

面向可靠性评估的两阶段聚类风-光-荷典型场景生成方法

李春燕¹, 赵晨宇¹, 胡 博¹, 陈正宇², 廖庆龙², 万凌云², 谢开贵¹

(1. 输配电装备及系统安全与新技术国家重点实验室, 重庆大学, 重庆 400044;

2. 国网重庆市电力公司电力科学研究院, 重庆 404100)

摘要: 由于风光出力 and 负荷的不确定性, 高比例可再生能源电力系统可靠性评估需要大量输入场景以保证评估精度。为降低计算复杂度, 选取适用的典型场景至关重要。传统典型场景生成直接采用聚类算法对输入场景聚类, 因此面对高维数据处理困难且无法保留负荷极端场景, 生成可靠性评估典型场景的效果不理想。为选取适用于可靠性评估的风-光-荷典型场景, 研究建立了基于密度的噪声应用空间聚类(DBSCAN)和改进 K 均值两阶段聚类模型, 保留极端场景同时引入降维算法处理高维时序数据。算例表明该算法兼具高效性和有效性, 所得典型场景应用于电力系统可靠性评估得到可靠性指标精确度良好。

关键词: 分层抽样; 改进 K 均值聚类; 密度聚类; 可靠性评估; 典型场景; 序贯蒙特卡洛模拟

DOI: 10.12067/ATEEE.2101008

文章编号: 1003-3076(2021)05-0001-09

中图分类号: TM732

1 引言

“十三五”以来, 我国电力行业走上绿色高效转型发展的快速路。截至 2020 年 10 月底, 全国发电装机 21 亿 kW, 其中水电 3.7 亿 kW、风电 2.3 亿 kW、太阳能发电 2.3 亿 kW, 预测显示 2020 年底, 全国发电装机将达到 21.4 亿 kW, 非化石能源装机占比达到 43.7%, 比 2015 年提高 8.7%。与传统能源发电相比, 风光发电具有无污染、可再生和成本低廉等优势, 但受风力和光照等不确定因素影响同时具备波动性和间歇性的典型特征^[1]。因此大规模可再生能源并网将对电力系统安全稳定运行带来消极影响。评估含可再生能源的电力系统可靠性对推动电力系统发展意义深远^[2]。

目前, 国内外对含新能源电力系统的可靠性研究主要基于模拟法(如序贯蒙特卡洛法), 通过随机抽样刻画风/光等可再生能源的不确定性, 进而计算系统的可靠性指标。文献[3]采用蒙特卡洛法对含风电和光伏的电力系统长期性能进行评估; 文献[4]利用基于 Well-being 模型的蒙特卡洛模拟法分

别评估含不同风/光容量配置的发电系统的可靠性, 并研究风/光配置容量对可靠性的影响。文献[5]通过模拟风电、光伏出力场景, 研究风电、光伏的波动性和负荷的不确定性对电力系统可靠性造成的影响。基于序贯蒙特卡洛模拟的可靠性评估算法可以计及变量的时序相关特性, 实现更精确的可靠性指标计算, 然而其计算需要大量的场景输入, 运算时间较长, 计算复杂度较高。随着大数据技术在电力行业的应用与发展, 应用场景约简技术实现时间序列聚合, 选取较少的典型场景代表原始数据集, 可以在简化计算的同时保证计算结果的精度^[6]。

现有场景约简技术主要包含 K-means 聚类、层次聚类法、模糊 C 均值聚类及谱聚类等。文献[7]基于 K-means 聚类对风电场历史故障数据进行分析, 通过典型场景刻画了故障历史数据特征和类别属性, 验证了采用典型场景集代表全数据集的可行性。文献[8]采用最近邻聚类方法对场景进行削减, 计算速度快, 简便易行。但由于以上算法的原理较为简单, 应用于高维数据时存在聚类结果不稳定、平滑原始时序数据的波动特征等弊端。文献[9]根

收稿日期: 2021-01-07

基金项目: 国家电网公司总部科技项目: 基于大数据的电力系统可靠性评估关键技术研究及应用(5100-201999332A-0-0-00)

作者简介: 李春燕(1976-), 女, 重庆籍, 副教授, 博士生导师, 博士, 研究方向为可再生能源消纳和需求响应研究; 赵晨宇(1995-), 女, 山西籍, 硕士研究生, 研究方向为高比例可再生能源参与的电力市场建模。

据层次聚类得到各集合的容量,对每个集合的代表场景进行加权,获得能够反映输入数据波动特征的典型场景集。虽然层次聚类可以获得相对稳定的典型场景集,但该算法时间复杂度大,结果依赖合并点和分裂点的选择。

对可靠性评估的典型场景研究目前主要面向元件状态场景和风光荷输入场景。由于风光荷场景具有强不确定性,对风光荷典型场景进行筛选,以典型场景代表全场景可以一定程度上描述其不确定性。考虑风电、光伏出力与负荷之间的相关性具有重要价值。文献[10]提出一种基于改进 K-means 的典型场景集选取方法,对包括风功率和电力负荷的时序数据集进行聚类分析和场景优选,计及了负荷和风电出力相关性。虽然该算法实现了风-荷联合典型场景选取,但 K-means 算法处理高维数据时效果有限,而考虑光伏出力会进一步增加数据的维度。

上述研究为典型场景的提取提供了思路,但应用于电力系统可靠性评估时,仍存在两方面问题:①风电、光伏出力具有较强的波动性,且风电具有反调峰特性,风电、光伏和负荷三类数据存在明显的时序相关性,因此需要对风-光-荷高维数据进行场景聚类,同时保留时序特性;②聚类算法会损失极端场景。而在可靠性评估过程中,高负荷水平、低电源出力等极端场景对可靠性评估结果有不可忽视的影响。

基于上述研究,本文提出一种面向可靠性评估的风-光-荷典型场景集选取方法,对包括风电出力、光伏出力和电力负荷的时序数据集进行降维和两阶段聚类,生成考虑风、光、负荷相关性的分层时序典型场景集,对场景集进行优选,得到考虑极端场景对可靠性评估影响的典型场景集。应用于电力系统可靠性评估当中,以传统时序负荷仿真法计算结果作为参照,对比分析算法的计算速度及各项指标误差。

2 问题描述和研究思路

为实现面向可靠性评估的风-光-荷典型场景生成,目前主要存在以下 3 个问题:

(1)为使聚类结果适用于可靠性评估,需要选择合适的曲线特征对风-光-荷场景进行聚类。

(2)为保留风-光-荷场景的风、光、负荷相关性及时序特征,需要对高维数据进行聚类,现有聚类算法对高维数据直接聚类效果并不理想,需要采用多次聚类和降维算法改进聚类算法。

(3)现有研究多采用聚类中心作为典型场景,由于聚类算法的特性,极端场景一般不作为聚类中心出现,但电力系统可靠性评估对负荷水平较高的场景非常敏感,忽略极端场景会明显影响评估结果的精度。因此需要计及极端场景的典型场景优选方法。

为解决以上问题,本文基于 DBSCAN 和 K-means 聚类提出了一种适用于含新能源电力系统可靠性评估的典型场景生成的两阶段聚类方法。首先,选用净负荷持续曲线作为风-光-荷场景特征,基于分层抽样的思想,采用 DBSCAN 聚类算法将风-光-荷场景分层;然后采用 SAX 降维算法对原始风-光-荷时序曲线进行降维,使用 K-means 聚类算法对每层曲线进行分别聚类,得到保留时序特征的风-光-荷场景聚类结果。最后依据基于核密度估计发电系统状态的可靠性评估结果指标对各层曲线集进行场景优选,生成计及极端场景影响的时序风-光-荷典型场景集,应用于含新能源的电力系统可靠性评估,具体流程如图 1 所示。

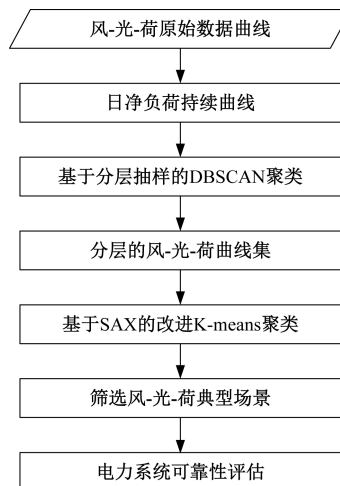


图 1 基于分层抽样的典型场景生成方法

Fig.1 Stratified sampling based typical scenario generation method

3 基于 DBSCAN 和改进 K-means 的风-光-荷典型场景生成

运用典型场景进行可靠性评估可以大大削减计算量,提高计算效率。典型场景生成的关键点和难点在于如何对大量原始场景进行削减,实现减小计算负担的同时保证削减后的场景集尽可能逼近原始场景。本文目的是寻找适用于可靠性评估的典型场景,因此需要对原始场景中可靠性评估相关特征进行发掘,利用特征筛选典型场景。

3.1 基于 DBSCAN 分层的日净负荷持续曲线分类

分层抽样法是将总体单位按其属性特征分成若干层,在层中按一定比例随机抽取样本单位。本方法通过划类分层,增大了各层级中单位间的共同性,使得样本代表性较好,抽样误差较小。因此更适用于总体情况复杂,各单位之间差异较大,单位较多的情况。

电力系统的失负荷量与失负荷概率与系统某时刻的净负荷水平有着不可分割的关系。净负荷指某一时刻系统的用电负荷与同时刻新能源出力差值,由于某一时刻可靠性评估指标与该时刻负荷水平直接相关,净负荷值越大,失负荷概率越高,切负荷量也越大。为舍弃负荷曲线冗余特征,提高计算效率,本文首先基于密度对日净负荷持续曲线进行聚类,得到依据负荷水平的负荷样本分层,从而确定负荷输入的优先级,约简对可靠性评估指标影响较小的场景,提高计算效率。

在将每天各时刻的净负荷按照从大到小的顺序进行重排形成日净负荷持续曲线后,本文引入并改进基于密度的有噪应用中的空间聚类算法(Density-Based Spatial Clustering of Applications with Noise, DBSCAN)算法对日净负荷持续曲线进行聚类。DBSCAN 算法是一种基于密度的聚类算法。此算法不需要预先指定簇的个数,而是根据样本数据点分布密度进行聚类,因此可以识别数据量稀少的离群值,应用在场景削减中可以保留极端场景。

DBSCAN 算法需要设定两个重要参数:DB 领域半径值(Epsilon, Eps)和领域密度阈值(Minimum Points, MinPts)。 Eps 表示个体之间距离临界值, $MinPts$ 表示临界距离半径中个体数量的临界值。传统 DBSCAN 算法中 Eps 和 $MinPts$ 由经验设置再根据聚类结果进行调整,存在较大的盲目性。本文采用绘制 k-距离曲线方法^[13],选取 k-距离曲线图明显拐点位置为聚类参数确定 Eps 。 $MinPts$ 的选取遵从原则:

$$MinPts = \frac{1}{N_p} \sum_{i=1}^{N_p} s_i \quad (1)$$

式中, s_i 为点 i 的 Eps 领域内个体的数量; N_p 为 Eps 领域个数。对于样本数据集 $D=(X_1, X_2, \dots, X_m)$,其中 X 为风-光-负荷原始场景数据, m 为原始场景个数,DBSCAN 算法流程如图 2 所示。

利用上述算法对日净负荷持续曲线进行聚类实现对风-光-荷原始场景集 D 的分层,得到分层场景

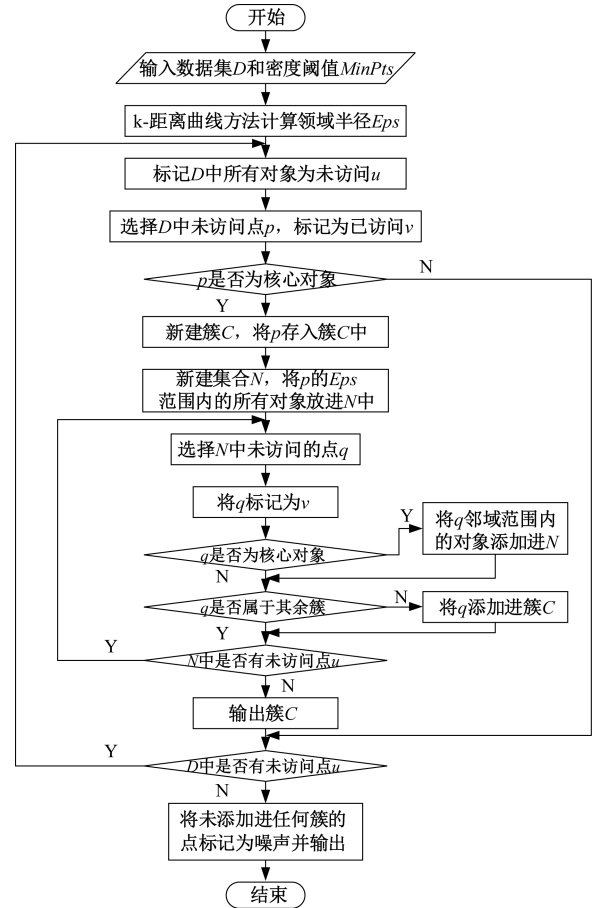


图 2 DBSCAN 算法流程图

Fig.2 Flow chart of DBSCAN algorithm

集 $\{D_1, D_2, \dots, D_s\}$, 所得不同场景集在峰值负荷大小及负荷分布上有明显分级。

3.2 基于优化 SAX 的 K-means 聚类算法

基于分层场景集 $\{D_1, D_2, \dots, D_s\}$, 本文采用改进的 K-means 聚类算法对每一层风-光-荷场景进行削减, 得到保留时序特征的分层典型场景集。

K-means 聚类算法是一种简单、高效的无监督聚类分析算法。随机选取空间中 k 个点为中心进行聚类, 计算每个对象与 k 个点中最靠近聚类中心的对象归类。通过迭代的方法, 逐次更新各聚类中心的值, 直至所有数据点到其所在类中心的距离和达到最小值。K-means 聚类使用方法简便, 且具有较好的计算效率, 但该算法对高维数据聚类效果不理想。

为了实现对高维海量的风电、光伏和负荷数据聚类, 同时保留其时序特征, 需要对数据进行降维处理。符号聚合近似 (Symbolic Aggregate Approximation, SAX) 是一种将连续时间序列转化为离散字符

变量的算法,可以有效地降维、降噪,且结果更为直观,具有处理速度快,便于状态分析等优势,在异常数据检测,模式识别等领域应用广泛^[11,12]。本文主要介绍 SAX 参数优化的相关部分。

以图 3 所示的曲线为例,对其进行分段数 $w = 3$,字符数 $l = 3$ 的 SAX 转化。首先将具有 n 维特征的单条原始负荷曲线 $\mathbf{X} = [x_1, x_2, \dots, x_n]$ 分段聚合近似(Piecewise Aggregate Approximation, PAA)为 w

段 $\bar{\mathbf{X}} = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_w]$,按式(2)计算第 i 个分段值;再利用式(3)对其字符化,其中 $l'_1 = a, l'_2 = b, l'_3 = c$,以此类推。

$$\bar{x}_i = \frac{w}{n} \sum_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i} x_j \quad (2)$$

$$L_i = l'_j, \text{ if } \beta_{j-1} \leq \bar{x}_i \leq \beta_j \quad (3)$$

式中, β 为字符断点值; L 为负荷曲线 \mathbf{X} 转化后的离散状态表示。

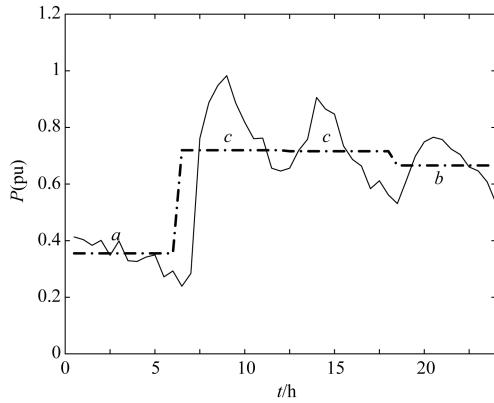


图 3 风电出力曲线的 SAX 表达

Fig.3 SAX expression of WP curve

本文从准确度 A 、信息量 E 、化简率 R 三个角度,综合考虑负荷曲线 SAX 表达的效果。其含义如下:准确度 A 反映分段后的风-光-荷曲线表征原曲线的能力;信息量 E 采用信息熵衡量分段后的曲线还原原曲线的能力,信息熵越小,则通过现有信息进行预测时的确定性越大,其所含信息量越大;化简率 R 则表征原始曲线的被压缩程度。通过三个指标对算法效果进行评估,达到综合效果最优,即得到最优的风-光-荷曲线降维表达。

将风-光-荷曲线 SAX 表达的优化过程转化为多目标优化问题,其中目标函数为:

$$Objective = \max(A, E, R) \quad (4)$$

其中

$$A = \frac{\text{cov}(\bar{X}_i, X_i) + 1}{2} \quad (5)$$

$$E = \sum_{i=1}^w p_i \log_2 p_i \quad (6)$$

$$R = \mu \left(1 - \frac{l}{l_m}\right) + (1 - \mu) \left(1 - \frac{w}{w_m}\right) \quad (7)$$

$$2 \leq l \leq l_m \quad (8)$$

$$2 \leq w \leq w_m \quad (9)$$

式中, $\text{cov}(\bar{X}_i, X_i)$ 代表负荷曲线 PAA 分段近似后的值 \bar{X}_i 与原始负荷曲线 X_i 的相关系数,由于维度不同,本文采用 \bar{X}_i 进行 spline 插值后形成与 X_i 相同维度的序列再进行相关系数计算; p_i 为字符 i 在负荷 X_i 出现的概率; l_m 为字符库中的字符总数; w_m 为设置的最大分段数,本文取 $l_m = w_m = 10$; μ 为两个参数化简的权重系数,本文取 $\mu = 0.5$ 。

针对多目标整数规划问题,相较于其他算法,粒子群算法更为简单有效,具有精度高、收敛快等优势^[14]。本文采用基于模拟退火的粒子群算法,其在搜索过程中具有概率突跳的能力,能够有效地避免搜索过程陷入局部极小解的情形^[15]。

改进后的参数优化算法流程主要有以下 7 个步骤:

Step 1: 输入粒子数目 N , 学习因子 c_1, c_2 , 退火常数 λ , 最大迭代次数 M , 初始温度 T_0 ;

Step 2: 随机初始化种群位置和速度;

Step 3: 利用目标函数计算每个粒子适应度并记录最优个体 P_g ;

Step 4: 将 T 赋值为初始温度 T_0 , i 赋值为 1;

Step 5: 确定当前温度,并通过轮盘赌策略确定全局最优替代值 P'_g ;

Step 6: 更新粒子位置和速度,计算新目标值并更新各粒子适应度值;

Step 7: 判断 i 是否大于等于 M 或满足阈值,若符合条件,结束算法;若不符合, i 自增 1 后返回 Step 5,直到 i 满足退出条件。

其中粒子数 $N = 8$ 。通过随机初始化生成代表分段数和字符数的粒子集后,根据目标函数和自变量约束进行优化。在退火过程中不仅接受较优的解,也以一定的概率接受较差的解,同时这种概率受到温度参数的控制。

3.3 基于核密度估计的聚类中心典型场景筛选

3.2 节所述算法所得聚类中心为数据分布中

心,为考虑可靠性评估相关因素,需要依据场景对应的可靠性指标在聚类所得场景集中进行典型场景筛选。

本文基于核密度估计参数估计建立元件故障模型^[16]拟合发电机停运容量分布,对 IEEE RTS-79 标准测试系统进行发电系统快速可靠性评估得到第 i 类对应的日平均停电时间期望为:

$$LOLE_i = \sum_{j=1}^{n_i} LOLE_{ij} / n_i \quad (10)$$

式中, $LOLE_{ij}$ 为第 i 类第 j 天对应的日停电时间期望; n_i 为第 i 类所包含风-光-荷联合曲线条数。

$$|LOLE_i, LOLE_{ij'}| = \min |LOLE_i, LOLE_{ij}| \quad (11)$$

则第 i 类场景集典型场景选取 j' 作为此类的典型场景输入可靠性评估算法。

4 典型场景生成算法总体流程

本文首先选取 DBSCAN 聚类算法对日净负荷持续曲线进行聚类,能够实现对大量数据的快速初步聚类,得到不同净负荷水平的场景集。然后采用改进 K-means 聚类算法,以对初步聚类结果进行高质量的二次聚类,同时保留风-光-荷出力的时序特征,最后对聚类结果进行筛选,得到适用于可靠性评估的风-光-荷典型场景集。整体算法流程如图 4 所示。

在聚类评价中采用 DB (Davies-Bouldin) 指标作为算法评价指标, DB 指标计算公式如式 (12) 所示,其值越小,则类间的相似度越低,类中相似度越高,聚类效果越好。

$$DB = \frac{1}{c} \sum_{i=1}^c \max_{j=1 \dots k, j \neq i} \left(\frac{W_i + W_j}{C_{ij}} \right) \quad (12)$$

式中, c 为聚类数; W_i 为 i 类内数据点到聚类中心 C_i 的平均距离; W_j 为 j 类内数据点到聚类中心 C_j 的平均距离; C_{ij} 为聚类中心 i 与 j 之间的距离。

5 算例分析

5.1 算例介绍

本文算例采用的平台为 1.7 GHz CPU 和 8 GB 内存的计算机,使用 MATLAB 2014a 进行算法仿真。为减少数据异常及缺失对数据分布的影响,保证聚类结果的正确性,本文已去除全零负荷并采用分段多项式拟合对曲线进行平滑处理,详见文献[17]。

5.2 算法性能测试结果

为测试算法的高效性,本文采用 IEEE 8 736 h

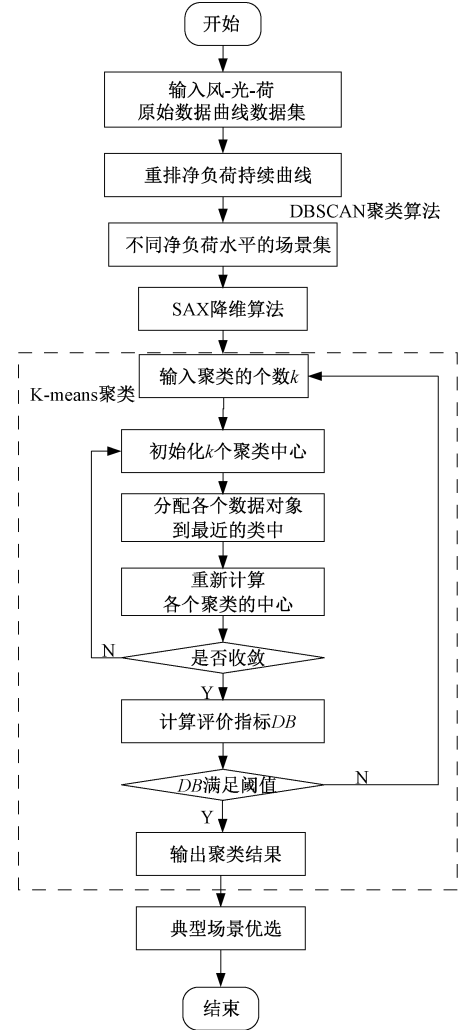


图 4 基于 DBSCAN 和改进 K-means 的两阶段聚类流程图

Fig.4 Flowchart of two-stage clustering based on DBSCAN and improved K-means

标准负荷数据^[18],共计 364 条负荷曲线进行聚类,对以下 4 种算法进行对比:

方法 1:未改进的 K-means 聚类算法;

方法 2:AP 聚类算法;

方法 3:文献[13]中自 FCM 算法;

方法 4:本文所述 DBSCAN 和 K-means 两阶段聚类算法。

为了算法的统一性,方法 1,2,3 均为在 Matlab 自带 K-means 算法的基础上进行实现。由于传统 K-means 算法需要事先输入聚类数目,为了方便比较,令方法 1 和方法 3 的聚类数 $c=15$ 。方法 4 两阶段聚类最终聚类数为 15。对上述四种算法的 DB 值及总计算时间分别进行计算,得到结果如表 1 所示。

表 1 算法结果对比

Tab.1 Algorithm performance comparison

算法	方法 1	方法 2	方法 3	方法 4
DB	3.128 1	3.019 7	2.450 9	2.124 0
聚类数/个	15	11	15	15
降维时间/s	0	0	0	0.015
总计算时间/s	16.062	63.031	26.374	46.375

方法 1 由于算法复杂度低,计算时间最短,然而传统 K-means 算法在数据迭代过程中收敛慢,在本文数据下,达到聚类迭代最大次数时,聚类中心尚未达到稳定,导致 K-means 算法的 DB 指标值高于其他算法;方法 2 中 AP 聚类通过相似度矩阵进行聚类,具有自适应的聚类数, DB 指标优于传统 K-means 算法,但此类算法复杂度较高,所需总计算时间明显长于其他算法。方法 3 所需计算时间短, DB 指标更优,算法表现上较好。方法 4 改进之后由于增加了一次聚类及降维过程,总计算时间有一定增长,但仍小于方法 2,且 DB 指标明显优于其他算法。结合 DB 指标和计算时间看,方法 3 和方法 4 较有优势。

为测试算法的有效性,本文将以上四种聚类算法的结果作为典型场景输入 IEEE RTS-79 标准系统中,利用枚举法进行发电系统可靠性评估,所得结果如表 2 所示。

表 2 基于状态枚举法的发电系统可靠性评估结果

Tab.2 Reliability evaluation result of power generation system based on enumeration method

算法	全部输入	方法 1	方法 2	方法 3	方法 4
<i>EENS</i> (MW·h)/a	1 120.9	991.2	803.6	870.3	1 066.2

表 2 中, *EENS* 指标 (Expected Energy Not Supplied) 表示测试系统运行一年的失负荷量。由表 2 可知,方法 2 和方法 3 所得 *EENS* 指标与全部场景输入所得差值非常大,方法 1 差值较小,且差值随聚类数的增大逐渐减小,方法 4 差值最小,原因主要是:①由于可靠性评估的失负荷状态主要发生在负荷水平较高的情况下,而在聚类过程中,负荷水平较高的场景往往是类别中的边缘点,通常不能作为聚类中心,因此以聚类中心作为典型场景评估所得年失负荷量会明显偏低;②方法 1、3、4 都需要人为指定聚类数目,聚类数的多少会明显影响可靠性评估的结果。

另外,方法 3 算法对满足正态分布的数据聚类

效果较好,但由于季节、天气等因素影响,用电数据并不呈正态分布,因此方法 3 用于用电数据聚类效果一般。而方法 4 算法简单,可以通过经验结合枚举法找出最优聚类数,且聚类结果较好。结合计算效率和有效性,方法 4 较有优势。

根据比较可以看出, DBSCAN 和改进 K-means 两阶段聚类算法在实际应用中,能够获得较高的聚类效率和较好的应用效果。

5.3 IEEE 标准系统测试结果

数据集来自比利时某电力公司官方网站 2017~2018 年负荷、风力发电数据及某光伏电站一年光伏发电上网数据^[19],选取负荷曲线,风电出力曲线及光伏出力曲线各 365 条,并对有效曲线进行了离差归一化处理。图 5 给出了数据集对应净负荷持续曲线 DBSCAN 聚类后的聚类中心。

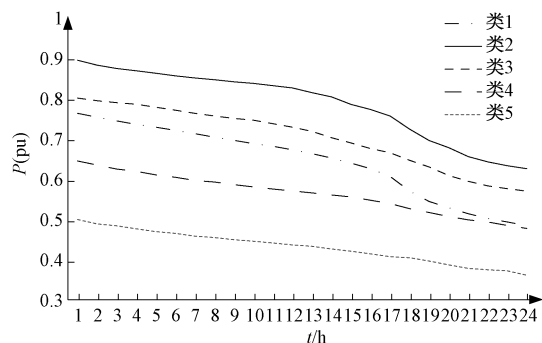


图 5 净负荷持续曲线聚类结果

Fig.5 Result of net load duration curve clustering

从图 5 可见,类 1 到类 5 的按峰值负荷大小有明显阶梯状分层,类 2 最大,类 5 最小。类 1、2、3 场景峰值负荷较大,用电峰谷差也较为明显,类 4、5 场景负荷水平较低,用电情况较为稳定。以上结果表明采用净负荷持续曲线的 DBSCAN 聚类可以对风光-荷场景实现较好的分层效果。

将各层聚类中心代入 IEEE RBTS-6 测试系统采用枚举法进行发电系统可靠性评估计算,得到可靠性指标如表 3 所示。

表 3 各层聚类中心可靠性评估结果

Tab.3 Reliability indices of clustering centers

指标	<i>LOLP</i>	<i>LOLE</i> /((MW·h)/天)
类 1	0.007 397	21.274 6
类 2	0.021 043	84.289 1
类 3	0.001 923	5.138 15
类 4	0.000 105	0.175 541
类 5	0	0

表3中, $LOLP$ 指标(Loss of Load Probability)为测试系统一天的失负荷概率, $LOLE$ 指标(Loss of Load Expectation)为测试系统一天的失负荷量。由表3可知,类2聚类中心作为输入计算得失负荷量最大,类5聚类中心计算得失负荷量最小。净负荷持续曲线峰值越大,计算得失负荷量越大。由此可见,净负荷大小与失负荷量呈正相关,本文对净负荷持续曲线分层对后续可靠性评估计算正确有效。

表3中分类结果的原始数据集即为根据净负荷持续曲线分层后的场景集,利用改进 K-means 分别对每一层进行聚类,经过多次试验,本文设 $k=3$ 。聚类结果为5层,每层3类,共15类场景集。对所得场景集进行筛选,用基于核密度估计的可靠性评估算法在每一类中筛选出最终适用于可靠性评估的典型场景,舍弃部分相似场景后结果如图6所示,各典型场景出现概率如表4所示。

表4 各典型场景对应概率

Tab.4 Probability of typical scenarios

场景	1	2	3	4	5	6	7
概率	0.082	0.096	0.074	0.186	0.041	0.071	0.003
场景	8	9	10	11	12	13	
概率	0.126	0.074	0.104	0.033	0.052	0.058	

本文对 IEEERTS-79 标准测试系统进行改进,在节点7添加风电出力输入,节点22添加光伏出力输入,该系统接线图如附图1所示。改进后测试系统上采用图6中所示的时序典型场景及各场景对应概率,依概率进行基于序贯蒙特卡洛抽样的可靠性评估,得到的可靠性指标与输入一年365天风-光-荷数据对比如表5所示。

表5 典型场景与全场景输入的序贯蒙特卡洛仿真结果对比

Tab.5 Comparison of sequential Monte Carlo simulation results using typical scenarios and full scenarios

场景输入	全场景	典型场景
$LOLE/((MW \cdot h)/天)$	9.902 1	9.832 2
$LOLF/(次/年)$	1.265 3	1.261 4
$LOLP$	0.001 1	0.001 1
$EENS/((MW \cdot h)/年)$	1 275.1	1 239.8
运行时间/s	248 579	108 524

表5中, $LOLF$ 指标(Loss of Load Frequency)为电力不足频率,即单位时间内停电的次数。从表5可见,典型场景输入计算所得 $EENS$ 指标对比全场景输入误差率为2.82%,较其他聚类算法所得典型

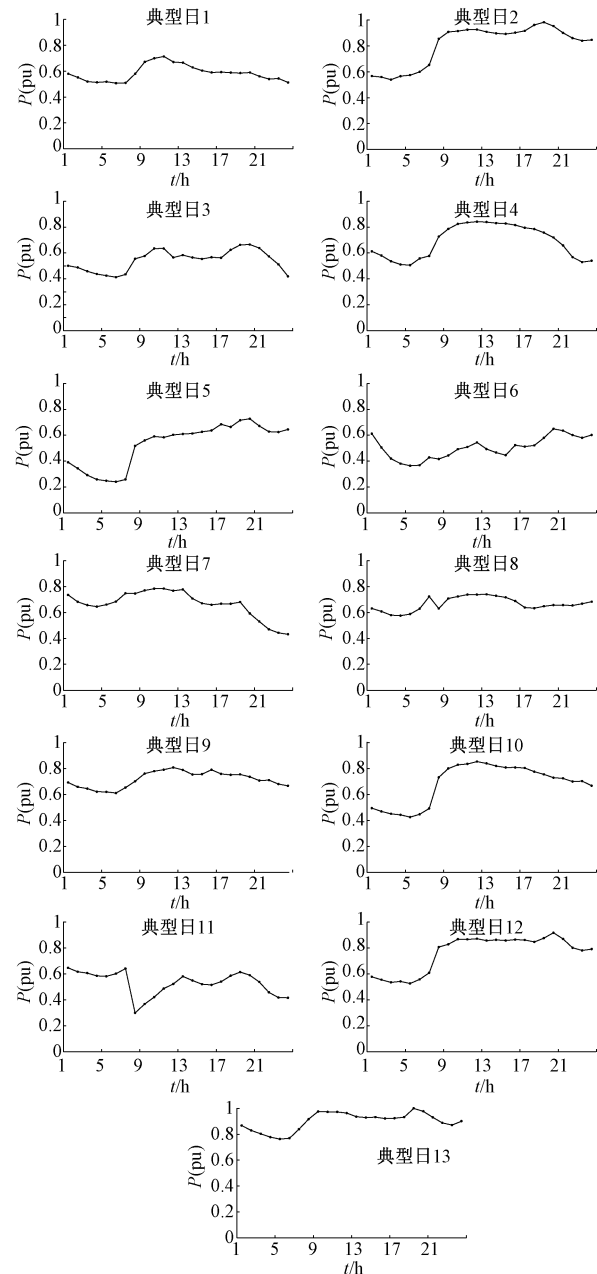


图6 DBSCAN 和 K-means 两阶段聚类后筛选所得典型日

Fig.6 Typical scenarios from DBSCAN and K-means two-stage clustering

场景输入误差较小,误差不为零的原因是聚类算法原理导致聚类过程中可能会丢失对可靠性评估指标影响较大的边缘点,采用本文算法选取典型场景可以一定程度减少此类误差,但不能完全消除。而运用典型场景进行可靠性评估较原方法计算时间缩短60%,显著提升了计算效率。由此可见,利用本文所提出的 DBSCAN 和改进 K-means 聚类典型场景生成方法所得的时序典型场景适用于可靠性评估计算,兼具高效性与准确性。

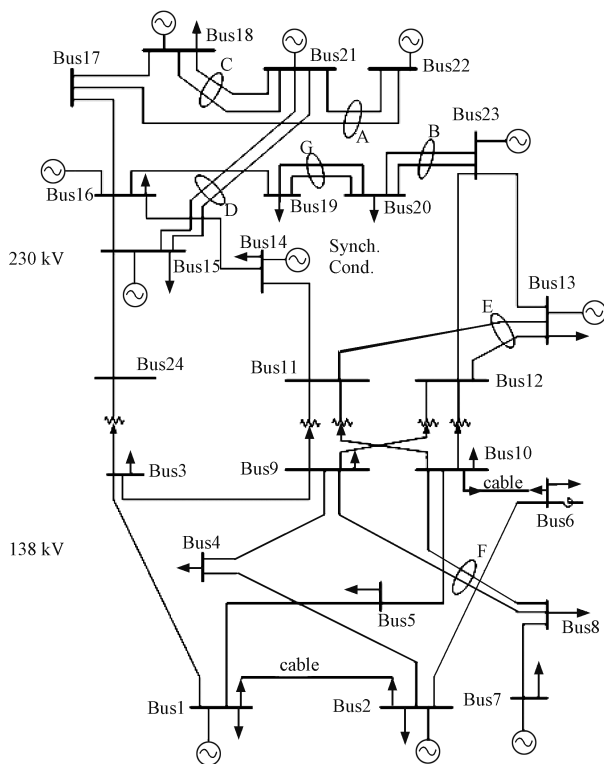
6 结论

本文提出了一种适用于含可再生能源电力系统可靠性评估的 DBSCAN 和改进 K-means 两阶段聚类典型场景生成方法。首先基于分层抽样的思想,采用密度聚类将净负荷持续曲线分层,然后采用 SAX 算法改进 K-means 聚类算法对风-光-荷原始曲线进行降维聚类,依据核密度估计可靠性评估 *LOLE* 指标对各层曲线集进行场景筛选,最终生成保留时序特征的风-光-荷典型场景。

该算法在保证聚类算法效率的同时,对高维数据聚类效果良好。最终所得时序风-光-荷典型场景用于电力系统序贯蒙特卡洛可靠性评估,显著提升算法效率的同时保证了可靠性评估的准确性。

场景约简技术作为一种稳定、高效的简化计算方法,在可靠性评估中具有很大的研究价值。下一步研究可以尝试关注极端场景,在聚类过程中保留边界点,细化高负荷水平场景的分类,从而进一步减小可靠性评估误差。

附录



附图1 IEEE RTS-79 标准测试系统接线图

App. Fig. 1 Diagram of IEEE RTS-79 reliability test system

参考文献 (References):

- [1] 蒋泽甫, 谢开贵, 胡博, 等 (Jiang Zefu, Xie Kaigui, Hu Bo, et al.). 风力发电系统可靠性评估解析模型 (Wind power generation analytical model for the reliability evaluation of power systems) [J]. 电力系统保护与控制 (Power System Protection and Control), 2012, 40 (21): 52-57, 95.
- [2] 汪海瑛, 白晓民, 许婧 (Wang Haiying, Bai Xiaomin, Xu Jing). 考虑风光储协调运行的可靠性评估 (Reliability assessment considering the coordination of wind power, solar energy and energy storage) [J]. 中国电机工程学报 (Proceedings of the CSEE), 2012, 32 (13): 13-20, 186.
- [3] Karki R. Reliability and cost evaluation of small isolated power systems containing photovoltaic and wind energy [D]. Saskatchewan: University of Saskatchewan, 2000.
- [4] 王凯, 栗文义, 李龙, 等 (Wang Kai, Su Wenyi, Li Long, et al.). 含风/光互补发电系统可靠性与经济性评估 (Fuzzy comprehensive evaluation of household electric energy consumption) [J]. 电工电能新技术 (Advanced Technology of Electrical Engineering and Energy), 2015, 34 (6): 52-56, 63.
- [5] Liu Yang, Nirmal-Kumar C Nair. A two-stage stochastic dynamic economic dispatch model considering wind uncertainty [J]. IEEE Transactions on Sustainable Energy, 2016, 7 (2): 819-829.
- [6] Björn B, Alexander K, Hagen S, et al. Time-series aggregation for synthesis problems by bounding error in the objective function [J]. Energy, 2017, 135: 900-912.
- [7] 卢志刚, 隋玉珊, 何守龙, 等 (Lu Zhigang, Sui Yushan, He Shoulong, et al.). 基于场景分析的含风电系统经济爬坡调度 (Economic/unit ramp dispatching for power grid integrated with wind power based on scenario analysis) [J]. 电工电能新技术 (Advanced Technology of Electrical Engineering and Energy), 2016, 35 (2): 38-43.
- [8] 赵国炫, 鲁文军, 喻锐, 等 (Zhao Guoxuan, Lu Wenjun, Yu Kun, et al.). 基于 K-means 聚类分析的风电场集电线路保护方法 (Protection method for wind farm collector line based on K-means clustering analysis) [J]. 电力系统及其自动化学报 (Proceedings of the CSU-EPSC), 2020, 32 (7): 39-46.
- [9] Nahmmacher P, Schmid E, Hirth L, et al. Carpe diem: A novel approach to select representative days for long-term power system modeling [J]. Energy, 2016, 112: 430-442.
- [10] 丁明, 解蛟龙, 刘新宇, 等 (Ding Ming, Xie Jiaolong, Liu Xinyu, et al.). 面向风电接纳能力评价的风资源/负荷典型场景集生成方法与应用 (The generation/typical scenario set generation method and application for wind power acceptance capacity evaluation) [J]. 电力系统及其自动化学报 (Proceedings of the CSU-EPSC), 2020, 32 (7): 39-46.

- tion method and application of wind resources/load typical scenario set for evaluation of wind power grid integration) [J]. 中国电机工程学报 (Proceedings of the CSEE), 2016, 36 (15): 4064-4072.
- [11] Lin J, Keogh E, Lonardi S, et al. A symbolic representation of time series, with implications for streaming algorithms [A]. ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery [C]. San Diego, California, 2003.
- [12] 李春燕, 蔡文悦, 赵溶生, 等 (Li Chunyan, Cai Wenye, Zhao Rongsheng, et al.). 基于优化 SAX 和带权负荷特性指标的 AP 聚类用户用电行为分析 (Customer behavior analysis based on affinity propagation algorithm with optimized SAX and weighted load characteristic indices) [J]. 电工技术学报 (Transactions of China Electrotechnical Society), 2019, 34 (S1): 368-377.
- [13] 李文杰, 闫世强, 蒋莹, 等 (Li Wenjie, Yan Shiqiang, Jiang Ying, et al.). 自适应确定 DBSCAN 算法参数的算法研究 (Research on method of self-adaptive determination of DBSCAN algorithm parameters) [J]. 计算机工程与应用 (Computer Engineering and Applications), 2019, 55 (5): 1-7, 148.
- [14] 孙建军, 张世泽, 曾梦迪, 等 (Sun Jianjun, Zhang Shize, Zeng Mengdi, et al.). 考虑分时电价的主动配电网柔性负荷多目标优化控制 (Multi-objective optimal control for flexible load in active distribution network considering time-of-use tariff) [J]. 电工技术学报 (Transactions of China Electrotechnical Society), 2018, 33 (2): 401-412.
- [15] 黄飞腾, 南余荣, 翁国庆, 等 (Huang Feiteng, Nan Yurong, Weng Guoqing, et al.). 计及监测可信度的扰动源定位粒子群算法 (Particle swarm optimization of disturbance source location with monitoring reliability) [J]. 电工技术学报 (Transactions of China Electrotechnical Society), 2016, 31 (16): 215-222.
- [16] 张夏菲 (Zhang Xiafei). 非参数核密度估计负荷模型在电网可靠性评估中的应用 (The application of non-parametric Kernel density estimation load model in power system reliability evaluation) [D]. 重庆: 重庆大学 (Chongqing: Chongqing University), 2010.
- [17] 龚钢军, 陈志敏, 陆俊, 等 (Gong Gangjun, Chen Zhimin, Lu Jun, et al.). 智能用电用户行为分析的聚类优选策略 (Clustering optimization strategy for electricity consumption behavior analysis in smart grid) [J]. 电力系统自动化 (Automation of Electric Power Systems), 2018, 42 (2): 58-63.
- [18] Grigg C, Wong P, Albrecht P, et al. The IEEE reliability test System-1996. A report prepared by the reliability test system task force of the application of probability methods subcommittee [J]. IEEE Transactions on Power Systems, 2002, 14 (3): 1010-1020.
- [19] <http://www.elia.be/en/grid-data> [EB/OL].

Two-stage clustering algorithm of typical wind-PV-load scenario generation for reliability evaluation

LI Chun-yan¹, ZHAO Chen-yu¹, HU Bo¹, CHEN Zheng-yu², LIAO Qing-long²,
WAN Ling-yun², XIE Kai-gui¹

(1. State Key Laboratory of Power Transmission Equipment & System Security and New Technology, Chongqing University, Chongqing 400044, China; 2. State Grid Chongqing Electric Power Research Institute, Chongqing 404100, China)

Abstract: To capture the uncertainty of renewable generations and load demand, reliability evaluation of power systems requires a large number of scenarios to ensure accurate results. Due to the computational cost issues, an efficient scenario reduction technique is crucial for the reliability evaluation. Traditional clustering algorithms are difficult to process high-dimensional data and can not retain extreme load scenarios, and the effect of generating typical scenarios for reliability evaluation is not ideal. In this paper, the DBSCAN&K-means double-stages clustering algorithm is proposed to generate the typical chronological scenario set for reliability evaluation, retaining extreme scenario and introducing dimensionality reduction algorithms to process high-dimensional time series data. Cased studies show that the algorithm in this paper is both efficient and effective, and the typical scenarios, used in power system reliability evaluation, can obtain more accurate reliability indices.

Key words: stratified sampling; improved K-means clustering; DBSCAN clustering; reliability evaluation; typical scenarios; sequential monte carlo simulation