法的可行性,选取美国 PJM 电力市场 2015~2019 年的负荷与风机出力数据,采样间隔为 1 h。进行缺失值处理等数据清洗后,总历史数据长度为 40 868 h。负荷及风电数据如图 5 所示。



(a) 负荷功率历史运行数据



(b) 风机功率历史运行数据

图 5 PJM 电力市场 2015~2019 历史风电、负荷运行数据

Figure 5 PJM electricity market 2015~2019 historical wind power, load operation data

4.1 基于高斯过程回归进行耦合场景生成

首先利用高斯过程对历史数据进行回归生成大量模拟运行数据,再对运行数据以 d 为单位进行时间划分生成基础场景集合。场景法运用的前提是根据历史数据生成大量的运行场景,为了加快计算速度,本文采取滚动回归的方式进行预测,以每 20 d 的历史数据作为训练集生成未来 5 d 的模拟运行数据。为了增强基础场景集对运行场景的表征能力,对生成的模拟运行数据加入扰动 ϵ , ϵ 符合均值为零的高斯分布,其标准差取高斯过程在对应点输出的预测值标准差。每个场景通过一个对应的维度为 $(24, 2)$ 的二维向量表示,2 个维度分别表示对应时间的负荷数据与风电数据。最终生成的基础场景集规模为 85 100。部分生成的基础负荷场景如图 6、7 所示。

由图 6、7 可知,负荷数据的波动性较弱,且不同场景之间关联性较强;而风电数据的波动性强,且不同场景之间的关联性较弱。接下来将通过子场景划分将全区间调度场景划分为几个子调度区间从而弱化风电数据的波动性,提升聚类有效性。

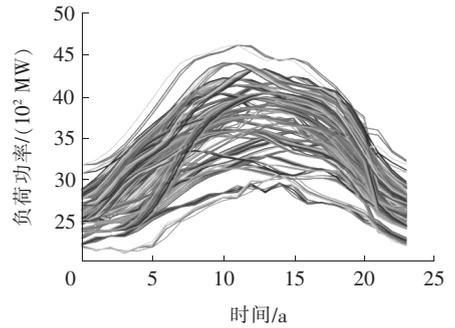


图 6 部分负荷基础场景集

Figure 6 Partial load base scenes

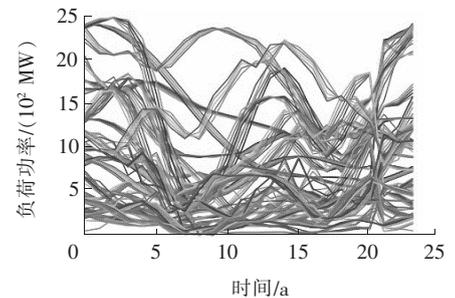


图 7 部分风电基础场景集

Figure 7 Basic wind power scenes

4.2 全调度区间典型场景集生成

在对场景集合进行聚类前,需要进行子场景划分,再对每段子场景分别聚类。子场景数目越大,典型场景对原始场景集合的表征效果就越好,但由于笛卡尔积连接将按照指数级增加场景数量,过大的子场景数目将会使得总典型场景规模过大。同时,子场景数目越小,总典型场景规模越小,但对原始场景集合的表征能力就越低。同时考虑到典型场景集规模与聚类效果的要求,本文取场景划分段数为 3。

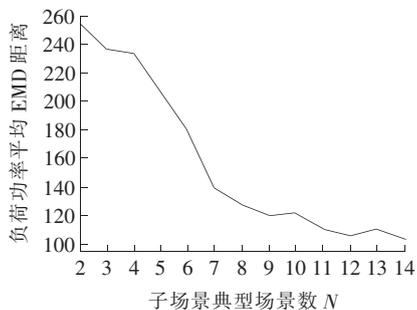
考虑到 03:00 为大部分场景的晚间负荷最低点,取场景起点时间为 3,以此保证大部分场景在分界点 2 端分别保持单调,降低子区间曲线复杂度,从而提升聚类效果。为减小单个子区间区间长度,本文取 2 个子场景分界点分别为 11、19,从而使得 3 个子场景等长且均为 8。

4.3 典型场景集评价及优化

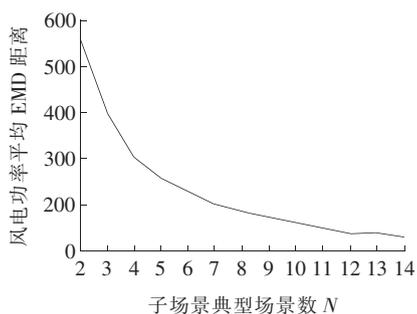
运用典型场景评价方法,采用等步长搜索法对每个子场景的典型场景数进行优化,结果如图 8 所示。

根据图 8(a)可知,负荷典型场景表征负荷原始场景的能力较好,子场景典型场景数 N 达到 9 后,

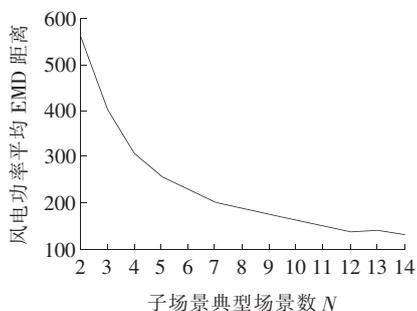
负荷功率平均 EMD 距离低于 120,并且随着典型场景数增加表征效果提升不明显;根据图 8(b)可知,风电典型场景表征原始场景的能力不如负荷数据,这是由于风电本身波动性较强的特点导致,但当典型场景数达到 10 后,风电功率平均 EMD 距离小于 150,且典型场景数增加带来的表征效果提升不明显;根据图 8(c)可知,当子场景典型场景数达到 7 后,风电、负荷功率 Pearson 相关性系数大于 0.35,而原始数据场景集中风电、负荷功率相关系数为 0.424 4,两者差距不明显,证明本文生成的典型场景较好地保留了原始场景中不确定性因素之间的耦合关系,具有一定科学性和准确性。且根据 Pearson 相关性系数的主要含义,相关系数为 0.00~1.00 时,2 变量为正相关,因此风电与负荷功率呈现正相关。



(a) 负荷功率平均 EMD 距离



(b) 风电功率平均 EMD 距离



(c) 风电、负荷 Pearson 相关性系数

图 8 典型场景数对聚类效果的影响

Figure 8 Effect of typical scene number on clustering effect

子典型场景数达到 10 后,场景数目增加对于聚类效果的提升能力显著降低。因此,本文取子典型场景数为 10。笛卡尔积连接后的总场景集合的评价指标汇总见表 1。由表 1 可得,生成的典型场景集合在不同时刻点的概率分布与原始数据的概率分布差异较小,并且很好的保留了原始数据集合中不同不确定因素之间的耦合关系。

表 1 典型场景集效果评价指标

Table 1 Effect evaluation indicators for typical scene sets

评价指标	得分
平均 EMD 距离负荷功率	121.435
平均 EMD 距离风电功率	171.032
负荷、风电 Pearson 相关性系数	0.349 1

5 结语

本文首先考虑了电网中不确定性因素之间的耦合关系,在改进核函数的基础上,利用高斯过程回归对电力系统中的多种不确定因素的耦合关系进行建模,生成模拟运行数据。然后,采用时序分段典型场景提取方法,划分总调度区间为若干子区间并分别进行中心点聚类,得出子区间带权典型场景并用笛卡尔积连接,生成全调度区间典型场景集,在以上基础上,应用 EMD 方法和超参数优化,进行典型场景提取效果评价。最后,基于美国 PJM 电力市场历史实际运行数据进行了算例分析,算例表明本文提取的典型场景能够很好地保留原始基础场景集合的概率分布特性,同时保留了风电、负荷等不确定性因素之间的耦合特性。本文方法为解决电力系统规划问题奠定了基础,为衡量规划方案优劣提供了新的途径。

参考文献:

[1] 李相俊, 马会萌, 姜倩. 新能源侧储能配置技术研究综述[J]. 中国电力, 2022, 55(1):13-25.
 LI Xiangjun, MA Huimeng, JIANG Qian. Review of energy storage configuration technology on renewable energy side[J]. Electric Power, 2022, 55(1):13-25.

[2] 王圆圆, 白宏坤, 李文峰, 等. 能源大数据应用中心功能体系及应用场景设计[J]. 智慧电力, 2020, 48(3):15-21+29.
 WANG Yuanyuan, BAI Hongkun, LI Wenfeng, et al.

- Function system and application scenario design of energy big data application center[J]. Smart Power, 2020, 48(3): 15-21+29.
- [3] 齐晖,程韧俐,史军,等.采用最优场景法求解含风电的无功优化模型[J].广东电力,2019,32(6):18-28.
QI Hui, CHEN Renli, SHI Jun, et al. Optimal scenario method for solving reactive power optimization model incorporating wind farms[J]. Guangdong Electric Power, 2019, 32(6): 18-28.
- [4] 林俐,费宏运,刘汝琛,等.基于分层聚类算法的地区风电出力典型场景选取方法[J].电力系统保护与控制, 2018, 46(7): 1-6.
LIN Li, FEI Hongyun, LIU Ruchen, et al. A regional wind power typical scenarios' selection method based on hierarchical clustering algorithm[J]. Power System Protection and Control, 2018, 46(7): 1-6.
- [5] 刘若凡,刘俊勇,刘友波,等.风电接入下断面极限传输功率运行规则的场景聚类提取方法[J].电力自动化设备, 2014, 34(10): 69-74+80.
LIU Ruofan, LIU Junyong, LIU Youbo, et al. Scenario clustering method for extracting TTC operating rules of transmission corridor with wind power[J]. Electric Power Automation Equipment, 2014, 34(10): 69-74+80.
- [6] 秦志龙,李文沅,熊小伏.含具有风速相关性风电场的发输电系统可靠性评估[J].电力系统保护与控制, 2013, 41(20): 27-33.
QIN Zhilong, LI Wenyuan, XIONG Xiaofu. Reliability assessment of composite generation and transmission systems with wind farms considering wind speed correlations[J]. Power System Protection and Control, 2013, 41(20): 27-33.
- [7] 王蓓蓓,刘小聪,李扬.面向大容量风电接入考虑用户侧互动的系统日前调度和运行模拟研究[J].中国电机工程学报, 2013, 33(22): 35-44.
WANG Beibei, LIU Xiaocong, LI Yang. Day-ahead generation scheduling and operation simulation considering demand response in large-capacity wind power integrated systems[J]. Proceedings of the CSEE, 2013, 33(22): 35-44.
- [8] 李珂,郇能灵,张沈习,等.考虑相关性的分布式电源多目标规划方法[J].电力系统自动化, 2017, 41(9): 51-57+199.
LI Ke, XUAN Nengling, ZHANG Shenxi, et al. Multi-objective planning method of distributed generators considering correlations[J]. Automation of Electric Power Systems, 2017, 41(9): 51-57+199.
- [9] 王洪涛,李晓刚,邹斌.基于贝叶斯网络刻画风—光—荷相关性的配电网概率潮流计算[J].中国电机工程学报, 2019, 39(16): 4753-4763+4977.
WANG Hongtao, LI Xiaogang, ZOU Bin. Probabilistic load flow calculation of distribution system based on bayesian network to depict wind-photovoltaic-load correlation[J]. Proceedings of the CSEE, 2019, 39(16): 4753-4763+4977.
- [10] 张楠,黄越辉,王晶,等.基于聚类算法的风电波动过程研究[J].电测与仪表, 2020, 57(6): 73-81.
ZHANG Nan, HUANG Yuehui, WANG Jing, et al. Analysis of wind power fluctuation characteristics of three-north regions based on clustering algorithm[J]. Electrical Measurement & Instrumentation, 2020, 57(6): 73-81.
- [11] 陈建华,王磊,巨云涛.基于 Mean-shift 聚类的大规模新能源对电网线损影响评估方法[J].供用电, 2021, 38(2): 59-64+76.
CHEN Jianhua, WANG Lei, JU Yuntao. Evaluation method of large scale new energy impact on power grid line loss based on Mean-shift clustering[J]. Distribution & Utilization, 2021, 38(2): 59-64+76.
- [12] 刘洁颖,刘俊勇,黄媛,等.基于面板数据有序聚类的主动配电网规划场景降维技术[J].电网技术, 2017, 41(4): 1132-1138.
LIU Jieying, LIU Junyong, HUANG Yuan, et al. Scene dimensionality reduction technology in active power distribution network planning based on orderly clustering of panel data[J]. Power System Technology, 2017, 41(4): 1132-1138.
- [13] 于晗,钟志勇,黄杰波,等.采用拉丁超立方采样的电力系统概率潮流计算方法[J].电力系统自动化, 2009, 33(21): 32-36.
YU Han, ZHONG Zhiyong, HUANG Jiebo, et al. A probabilistic load flow calculation method with latin hypercube sampling[J]. Automation of Electric Power Systems, 2009, 33(21): 32-36.
- [14] 董雷,孟天骄,陈乃仕,等.采用马尔可夫链-多场景技术的交直流主动配电网优化调度[J].电力系统自动化, 2018, 42(5): 147-153.

- DONG Lei, MENG Tianjiao, CHEN Shinai, et al. Optimized scheduling of AC/DC hybrid active distribution network using markov chains and multiple scenarios technique[J]. Automation of Electric Power Systems, 2018, 42(5): 147-153.
- [15] 朱文俊, 王毅, 罗敏, 等. 面向海量用户用电特性感知的分布式聚类算法[J]. 电力系统自动化, 2016, 40(12): 21-27.
- ZHU Wenjun, WANG Yi, LUO Min, et al. Distributed clustering algorithm for awareness of electricity consumption characteristics of massive consumers[J]. Automation of Electric Power Systems, 2016, 40(12): 21-27.
- [16] 宋易阳, 李存斌, 祁之强. 基于云模型和模糊聚类的电力负荷模式提取方法[J]. 电网技术, 2014, 38(12): 3378-3383.
- SONG Yiyang, LI Cunbin, YI Zhiqiang. Extraction of power load patterns based on cloud model and fuzzy clustering[J]. Power System Technology, 2014, 38(12): 3378-3383.
- [17] 黎学春, 高宇男, 李磊, 等. 基于自适应神经网络系统聚类的特殊负荷分类研究[J]. 电网与清洁能源, 2019, 35(1): 8-15.
- LI Xuechun, GAO Yunan, LI Lei, et al. Research on special load classification based on self-adaptive neural network hierarchical cluster[J]. Power System and Clean Energy, 2019, 35(1): 8-15.
- [18] 宋军英, 何聪, 李欣然, 等. 基于特征指标降维及熵权法的日负荷曲线聚类方法[J]. 电力系统自动化, 2019, 43(20): 65-76.
- SONG Junying, HE Cong, LI Xinran, et al. Daily load curve clustering method based on feature index dimension reduction and entropy weight method[J]. Automation of Electric Power Systems, 2019, 43(20): 65-76.
- [19] 孙毅, 冯云, 崔灿, 等. 基于动态自适应 K 均值聚类的电力用户负荷编码与行为分析[J]. 电力科学与技术学报, 2017, 32(03): 3-8.
- SUN Yi, FENG Yun, CUI Can, et al. Power user load code and behavior analysis based on dynamic adaptive K -means clustering[J]. Journal of Electric Power Science and Technology, 2017, 32(03): 3-8.
- [20] BISHOP C M. Pattern recognition and machine learning (information science and statistics) [M]. New York: Springer-Verlag, 2006.
- [21] 张建文, 杨晨, 冉懿, 等. 基于 PCA-GPQR 的电网负荷短期概率预测[J]. 电力系统及其自动化学报, 2020, 32(5): 24-29.
- ZHANG Jianwen, YANG Chen, RAN Yi, et al. Short-term load probability forecasting based on PCA-GPQR [J]. Proceedings of the CSU-EPSA, 2020, 32(5): 24-29.
- [22] 刘升伟, 王星华, 鲁迪, 等. 基于改进高斯过程回归的短期负荷概率区间预测方法[J]. 电力系统保护与控制, 2020, 48(1): 18-25.
- LIU Shengwei, WANG Xinghua, LU Di, et al. Electric load probabilistic interval prediction method based on improved gaussian process regression[J]. Power System Protection and Control, 2020, 48(1): 18-25.
- [23] 梁智, 孙国强, 卫志农, 等. 基于变量选择与高斯过程回归的短期负荷预测[J]. 电力建设, 2017, 38(2): 122-128.
- LIANG Zhi, SUN Guoqiang, WEI Zhinong, et al. Short-term load forecasting based on variable selection and gaussian process regression[J]. Electric Power Construction, 2017, 38(2): 122-128.
- [24] PUTNAM J. Computer vision: models, learning, and inference[J]. Computing Reviews, 2013, 54(6): 348-349.
- [25] 李航. 统计学习方法 [M]. 北京: 清华大学出版社, 2012.
- [26] HAE-SANG P, CHI-HYUCK J. A simple and fast algorithm for K -medoids clustering[J]. Expert Systems with Applications, 2009, 36(2): 3336-3341.
- [27] RUBNER Y, CARLO T, LEONIDAS J. G. The earth mover's distance as a metric for image retrieval[J]. International Journal of Computer Vision, 2000, 40(2): 99-121.
- [28] 曹文梁, 康岚兰. 高斯过程回归超参数自适应选择粒子群优化算法[J]. 合肥工业大学学报(自然科学版), 2019, 42(11): 1479-1484.
- CAO Wenliang, KANG Lanlan. Particle swarm optimization for adaptive hyper-parameters acquisition of gaussian process regression[J]. Journal of Hefei University of Technology(Natural Science), 2019, 42(11): 1479-1484.