

基于特性指标降维的日负荷曲线聚类分析

刘思¹, 李林芝¹, 吴浩¹, 孙维真², 傅旭华³, 叶承晋³, 黄民翔¹

(1. 浙江大学电气工程学院, 浙江省杭州市 310027;

2. 浙江电力调度控制中心, 浙江省杭州市 310007;

3. 浙江省电力公司经济技术研究院, 浙江省杭州市 310008)

Cluster Analysis of Daily Load Curves Using Load Pattern Indexes to Reduce Dimensions

LIU Si¹, LI Linzhi¹, WU Hao¹, SUN Weizhen², FU Xuhua³, YE Chengjin³, HUANG Minxiang¹

(1.College of Electrical Engineering, Zhejiang University, Hangzhou 310027, Zhejiang Province, China;

2.Zhejiang Electrical Power Dispatch and Control Center, Hangzhou 310007, Zhejiang Province,China;

3. Economic Research Institute of State Grid Zhejiang Electric Power Company, Hangzhou 310008, Zhejiang Province,China)

ABSTRACT: A new method, Pattern Index Clustering (PIC), originated from clustering techniques in data mining, is proposed for daily load curve clustering in this paper. PIC method reduces dimensions of sequential load curves with six load pattern indexes, such as load factor and daily peak-valley ratio. Weights of the indexes are gradually corrected with clustering-validity-based Delphi method. Hence, the method clusters curves with weighted Euclidean distance as similarity measurement, resulting in a performance better than classic clustering algorithms. Numerical examples show that PIC method has shorter run time, stronger robustness and better clustered load curves. Therefore it reflects characteristics of typical load curves more directly.

KEY WORDS: load curve classification; load pattern indexes; dimension reduction; weighted Euclidean distance; clustering validity

摘要: 运用数据挖掘中的聚类技术对电力系统日负荷曲线进行分析, 提出一种基于特性指标降维的日负荷曲线聚类方法——特性指标聚类(pattern index clustering, PIC), 通过负荷率、日峰谷差率等 6 个日负荷特性指标对日负荷曲线进行降维处理, 利用基于聚类有效性修正的德尔菲方法配置各指标权重, 以加权欧式距离作为相似性判据, 对日负荷曲线进行聚类。算例结果表明所提方法运行时间短, 鲁棒性好, 提高了负荷曲线聚类质量, 能直观反映典型负荷曲线的特点。

关键词: 负荷曲线分类; 负荷特性指标; 降维; 加权欧式距离; 聚类有效性

DOI: 10.13335/j.1000-3673.pst.2016.03.020

基金项目: 国家电网公司科技项目(面向城乡差异化的空间负荷预测及统一应用模型研究, No.5211JY150016)。

Project Supported by State Grid Science & Technology Project (5211JY150016)。

0 引言

分布式能源的广泛接入, 电动汽车等多元化负荷的大量投入, 在优化能源结构的同时, 也使得局部区域出现负荷紧张、可再生能源消纳困难等问题^[1-3]。通过数据挖掘中的聚类技术对电力用户负荷曲线进行分类, 依据其不同的用电特性, 在电力市场中引入合理的需求侧响应(demand response, DR)^[4-6], 可优化区域峰谷差, 平衡电网供需缺口, 有助于解决以上问题^[7]。

聚类算法应用于负荷曲线分类, 主要包括 k-means、层次聚类、模糊 C 均值(fuzzy C-means, FCM)、自组织映射神经网络(self-organizing map, SOM)等几种算法(含改进算法)及其组合^[8-11], 通常以负荷曲线每个采集点的负荷数据作为输入, 对数据规范化处理后, 直接或间接以距离作为相似性判据, 对用户进行分类。

文献[12]指出, 以距离作为相似性判据, 其实质是对象间的相近程度, 然而相似性还依赖于对象内在性质。对于作为时间序列的负荷曲线来说, 时序负荷数据易受气温气候、收入、电价政策等多种因素影响, 这些影响结果作为时序数据的内在特征, 无法通过距离得到充分反映, 不能完全保证时间序列的形态或轮廓的相似性。并且, 对于日负荷曲线这类有明显负荷形状的曲线, 在高维情况下会表现出不理想的等距性^[13], 文献[14]指出对于数据集集合中的对象, 维数越高, 距离测度的意义就越小。针对上述问题, 文献[15]利用高斯滤波器对负荷曲线进行平滑处理, 弱化了负荷数据中的“噪声”干扰, 虽提升了聚类质量, 但使得负荷数据失真, 且

聚类结果与滤波器选择和负荷数据质量密切相关。文献[16]采用截断的离散傅里叶变换系数来对负荷曲线降维,有一定改进,但计算时间偏长。文献[17-18]分别利用负荷序列的差分方程和夹逼准则改进相似性判据,取得了一定效果,但增大了计算的复杂度。

综合考虑上述情况后,为充分反映负荷间的相似性,兼顾运算效率,本文提出一种基于特性指标降维的日负荷曲线聚类方法——特性指标聚类(pattern index clustering, PIC),首先利用物理意义明确的日负荷特性指标对时序负荷曲线进行特征降维,然后提出基于聚类有效性修正的德尔夫法,对日负荷特性指标进行权重配置,最后以加权欧式距离为相似性判据,对日负荷曲线进行分类。算例表明,该算法运行时间短,鲁棒性好,能够提供更为合理的用户分类,便于充分掌握用户的负荷特性。

1 负荷特性指标选取

电力系统负荷特性是指用电负荷的特点和性质,可通过相关特性指标来体现。目前,电力系统负荷特性指标在国际上还没有统一的标准,我国常使用负荷率、峰谷差、平均负荷等 15 个指标^[19]分析各类用电负荷的特点和性质。

在年、月、日等各时间跨度的负荷特性分析中,日负荷特性因数据量巨大,异常数据、坏数据、畸变数据等干扰因素多,易受气温气候、生活习惯等的影响,分析难度最大。文献[20]从全天、峰期、平期、谷期 4 个角度,选取 6 种常用日负荷特性指标,较全面地反映了各类用户的用电特性。各指标定义和物理意义如表 1 所示。

本文将表 1 所列 6 个特性指标作为数据降维的依据。由表 1,指标 a_1 — a_3 从不同方面反映用户全

天的用电特性, a_4 — a_6 从不同时段反映用户的用电特性,各指标侧重角度和物理意义不同,用来反映用户用电特性时其重要程度也不同。记 6 维日负荷特性指标权重向量为 W , $W = [w_1, w_2, w_3, w_4, w_5, w_6]$, 满足 $\sum_{j=1}^6 w_j = 1$ 。本文所提权重配置方法详见 2.4 节。

2 基于特性指标降维的 PIC 算法

2.1 数据预处理

1) 数据选择。从相关单位得到 n 条日负荷曲线,设每条日负荷曲线有 q 个量测数据, S 为 n 条日负荷曲线构成的 $n \times q$ 阶初始负荷曲线矩阵。

2) 数据清理。首先对 S 中每条负荷曲线的负荷数据进行检查,查找缺失和异常数据(主要包括负荷骤降、暴增或为负等数据)。当某负荷曲线的数据缺失量和异常量达到采集量的 10% 或以上,认为该曲线无效,记剔除的无效曲线总数为 λ 。然后,利用平滑修正公式对该点的数据值进行修正和替代。设第 i 条负荷曲线的第 j 个采集点的负荷数据 $P_{i,j}$ 为缺失或异常数据,则修正值 $P'_{i,j}$ 为

$$P'_{i,j} = \frac{\sum_{k=1}^{k_1} P_{i,j-k} + \sum_{l=1}^{l_1} P_{i,j+l}}{k_1 + l_1} \quad (1)$$

式中: k 表示向前采集; l 表示向后采集; k_1 、 l_1 视具体情况而定,一般可取 5~10。经数据清理后,记 X 为 $(n-\lambda)$ 条有效日负荷曲线构成的 $(n-\lambda) \times q$ 阶有效负荷曲线矩阵。

3) 数据降维。利用表 1 中 6 个日负荷特性指标对有效负荷曲线矩阵 X 进行特征降维,得 $(n-\lambda) \times 6$ 阶特征降维矩阵,记为 Y 。

2.2 PIC 算法处理过程

PIC 算法以特征降维矩阵 Y 为输入,以加权欧式距离为相似性判据,进行聚类,其处理过程如下:

1) 设 k 为聚类数,在特征降维矩阵 Y 中随机选择 k 个样本,将 k 个样本值分别赋给初始聚类中心 $\bar{m}_j (j=1, 2, \dots, k)$, \bar{m}_j 为第 j 类聚类中心, $\bar{m}_j = [\bar{m}_{j1}, \bar{m}_{j2}, \bar{m}_{j3}, \bar{m}_{j4}, \bar{m}_{j5}, \bar{m}_{j6}]$ 。

2) 设 y_i 为第 i 个样本, $y_i = [y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5}, y_{i6}]$ 对特征降维矩阵 Y 中的所有 $y_i (i=1, 2, \dots, p)$ 依次计算到各类中心 \bar{m}_j 的加权欧式距离。

$$d(i, j) = \sqrt{w_1 |y_{i1} - \bar{m}_{j1}|^2 + w_2 |y_{i2} - \bar{m}_{j2}|^2 + \dots + w_6 |y_{i6} - \bar{m}_{j6}|^2} \quad (2)$$

表 1 日负荷特性指标
Tab.1 Daily load pattern indexes

时段	指标	定义	物理意义
全天 00:00—24:00	负荷率	$a_1 = P_{av} / P_{max}$	反映全天负荷变化
	最高利用小时率	$a_2 = P_{sum} / (24P_{max})$	反映时间利用效率
	日峰谷差率	$a_3 = \frac{P_{max} - P_{min}}{P_{max}}$	反映电网调峰能力
峰期 08:00—11:00 18:00—21:00	峰期负载率	$a_4 = P_{av,peak} / P_{av}$	反映峰期负荷变化
	平期负载率	$a_5 = P_{av,sh} / P_{av}$	反映平期负荷变化
谷期 22:00—24:00 00:00—06:00	谷期负载率	$a_6 = P_{av,val} / P_{av}$	反映谷期负荷变化

注: P 代表负荷,下标 sum、av、max、min 分别代表总量、均值、最大值、最小值, peak、val、sh 分别代表峰期、谷期、平期。

3) 找出所有 y_i 关于 $\bar{m}_j (j=1,2,\dots,k)$ 的最小距离 $\min(d(i,j))$, 将 y_i 归入到和 \bar{m}_j 距离最小的类中。

4) 设 n_j 是第 j 类样本的个数, $y_{i,j}$ 为第 j 类中第 i 个样本, 重新计算各类聚类中心。

$$\bar{m}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{i,j} \quad (j=1,2,\dots,k) \quad (3)$$

5) 设 t 为迭代次数, 按下式计算 Y 中所有样本的平方误差 $J_c(t)$, 并与前一次误差 $J_c(t-1)$ 比较。

$$J_c = \sum_{j=1}^k \sum_{i=1}^{n_j} \|y_{i,j} - \bar{m}_j\|^2 \quad (4)$$

6) 若 $J_c(t) - J_c(t-1) < 0$ 转 2), 否则算法结束。

2.3 聚类有效性检验

聚类有效性是指对聚类结果进行评价, 以确定最适合特定数据集的划分^[21]。Silhouette 指标^[22-23]是一种反映类内紧密性和类间分离性测度的常用有效性指标, 当聚类数未知时, 可用于确定最佳聚类数和评价聚类质量。其定义如下:

设 p 个对象的数据集 D 被划分为 k 类, 对于第 j 类中第 i 个样本, 定义 $\hat{d}_a(i)$ 为样本 i 与类内所有其他样本的平均距离, 用以表征类内紧密程度, 该值越小, 类内越紧凑; 定义 $\tilde{d}_b(i)$ 为样本 i 到非同类所有样本的最小平均距离, 用以表征类间分散程度, 该值越大, 类间越分散。则样本 i 的 Silhouette 指标 $I_{\text{Sil}}(i)$ 定义为

$$I_{\text{Sil}}(i) = \frac{\tilde{d}_b(i) - \hat{d}_a(i)}{\max(\hat{d}_a(i), \tilde{d}_b(i))} \quad (5)$$

样本 i 的 Silhouette 指标值 $I_{\text{Sil}}(i)$ 在 $[-1,1]$ 范围内变动。 $\hat{d}_a(i)$ 越小, $\tilde{d}_b(i)$ 越大, 此时 $I_{\text{Sil}}(i)$ 越接近 1, 样本 i 所属类 j 的类内紧密性和类间分离性好, 聚类质量好。当 $\tilde{d}_b(i) < \hat{d}_a(i)$, 此时 $I_{\text{Sil}}(i) < 0$, 样本 i 距非同类样本的距离小于类内样本的距离, 聚类失效。

所有样本的 Silhouette 指标均值 I_{Silmean} 可用于评价聚类质量, 其值越大表示聚类质量越好, 其值为负表示聚类失效。进行类数估计时, I_{Silmean} 最大值对应的类数 k 称为最佳聚类数, 记作 k^* 。

$$I_{\text{Silmean}} = \frac{1}{p} \sum_{i=1}^p I_{\text{Sil}}(i) \quad (6)$$

基于聚类有效性指标确定 PIC 算法最佳聚类数及进行算法性能评价的流程如图 1 所示, 其中, k_{\min} 和 k_{\max} 为设定的最小和最大聚类数, 通常 $k_{\min} = 2$, $k_{\max} = \sqrt{0.5p}$ 。

2.4 特性指标权重配置

日负荷特性指标权重配置的差异, 会影响整个

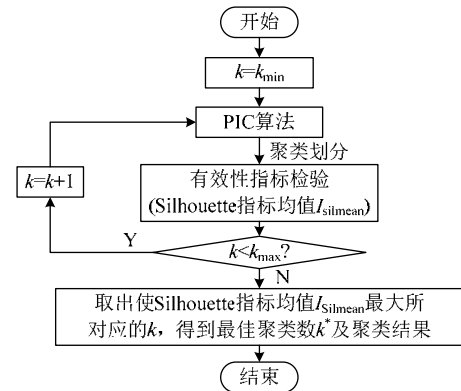


图 1 基于有效性指标确定 PIC 算法最佳聚类数的流程
Fig.1 Flow chart of obtaining the best clustering number based on PIC method

聚类结果及算法的可靠性。德尔菲法^[24]是根据专家经验直接设定权重参数的赋权方法, 通常采用算术平均法综合各专家意见, 该方法存在一定主观性, 不加筛选的赋权可使综合结果出现明显偏差。本文提出一种基于聚类有效性修正的德尔菲法权重配置方法, 称为 Validity-Delphi, 用于配置日负荷特性指标权重。该方法通过聚类有效性指标检验各专家意见的合理性, 剔除使聚类结果失效或出现明显偏差的专家意见, 保证聚类质量好的专家权重在综合结果时占优, 从而实现特性指标权重的自动调整。具体步骤如下:

1) 根据德尔菲法“背对背决策”思想邀请 n 位专家分别对 6 维日负荷特性指标权重进行赋值。第 m 位专家所赋权重为 $W_m = [w_{m1}, w_{m2}, \dots, w_{m6}]$, 其中 $m=1,2,\dots,n$, 满足 $\sum_{j=1}^6 w_{mj} = 1$ 。

2) 以第 m 位专家所赋权重作为特性指标权重, 利用 PIC 算法进行聚类, 得到最佳聚类数 k_m^* 及对应 Silhouette 指标均值 $I_{\text{Silmean}}(m)$ 。

3) 求取 n 个最佳聚类数均值 \bar{k}^* (取整), 比较各组专家权重下的最佳聚类数 k^* 和对应 Silhouette 指标均值 I_{Silmean} , 剔除使聚类失效 ($I_{\text{Silmean}} < 0$) 和使最佳聚类数 k^* 出现明显偏差 ($|\bar{k}^* - k_m^*| > 1$) 的 η 组专家权重, 得到剔除后剩余 $n-\eta$ 组专家权重的权重矩阵:

$$W = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_{n-\eta} \end{bmatrix} \quad (7)$$

4) 基于 Silhouette 指标均值分别计算剩余 $n-\eta$ 组专家权重的影响因子 β , 得到影响因子向量 $\beta = [\beta_1, \beta_2, \dots, \beta_{n-\eta}]$, 其中第 l 位专家权重的影响因子 β_l 为

$$\beta_l = \frac{I_{\text{Silmean}}(l)}{\sum_{i=1}^{n-\eta} I_{\text{Silmean}}(i)} \quad (l=1,2,\cdots,n-\eta) \quad (8)$$

5) 记修正特性指标权重矩阵 W^* 为

$$W^* = \beta W = [w_1^*, w_2^*, w_3^*, w_4^*, w_5^*, w_6^*] \quad (9)$$

将修正特性指标权重矩阵 W^* 用于 PIC 算法聚类,能够更大程度反映不同物理意义特性指标的重要程度,得到更准确的日负荷曲线聚类结果。

2.5 聚类结果及评价

以修正特性指标权重 W^* 为基准,利用 PIC 算法对特征降维矩阵 Y 重新聚类的结果,标记有效负荷曲线 X 所属分类,得各类用户的典型负荷曲线(同类负荷曲线的中心线)。

利用聚类有效性指标、算法运行时间等标准对聚类质量和算法性能进行评价,并从工程角度校验和评价聚类结果的合理性。

3 算例分析

3.1 电网实际日负荷曲线聚类

以中国某市 2015 年 4 月某工作日实测 2394 个用户的日负荷曲线为研究对象,每 15min 采集一次,共计 96 个量测点。经数据清理后,本算例共含 2355 条有效日负荷曲线。

采用基于聚类有效性修正的德尔菲法,邀请 6 位专家对特性指标赋权,各组专家权重见附录表 1。各专家权重下利用 PIC 算法的聚类结果见表 2。

表 2 各专家权重下 PIC 算法聚类结果			
Tab.2 Clustering results based on PIC method under each expert weight			
权重	电网负荷曲线		
	最佳聚类数 k^*	Silhouette 指标均值 I_{Silmean}	影响因子 β
专家权重 1	5	0.3445	—
专家权重 2	3	0.4285	0.2008
专家权重 3	3	0.4192	0.2008
专家权重 4	3	0.4289	0.2010
专家权重 5	3	0.4287	0.2010
专家权重 6	3	0.4285	0.2008

由表 2 知,各专家权重聚类结果不尽相同。专家权重 1 下最佳聚类数为 5,而其他 5 组权重下最佳聚类数为 3,其 I_{Silmean} 值小于其他 5 组权重下的 I_{Silmean} 值,聚类结果出现较大偏差,故剔除专家 1 意见。这说明各特性指标重要程度不尽相同,不应等权处理。按式(9)综合其余 5 组专家意见,得修正特性指标权重 W^* ,见附录表 2。

利用以修正特性指标权重 W^* 为基准的 PIC 算法对日负荷曲线分类,将聚类结果与以 96 个量测点的数据为输入,经极大值归一化后利用 k-means

算法^[8]聚类的结果作对比。

由图 2 聚类有效性指标检验结果可知,当分类数为 3 时,2 种算法 Silhouette 均值均最大,即 2 者最佳聚类数均为 3。从图 3、4 可知,PIC 算法聚类结果中属于一、二、三类的曲线数目依次为 1080、224、1051,k-means 算法聚类结果中属于一、二、三类的曲线数目为 1069、243、1043,其分类数目基本一致;再依次比较每条负荷曲线在 2 种算法中的分类结果,其分类一致的比率为 93.7%,分类结果高度一致。

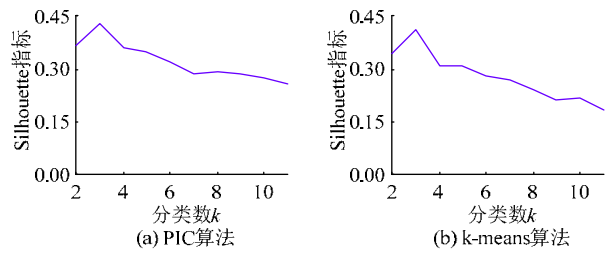


图 2 基于聚类有效性确定最佳聚类数
Fig.2 Obtaining the best clustering number based on clustering validity index

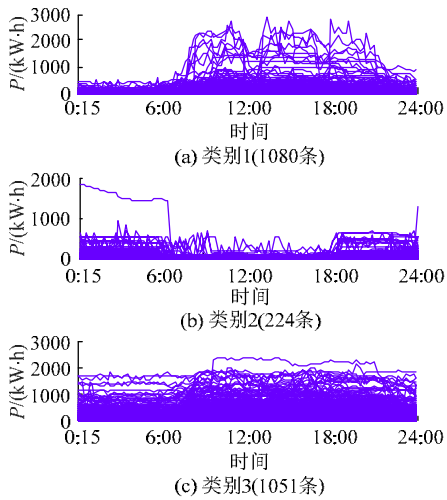


图 3 基于 PIC 算法的日负荷曲线聚类结果
Fig.3 Load curves clustering results of PIC method

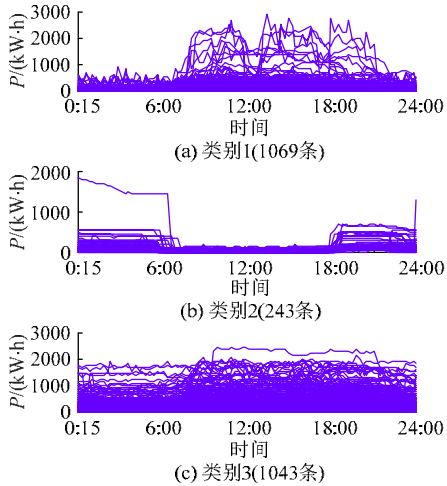


图 4 基于 k-means 算法的日负荷曲线聚类结果
Fig.4 Load curves clustering results of k-means method

由图 5, 2 种算法提取的典型负荷曲线, 负荷其形态十分相似。2 种算法提取的一、二、三类典型负荷曲线均呈双峰、避峰、峰平期 3 种类型。一类双峰型曲线, 多为大工业、政府、企业办事机构等行业用户, 工作时间固定且规律; 二类避峰型曲线较为特殊, 主要包括制冰厂、面包厂等工厂, 呈现明显的行业特点; 三类峰平期型曲线, 主要包括制衣厂等公司, 白天工作量较大, 午间负荷因员工休息略有波动。因此, 从工程角度来看, PIC 算法分类结果合理、可以应用。

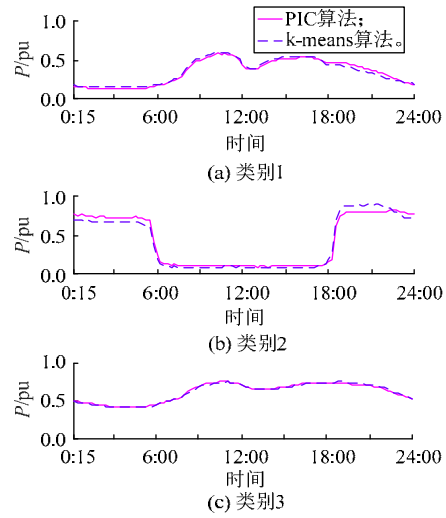


图 5 PIC 算法和 k-means 算法提取的典型负荷曲线对比
Fig. 5 Comparison of typical load curves between PIC and k-means method

由表 3, 通过聚类有效性指标和程序运行时间综合比较 2 种算法的性能, 在聚类结果一致的情况下, 使用 PIC 算法聚类的 Silhouette 指标均值大于 k-means 算法的指标均值, 其聚类质量更优, 且程序运行时间仅为 k-means 算法的 1/5 左右。综上, PIC 算法用于负荷曲线聚类不仅可行、合理, 而且提高了聚类效率和质量。

表 3 PIC 算法和 k-means 算法聚类结果性能对比 Tab.3 Clustering property comparison between PIC and k-means method					
算法	最佳聚类数 k^*	Silhouette 指标均值 $I_{Silmean}$	程序运行时间/s		
			数据降维	主程序	合计
PIC	3	0.4289	0.28	9.837	10.117
k-means	3	0.4098	—	46.193	46.193

注: 本文算例中 PIC 算法和 k-means 算法的 Silhouette 指标均值 $I_{Silmean}$ 均以极大值归一化后的多维负荷数据为测算基准。

3.2 模拟 8 类典型日负荷曲线聚类

基于文献[25]分析提取的 8 类典型日负荷曲线(后文图 10), 分别在每类典型日负荷曲线的每个采集点上添加比例为 r 的随机噪声, 模拟得分属 8 类、每类 125 条总计 1000 条的日负荷曲线。

改变噪声比例 r , 分别利用 PIC 算法和 k-means

算法对其聚类, 若聚类后该曲线所属类别与初始分类相同, 则认为分类准确, 定义分类准确率为分类准确的负荷曲线总数占全部负荷曲线的百分比, 继而利用最佳聚类数、Silhouette 指标均值、分类准确率 3 个指标检验算法的鲁棒性, 结果如表 4、图 6 所示。

表 4k-means 算法和 PIC 算法鲁棒性比较 Tab. 4 Comparison of robustness between k-means and PIC method						
噪声比例 $r/\%$	PIC 算法			k-means 算法		
	最佳分类数	Silhouette 指标均值	准确率/%	最佳分类数	Silhouette 指标均值	准确率/%
5	8	0.9919	100	8	0.9919	100
10	8	0.9700	100	7	0.8977	87.5
15	8	0.9383	100	7	0.8717	87.5
20	8	0.8981	100	6	0.8529	75.0
25	8	0.8547	99.9	7	0.8047	87.5
30	8	0.8133	99.8	7	0.7680	87.3
35	7	0.7660	87.4	7	0.7324	87.2
40	4	0.5954	50.0	5	0.6959	62.5

注: 各噪声比例下 PIC 算法以修正特性指标权重向量(附录表 2)为基准, 各专家特性指标权重向量见附录表 1。

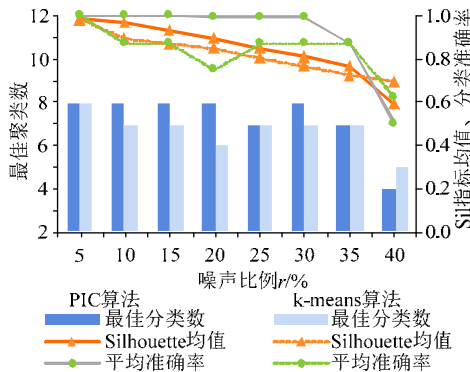


图 6 k-means 算法和 PIC 算法鲁棒性比较
Fig. 6 Comparison of robustness between k-means and PIC method

由表 4 及图 6 可知, 对于各单一算法, 当噪声比例增加时, Silhouette 指标均值、分类准确率下降, 最佳聚类数出现偏差, 表明 3 种指标可靠。对于 PIC 算法, 当噪声比例不超过 30%, 最佳聚类数一直为 8, Silhouette 均值大于 0.8, 分类准确率接近 100%, 算法鲁棒性好。而对于 k-means 算法, 当噪声比例增大至 10%, 最佳聚类数不再为 8, 分类准确率也波动变化, 受噪声干扰明显, 鲁棒性差。

当噪声比例为 30%, 1000 条模拟负荷曲线原始分类见图 7, 2 种算法聚类结果分别如图 8、9 所示。

由图 9 可知, 利用 k-means 算法聚类最佳聚类数为 7, 对比图 7, 该算法未能区别出第 7 和第 8 类曲线, 导致分类出现明显偏差。对比图 7 和 8, 利用 PIC 算法进行聚类, 仅第一类中 2 条负荷曲线分类错误, 分类正确率为 99.8%, 鲁棒性好。

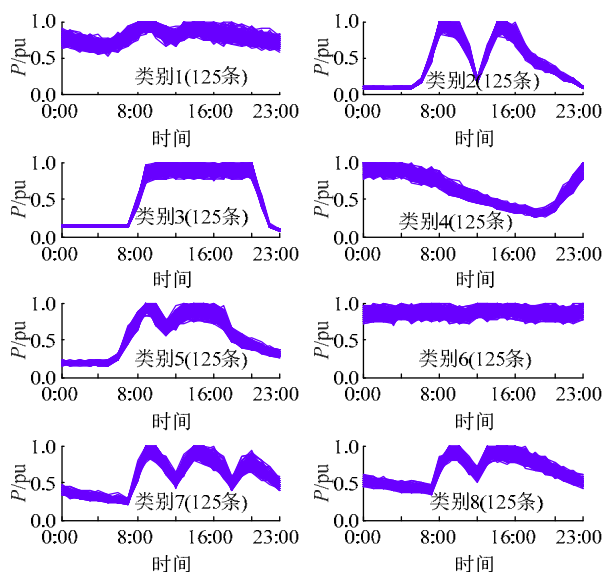
图7 模拟 1000 条典型负荷曲线($r=30\%$)

Fig.7 Simulation of 1000 typical load curves

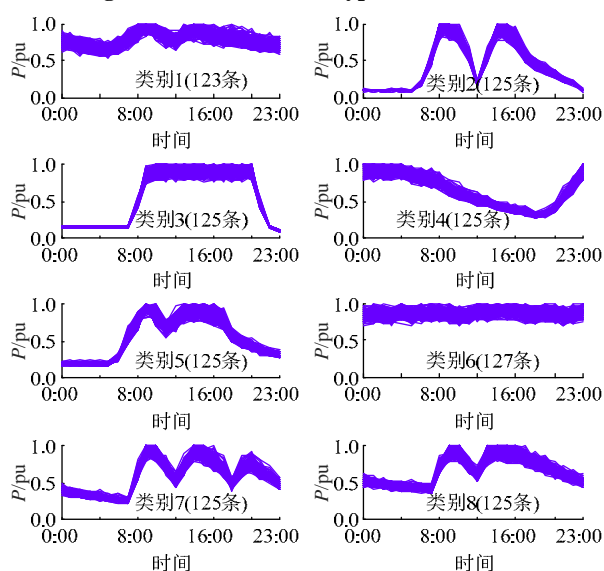
图8 PIC 算法聚类结果($r=30\%$)

Fig.8 Clustering results with PIC method

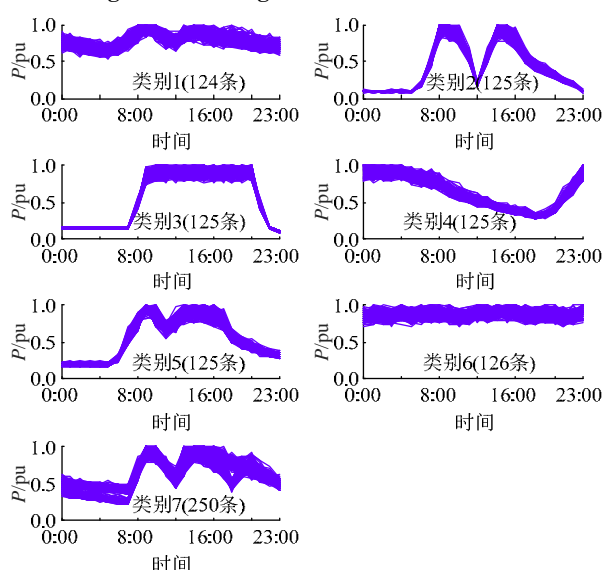
图9 k-means 算法聚类结果($r=30\%$)

Fig.9 Clustering results with k-means method

将基于 PIC 算法提取的 8 类典型负荷曲线与初始 8 类典型负荷曲线进行对比,如图 10 所示,两者负荷曲线形态高度相似。综上,所提方法对典型日负荷曲线提取效果优越,高度还原了初始典型日负荷曲线的特征和特点,算法性能稳定,耐噪声干扰,鲁棒性好。

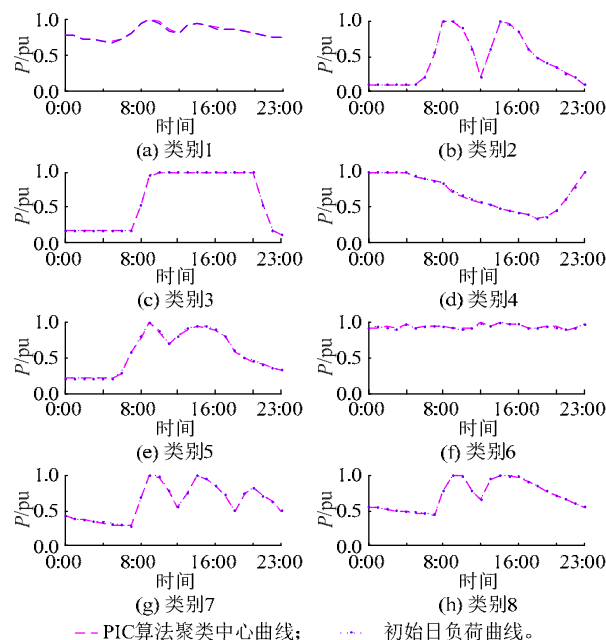
图10 PIC 算法聚类中心曲线与初始 8 类典型负荷曲线对比图($r=30\%$)

Fig.10 Comparison of typical load curves with PIC algorithm and the initial 8 kinds of load curves

4 结论

本文提出一种基于特性指标降维的日负荷曲线聚类方法,通过特征降维、权重配置及采用加权欧式距离作为相似性判据,算例表明该方法应用于负荷曲线分类可行,算法鲁棒性好,且大幅缩减了程序运行时间,提高了聚类质量,提取的典型日负荷曲线正确、合理。

日负荷特性指标选取是所提方法的基础和关键,基于本文所选 6 个特性指标进行聚类分析,取得一定效果,构建满足多层次、精细化划分要求的负荷特性指标体系是下一阶段的研究重点。

附录见本刊网络版(<http://www.dwjs.com.cn/CN/volumn/current.shtml>)。

参考文献

- [1] 崔强,王秀丽,王维洲.考虑风电消纳能力的高载能用户错峰峰谷电价研究[J].电网技术,2015,39(4):946-952.
Cui Qiang, Wang Xiuli, Wang Weizhou. Stagger peak electricity price for heavy energy-consuming enterprises considering improvement of wind power accommodation[J]. Power System Technology, 2015, 39(4): 946-952(in Chinese).
- [2] 葛少云,郭建伟,刘洪,等.计及需求侧响应及区域风光出力的电动汽车有序充电对电网负荷曲线的影响[J].电网技术,2014,

- 38(7): 1806-1811 .
GeShaoyun ,GuoJianyi ,Liu Hong ,et al .Impacts of electric vehicle's ordered charging on power grid load curve considering demand side response and output of regional wind farm and photovoltaic generation[J] .Power System Technology ,2014 ,38(7) :1806-1811(in Chinese) .
- [3] 谢胤喆,郭瑞鹏.考虑风电机组无功特性的安全约束机组组合方法[J].电力系统自动化,2012,36(14):113-118.
XieYinzhe ,GuoRuipeng . Security-constrained unit commitment considering reactive power performance of wind turbine generator[J] . Automation of Power Systems ,2012 ,36(14) :113-118(in Chinese) .
- [4] KIRSCHEN D S . Demand-side view of electricity markets[J] . IEEE Transactions on Power Systems ,2003 ,18(2) :520-527 .
- [5] Yi Wang ,Qixin Chen ,Chongqing Kang ,et al . Load profiling and its application to demand response :a review[J] .Tsinghua Science and Technology ,2015 ,20(2) :117-129 .
- [6] 张钦,王锡凡,王建学,等.电力市场下需求响应研究综述[J].电力系统自动化,2008,32(3):97-106.
Zhang Qin ,Wang Xifan ,Wang Jianxue ,et al . Survey of demand response research in deregulated electricity markets[J] .Automation of Power Systems ,2008 ,32(3) :97-106(in Chinese) .
- [7] 田世明,王蓓蓓,张晶.智能电网条件下的需求响应关键技术[J].中国电机工程学报,2014,34(22):3576-3589.
TianShiming ,Wang Beibei ,Zhang Jing . Key technologies for demand response in smart grid[J] . Proceedings of the CSEE ,2014 ,34(22) :3576-3589(in Chinese) .
- [8] 刘莉,王刚,翟登辉.k-means 聚类算法在负荷曲线分类中的应用[J].电力系统保护与控制,2011,39(23):65-68.
Liu Li ,Wang Gang ,ZhaiDenghui .Application of k-means clustering algorithm in load curve classification[J] .Power System Protection and Control ,2011 ,39(23) :65-68(in Chinese) .
- [9] 杨浩,张磊,何潜,等.基于自适应模糊C均值算法的电力负荷分类研究[J].电力系统保护与控制,2010,38(16):111-115.
Yang Hao ,Zhang Lei ,He Qian ,et al . Study of power load classification based on adaptive fuzzy C means[J] . Power System Protection and Control ,2010 ,38(16) :111-115(in Chinese) .
- [10] 陈伯成,梁冰,周越博,等.自组织映射神经网络(SOM)在客户分类中的一种应用[J].系统工程理论与实践,2004,3(8):8-14.
Chen Bocheng ,Liang Bing ,Zhou Yuebo ,et al . An application of SOM neural network in customer classification[J] . Systems Engineering-Theory & Practice ,2004 ,3(8) :8-14(in Chinese) .
- [11] Tiefeng Zhang ,Guangquan Zhang ,Jie Lu ,et al . A new index and classification approach for load pattern analysis of large electricity customers[J] . IEEE Transactions on Power Systems ,2012 ,27(2) :153-160 .
- [12] 李桂林,陈晓云.关于聚类分析中相似度的讨论[J].计算机工程与应用,2004(31):64-66.
Li Guilin ,ChenXiaoyun . The discussion on the similarity of analysis[J] . Computer Engineering and Applications ,2004(31) :64-66(in Chinese) .
- [13] Piao M ,Shon H S ,Lee J Y ,et al . Subspace projection method based clustering analysis in load profiling[J] . IEEE Transactions on Power Systems ,2014 ,29(6) ,2628-2635 .
- [14] Parsons L ,Haque E ,Liu H . Subspace clustering for high dimensional data : a review[J] . SIGKDD Explore ,2004 ,6(1) :90-105 .
- [15] 彭显刚,赖家文,陈奕.基于聚类分析的客户用电模式智能识别方法[J].电力系统保护与控制,2014,42(19):68-73.
PengXiangang ,Lai Jiawen ,Chen Yi . Application of clustering analysis in typical power consumption profile analysis[J] . Power System Protection and Control ,2014 ,42(19) :68-73(in Chinese) .
- [16] SALAHUDDIN A Azad ,ABM Shawkat Ali ,Wolfs P . Identification of typical load profiles using k-means clustering algorithm[C]//2014 Asia-Pacific World Congress on Computer Science and Engineering . Nadi : Asia-Pacific World Congress ,2014 :1-6 .
- [17] 张智晟,孙雅明,张世英,等.基于数据挖掘多层次细节分解的负荷序列聚类分析[J].电网技术,2006,30(2):51-56.
Zhang Zhisheng ,Sun Yaming ,Zhang Shiyong ,et al . Clustering analysis of electric load series using clustering algorithm of multi-hierarchy and detailed decomposition based on data mining[J] . Power System Technology ,2006 ,30(2) :51-56(in Chinese) .
- [18] 贾慧敏,何光宇,方朝雄,等.用于负荷预测的层次聚类和双向夹逼结合的多层次聚类法[J].电网技术,2007,31(23):33-36.
JiaHuimin ,He Guangyu ,Fang Chaoxiong ,et al . Load forecasting by multi-hierarchy clustering combining hierarchy clustering with approaching algorithm in two directions[J] . Power System Technology ,2007 ,31(23) :33-36(in Chinese) .
- [19] 国家发展改革委,国家电网公司.电力需求侧管理工作指南[M].北京:中国电力出版社,2007:237-245.
- [20] 赵岩,李磊,刘俊勇,等.上海电网需求侧负荷模式的组合识别模型[J].电网技术,2010,34(1):145-151.
Zhao Yan ,Li Lei ,Liu Junyong ,et al . Combinational recognition model for demand side load profile in Shanghai power grid[J] . Power System Technology ,2010 ,34(1) :145-151(in Chinese) .
- [21] 王开军.基于数据几何特性的概率推理和统计学习研究[D].西安:西安电子科技大学,2008.
- [22] Han J ,Kamber M . Data mining : concepts and techniques[M] . Beijing : China Machine Press . 2001 :223-262 .
- [23] 周世兵,徐振源,唐旭清.新的k-均值算法最佳聚类数确定方法[J].计算机工程与应用,2010,46(16):27-31.
Zhou Shibing ,XuZhenyuan ,Tang Xuqing . New method for determining optimal number of clusters in k-means clustering algorithm[J] .Computer Engineering and Applications ,2010 ,46(16) :27-31 .
- [24] 刘学毅.德尔菲法在交叉学科研究评价中的运用[J].西南交通大学学报:社会科学版,2007,42(2):21-25.
Liu Xueyi . Delphi technique in the assessment of interdisciplinary research[J] . Journal of Southwest Jiaotong University Social Sciences ,2007 ,42(2) :21-25(in Chinese) .
- [25] 黄宇腾,侯芳,周勤,等.一种面向需求侧管理的用户负荷形态组合分析方法[J].电力系统保护与控制,2013,41(13):20-25.
Huang Yuteng ,Hou Fang ,Zhou Qin ,et al . A new combinational electrical load analysis method for demand side management[J] . Power System Protection and Control ,2013 ,41(13) :20-25(in Chinese) .



刘思

收稿日期:2015-07-05。

作者简介:

刘思(1991),女,硕士研究生,主要研究方向为电力系统规划、电力系统需求侧管理,E-mail: candyls@zju.edu.cn;

李林芝(1991),男,博士研究生,主要研究方向为负荷建模算法与电力系统连锁故障,E-mail: lilinzhi@zju.edu.cn;

吴浩(1973),男,副教授,主要研究方向为电力系统运行与控制。E-mail: wuhao@zju.edu.cn;

孙维真(1963),男,高级工程师,从事电力系统运行、控制和电力市场分析方面的研究;

傅旭华(1972),男,高级工程师,研究方向为电力系统规划;

叶承晋(1987),男,工程师,主要研究方向为电力系统规划;

黄民翔(1945),男,通信作者,教授,主要研究方向为电力系统规划、电力市场。

(责任编辑 李兰欣)