

DOI: 10.19783/j.cnki.pspc.200713

基于变分模态分解和密度峰值快速搜索的 电力负荷曲线可控聚类模型

谷紫文¹, 李鹏¹, 郎恂¹, 喻怡轩¹, 沈鑫², 曹敏²

(1. 云南大学信息学院, 云南 昆明 650091; 2. 云南电网有限责任公司电力科学研究院, 云南 昆明 650217)

摘要: 电力负荷曲线作为一种非平稳信号, 可以看作由宽平稳的低频分量和非平稳的高频分量构成。针对负荷数据的时间多粒度构成特点, 提出了一种基于变分模态分解和密度峰值快速搜索的负荷可控聚类模型。原始负荷曲线通过变分模态分解算法被分解为低频, 中频和高频三个模态分量。首先, 利用负荷曲线的低频模态分量实现簇间的时间粗粒度聚类。然后, 在子类中添加中频分量实现簇内的时间细粒度聚类。使用 OpenEI 数据集对所提模型进行了有效性验证, 并与不同聚类算法对原始负荷数据直接聚类进行对比。实验结果表明该模型可以实现不同时间颗粒度的合理聚类。

关键词: 负荷曲线聚类; 变分模态分解; 密度峰值聚类; 智能电网; 数据驱动

A controllable clustering model of the electrical load curve based on variational mode decomposition and fast search of the density peak

GU Ziwen¹, LI Peng¹, LANG Xun¹, YU Yixuan¹, SHEN Xin², CAO Min²

(1. School of Information Science and Engineering, Yunnan University, Kunming 650091, China;

2. Electric Power Research Institute, Yunnan Power Grid Co., Ltd., Kunming 650217, China)

Abstract: As a kind of non-stationary signal, the electrical load curve can be regarded as composed of a wide-stationary low-frequency component and a non-stationary high-frequency component. Aiming at the temporal multi-granularity characteristics of load data, a controllable clustering model of the electrical load curve based on variational mode decomposition and fast search of the density peak is proposed. The original load curve is decomposed into three modal components of low-, intermediate-, and high-frequency by variational mode decomposition. First, the low-frequency modal components of the load curve are used to achieve coarse-grained clustering among clusters. Then, the intermediate-frequency components are superimposed on the subclasses to achieve fine-grained clustering within the cluster. The validity of the proposed model is verified using the OpenEI data set and compared with different clustering algorithms for the direct clustering of the original load data. The experimental results show that the model can achieve reasonable clustering at different time granularities.

This work is supported by the National Natural Science Foundation of China (No. 61763049) and the Key Project of Applied Basic Research of Yunnan Province (No. 2018FA032).

Key words: load curve clustering; variational mode decomposition; density peak clustering; smart grid; data driven

0 引言

随着主动配电网智能化的发展, 需要基于海量用户数据分析用户行为特性, 从而实现多能互补及

优化^[1]。负荷特性分析是电网智能化发展的重要环节, 有助于掌握用户用电行为习惯, 从而进行用电模式分析、社会属性辨识、非侵入式负荷监测等研究^[2]。如何以数据驱动的方式对电力用户进行有效分类, 是目前智能电网数据挖掘中一个广泛关注的问题^[3]。

负荷曲线作为用电特征最直观反映, 是分析

基金项目: 国家自然科学基金项目资助(61763049); 云南省应用基础研究重点课题(2018FA032)

用户用电行为的主要研究对象^[4]。现有的负荷曲线分类方法包括无监督学习和有监督学习。无监督学习主要包括 K-means^[5], 谱聚类^[6], 模糊 C 均值聚类^[7], 高斯混合模型聚类(GMM)^[8]等。有监督学习主要包括基于统计原理分类的贝叶斯模型^[9], 以及基于神经网络分类的极限学习机(ELM)^[10], 决策树^[11], 支持向量机(SVM)^[12]等。然而, 这些模型分类质量都与训练样本标签有着较高的依赖关系。基于这个问题, 近年来有学者提出将无监督学习与有监督学习相结合的方式进行负荷分类, 利用无监督算法获取样本标签, 然后将它们作为训练样本通过有监督算法完成模型的训练^[13-14]。因此, 聚类分析是负荷分类方面的一个重要环节。

在传统的聚类算法中, 基于划分的 K-means 算法易受初始聚类中心的影响, 并且确定聚类数目的理论缺少完善; 基于层次的聚类算法的计算时间复杂度过高, 并且当数据集过大时会使聚类结果陷入局部最优; 基于密度聚类的 DBSCAN 算法易受参数取值的影响^[15]。文献[16]提出的一种称为密度峰值快速搜索算法(Clustering by Fast Search and Find of Density Peaks, CFSFDP), 该算法能够自动确定聚类数目, 对算法参数取值较不敏感, 能快速完成聚类过程, 并且还能够实现任意形状的聚类。近年来, 已有不少学者针对该算法存在的问题进行了相应的改进^[17-18]。

为了提升聚类质量, 近年来, 多名学者提出了一些与聚类算法相结合的模型。文献[19]基于信息熵对负荷曲线进行分段聚合近似, 通过基于距离和形态的双尺度相似性度量方式的谱聚类算法实现了有效聚类。文献[20]以欧氏距离和皮尔逊相关系数作为度量方式, 通过内外双层聚类的 K-means 算法实现了用户的可控精细化分类。文献[21]通过基于 KNN 改进的快速密度峰值算法, 实现了海量负荷曲线的有效聚类。

从信号的角度分析, 负荷曲线由宽平稳的低频分量和非平稳的高频分量叠加组成非平稳信号。变分模态分解(Variational Mode Decomposition, VMD)是一种应用于非平稳信号的分解算法, 能够将原始信号分解成一系列频率递增的模态分量^[22]。相比较于经验模态分解(Empirical Mode Decomposition, EMD)^[23]和局部均值分解(Local Mean Decomposition, LMD)^[24], VMD 能够克服端点效应和模态混叠两大弊端^[25]。

本文考虑负荷信号的时间多粒度构成特点, 结合密度峰值聚类算法的优势, 提出了一种基于 VMD-CFSFDP 的可控聚类模型。原始负荷曲线被 VMD 分解为低频主要特征分量, 中频细节特征分

量和高频随机特征分量。CFSFDP 算法首先选取低频模态分量实现簇间主要特征聚类, 然后在子类低频分量的基础上添加中频分量实现簇内精细特征聚类。本文模型可以根据实际需求实现不同时间颗粒度的聚类分析, 通过与其他聚类模型的结果对比, 本文所提模型的结果簇内更加紧密, 簇间更加疏离。

1 基本理论

1.1 变分模态分解原理

变分模态分解算法 (Variational Mode Decomposition, VMD) 由 Dragomiretskiy 等提出, 它是一种自适应并且完全非递归的多分辨率信号分解技术。VMD 算法将原始非平稳信号 f 分解为 k 个具有不同中心频率 w_k 和有限带宽的相对平稳子信号 $\{\mu_1, \mu_2, \dots, \mu_k\}$ 。每一个子信号作为原始信号的一种带限固有模态分量(Band-Limited Intrinsic Mode Function, BLIMF), 能够反映原始信号在不同时间尺度下的结构特征。为了估算每一种模态分量, VMD 算法的求解过程概括如下。

Step1: 通过 Hibert 变换获取每种模态分量的单边频谱。

$$[\mathcal{D}(t) + \frac{j}{\pi t}] \otimes \mu_k(t) \quad (1)$$

式中: \otimes 表示卷积; μ_k 表示为第 k 个模态分量。

Step2: 将各个模态分量的频谱调制到各自预估中心频率的基频带。

$$[(\mathcal{D}(t) + \frac{j}{\pi t}) \otimes \mu_k(t)] e^{-jw_k t} \quad (2)$$

Step3: 通过计算各个解调信号的梯度平方 L^2 的范数, 估算每种模态分量的中心频率。因此, 约束变分问题表达为

$$\begin{cases} \min_{\{\mu_k\}, \{w_k\}} \left\{ \sum_k \left\| d_r [(\mathcal{D}(t) + \frac{j}{\pi t}) \otimes \mu_k(t)] e^{-jw_k t} \right\|_2^2 \right\} \\ \text{s.t. } \sum_k \mu_k = f \end{cases} \quad (3)$$

Step4: 引入惩罚因子 α 和拉格朗日惩罚算子 $\lambda(t)$, 约束性变分问题转为非约束性变分问题为

$$\begin{aligned} L(\{\mu_k\}, \{w_k\}, \lambda) = & \alpha \sum_k \left\| d_r [(\mathcal{D}(t) + \frac{j}{\pi t}) \otimes \mu_k(t)] e^{-jw_k t} \right\|_2^2 + \\ & \left\| f(t) - \sum_k \mu_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k \mu_k(t) \right\rangle \end{aligned} \quad (4)$$

Step5: 通过交替方向乘子法 (Altern + ating Direction Method of Multipilers, ADMM) 求解式(4)中的非约束变分问题。

1.2 密度峰值快速搜索算法

密度峰值快速搜索算法(Clustering by Fast Search and Find of Density Peaks, CFSFDP)由Rodriguez等提出,该算法能够自动地发现聚类中心,并且可以实现任意数据分布形状的高效聚类。CFSFDP基于两个基本假设:第一,聚类中心被低于自身密度的局部邻域数据点包围;第二,局部邻域数据点到其他聚类中心的距离相比到当前聚类中心较远。CFSFDP算法实现聚类的基本流程如下。

Step1: 计算每个样本点 x_i 的局部密度 ρ_i 。

$$\rho_i = \sum_j \chi(d_{ij} - d_c) \quad (5)$$

$$\chi(w) = \begin{cases} 1 & w \leq 0 \\ 0 & w > 0 \end{cases} \quad (6)$$

式中: d_{ij} 为样本点 x_i 与样本点 x_j 之间的欧式距离; d_c 为截断距离,其值过大将导致不同簇的合并,其值过小将导致同一簇的分裂。 d_c 应该使截断距离内的邻域样本数目为总体样本数目的 1%~2%。

为了降低 d_c 取值对聚类结果的影响,采用高斯核函数改进局部密度^[26]。

$$\rho_i = \sum_{j \neq i} \exp\left(-\frac{d_{ij}^2}{d_c}\right) \quad (7)$$

Step2: 计算样本间的拓扑关系及距离 δ_i 。

$$[\mathbf{x}_j, \delta_i] = \min_{j: \rho_j > \rho_i} (d_{ij}) \quad (8)$$

式中: \mathbf{x}_j 为密度大于 \mathbf{x}_i 且距离最近的样本,若 \mathbf{x}_i 是密度最大样本,则 $\delta_i = \max_j (d_{ij})$ 。

Step3: 选择样本密度较大且样本间距离较远的样本作为聚类中心,以 γ 值确定聚类中心。

$$\gamma_i = \rho_i \delta_i \quad (9)$$

Step4: 按照 γ 值大小的排列顺序,将剩余样本分配到它最近邻且密度比它大的样本所在簇。

1.3 聚类有效性评价指标

典型的聚类有效性评价指标为误差平方和(SSE)、Davies Bouldi Index(DBI)、轮廓系数(SC)等^[27-28]。假设数据集 \mathbf{X} 被划分为 K 个簇 $\{C_1, C_2, \dots, C_K\}$, 上述指标计算公式如下。

1) SSE 指标。该指标反映了所有簇的凝聚度的状况,其值越小说明聚类质量越高。

$$I_{SSE} = \sum_{i=1}^K \sum_{c \in C_i} \text{dist}^2(\bar{\mathbf{x}}_i, \mathbf{c}) \quad (10)$$

式中, $\bar{\mathbf{x}}_i$ 为 C_i 簇内的聚类中心。

2) DBI 指标。该指标反映了簇间的分散性和簇

内的紧凑型,其值越小说明聚类质量越高。

$$I_{DBI} = \frac{1}{K} \sum_{i=1}^K R_i \quad (11)$$

式中, R_i 表示第 i 类与其他类的最大形似程度。

3) SC 指标。该指标反映了簇之间的疏密程度,其值越接近 1 说明簇与簇之间越疏远。

$$I_{SC} = \frac{1}{N} \sum_{i=1}^N \frac{b(\mathbf{x}_i) - a(\mathbf{x}_i)}{\max\{a(\mathbf{x}_i), b(\mathbf{x}_i)\}} \quad (12)$$

式中: $a(\mathbf{x}_i)$ 表示 \mathbf{x}_i 到同一簇内其他样本的平均距离; $b(\mathbf{x}_i)$ 表示 \mathbf{x}_i 到其他簇内样本的最小平均距离。

2 负荷曲线可控聚类模型

2.1 CFSFDP 算法负荷曲线聚类问题

CFSFDP 算法能够自发的确定聚类数目和聚类中心,并实现样本所属类别的分配。然而, CFSFDP 算法应用于负荷曲线聚类时存在一些问题。图 1 为负荷曲线通过 T-SNE 数据可视化算法在二维空间的分布,形态 1 的负荷曲线分布比较稀疏,导致密度峰值点不清晰;有一部分形态 4 的负荷曲线分布在形态 3 的负荷曲线周围,这将导致 CFSFDP 算法发生一连串样本分配错误。

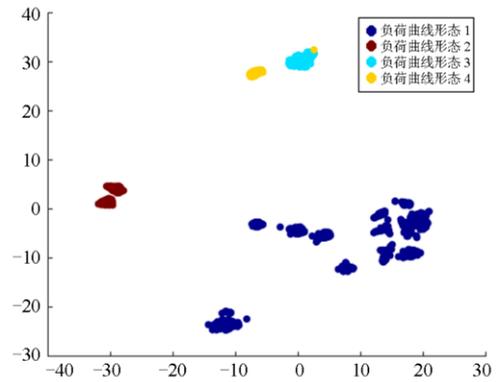


图 1 原始负荷曲线可视化分布

Fig. 1 Visual distribution of original load curves

由式(7)可知,样本密度 ρ 决定着 CFSFDP 算法的聚类质量。由于高频分量的干扰使样本分布稀疏时,导致密度峰值点不清晰,从而影响整体的聚类质量。图 2 为某用户连续 5 天工作日负荷标准化后的分解结果,低频模态分量能够反映了原始负荷的主要结构特征,随着模态分量的中心频率增大,曲呈现线的随机性变化越强。

设原始信号 \mathbf{X} 可以由 BLIMF1 分量 $\mathbf{x}^{\text{blimf1}}$ 和其他分量 \mathbf{x}^N 构成:

$$\mathbf{X} = \mathbf{x}^{\text{blimf1}} + \mathbf{x}^N \quad (13)$$

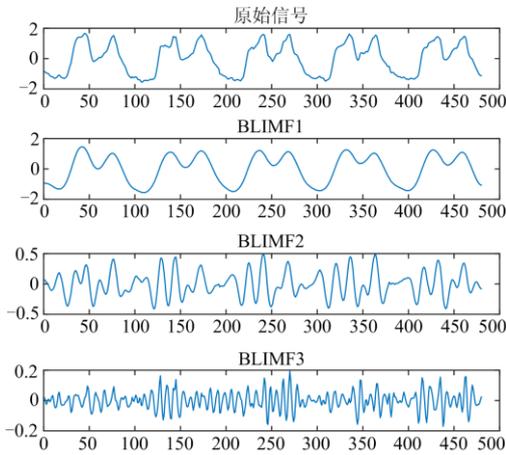


图2 负荷曲线 VMD 分解结果

Fig. 2 VMD decomposition result of load curves

通过原始信号与 BLIMF1 分量计算样本之间的距离，它们之间的关系为

$$dist(x_i, x_j) = dist(x_i^{blimf1}, x_j^{blimf1}) + N \quad (14)$$

式中， N 表示大于等于 0 的变量。

由于通过 BLIMF1 分量计算样本之间的距离更小，因此，BLIMF1 模态分量的分布比原始信号更密集。图 3 为负荷曲线 BLIMF1 分量通过 T-SNE 数据可视化算法在二维空间的分布，相同形态的负荷

曲线分布较为紧密，不同形态的负荷曲线没有发生混叠的现象。

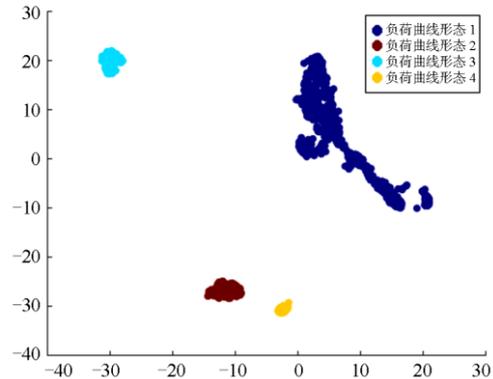


图3 负荷曲线低频分量可视化分布

Fig. 3 Visual distribution of low-frequency components of load curves

2.2 负荷曲线可控聚类模型

本文所提基于 VMD-CFSFDP 的可控聚类的总体框架如图 4 所示。首先 VMD 算法将负荷曲线分解为多个不同频率占优并且相对平稳的子序列，然后选择不同时间尺度的频率模态分量，最后通过 CFSFDP 算法实现不同时间尺度的可控聚类。

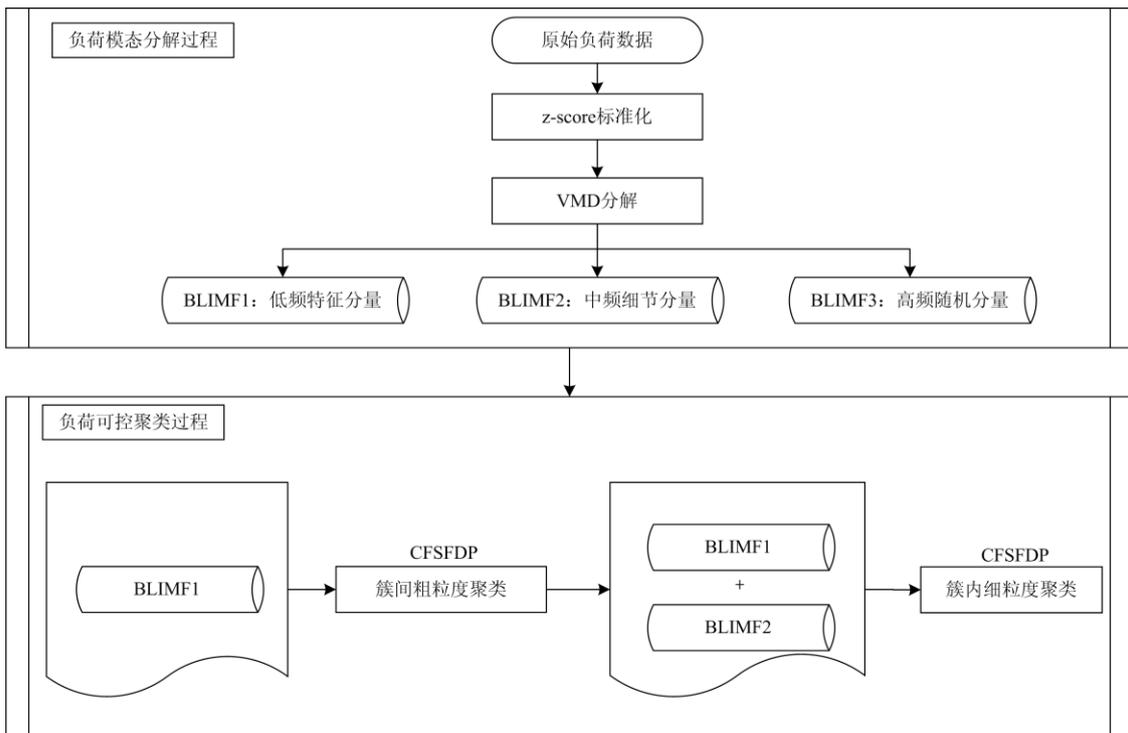


图4 基于 VMD-CFSFDP 可控聚类的总体框架

Fig. 4 Framework of controllable clustering model based on VMD-CFSFDP

1) 负荷模态分解过程

负荷曲线分解的目的是获取数据不同频率的模态分量，以低频子序列作为负荷曲线的主要特征分量，中频子序列作为负荷曲线的细节特征分量，高频子序列作为负荷曲线的随机特征分量。基于 VMD 的负荷分解流程如下所述。

Step1: 数据 z-score 标准化。标准化的目的是将不同量级的负荷统一化为同一个量级，保证数据之间的可比性。标准化后的数据 x^* 表示为

$$x^* = \frac{x - \mu}{\sigma} \quad (15)$$

式中： μ 为序列 x 的均值； σ 为序列 x 的标准差。

Step2: VMD 参数设定。负荷曲线的分解结果主要受 VMD 中两个参数的影响，惩罚因子 α 和分解层数 K 。惩罚因子越大，以原始数据的中心频率为中心，两侧的频率衰减越快，本文将 α 设为日负荷抽样点长度的 2 倍。分解层数 K 过大将导致原始信号过度分解，过小将导致模态混叠。本文中分解层数 K 设置为 3，主要原因是 BLIMF1 作为低频模态分量表现原始负荷的主要特征，BLIMF2 作为较高频模态分量反映原始负荷的细节特征，BLIMF3 作为高频模态分量体现原始负荷的随机特征。

Step3: VMD 信号分解。将某一用户 n 天采样频率为 d 的日负荷数据 $X^{n \times d}$ 转换为二维的时间序列 $X^{1 \times nd}$ ，然后通过 VMD 算法进行分解，最终得出 3 个模态分量。

2) 负荷可控聚类过程

负荷可控聚类的目的是根据不同的需求响应，选取合适的模态分量，通过 CFSFDP 算法实现不同时间颗粒度的聚类分析。基于 CFSFDP 的负荷可控聚类流程如图 5 所示。

Step1: CFSFDP 的参数设定。聚类结果主要受邻域样本数目 m 和截断距离 d_c 的影响。邻域样本数目 m 一般设为总体样本数目的 1%~2%。 d_c 通过 KNN 算法搜寻样本与邻域样本的距离确定。

$$d_c = \frac{1}{N} \sum_{j \in \text{KNN}(i)} \max(d_{ij}) \quad (16)$$

Step2: 簇间时间粗粒度聚类。选取低频 BLIMF1 模态分量，通过式(8)做出 γ 值分布并判断样本分布情况，选取 q 个样本作为聚类中心，输出初始聚类结果 $\{c_1, c_2, \dots, c_q\}$ 。

Step3: 簇内时间细粒度聚类。通过上一步粗时间颗粒度的聚类，虽然解决了样本分布稀疏时导致密度峰值点不突出的问题，但也导致了一些特征区

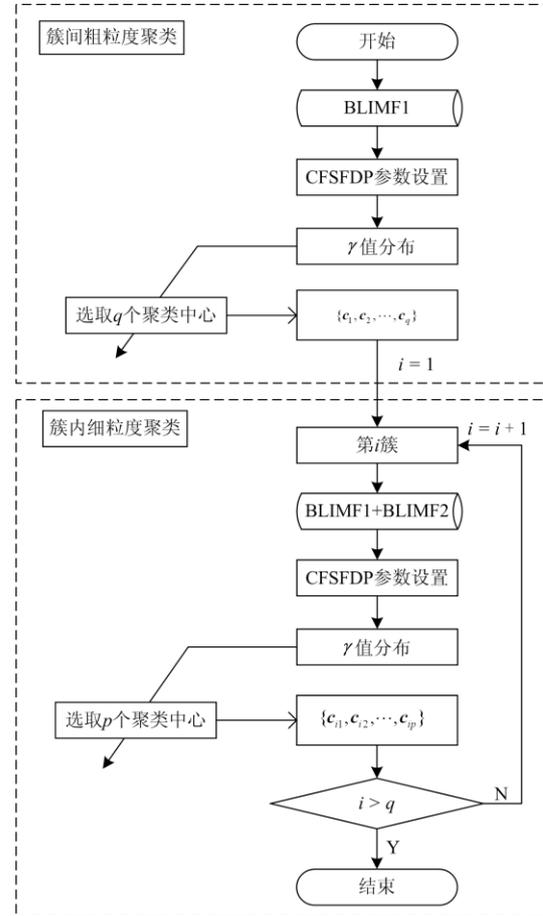


图 5 可控聚类流程

Fig. 5 Controllable clustering process

别度不大的簇发生了合并。为了实现精细聚类，基于上一步的聚类结果，在 c_i 簇内通过样本的 BLIMF1 分量和 BLIMF2 分量计算样本之间的密度及距离。根据式(8)做出 γ 值分布并判断样本分布情况，选择 p 个样本作为聚类中心，并输出该簇新划分的子类 $\{c_{i1}, c_{i2}, \dots, c_{ip}\}$ 。

3 案例分析

本文选取某网站公布的商业用户负荷数据作为案例，其涵盖 16 个行业的夏季(6~9 月份 85 个工作日)总共为 13 600 条日负荷曲线，采样时间间隔为 1 h。

3.1 负荷曲线可控聚类分析实例

3.1.1 负荷曲线簇间粗粒度聚类

选取原始负荷曲线的 BLIMF1 模态分量，通过图 6 的决策图，选取分布突出的 4 个样本作为聚类中心，聚类中心形态如图 7 所示，最终的聚类结果如图 8 所示，各类簇的用户信息如表 1 所示。负荷形态主要分为单峰、双峰和三峰 3 种形态。

第 1 类负荷形态为单峰型, 包括医院、办公楼、学校、超市、仓库等, 主要表现为白天用电量较大。第 2 类负荷形态为双峰 A 型, 仅包括酒店类, 主要表现为在 07:00 和 20:00 附近存在两个用电高峰, 并且这两个峰值幅度接近。第 3 类负荷形态为双峰 B 型, 仅包含中层公寓, 主要表现为在 07:00 和 19:00 附近存在两个用电高峰, 但是晚上的峰值幅度高于早上的峰值。在图 7 中, 两种双峰型负荷表现为不同的形态。第 4 类负荷形态为三峰型, 仅包含餐厅类, 主要表现为在 07:00、11:30 和 17:30 附近存在三个用电高峰, 并且这三个峰值幅度接近。

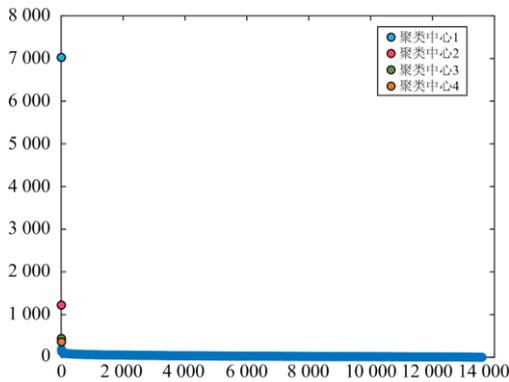


图 6 粗粒度聚类的决策图

Fig. 6 Decision graph for coarse-grained clustering

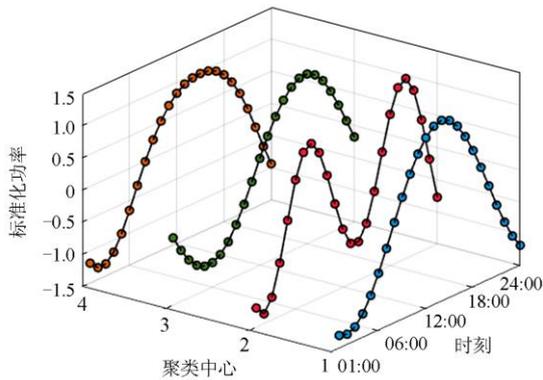


图 7 粗粒度聚类中心

Fig. 7 Coarse-grained cluster centers

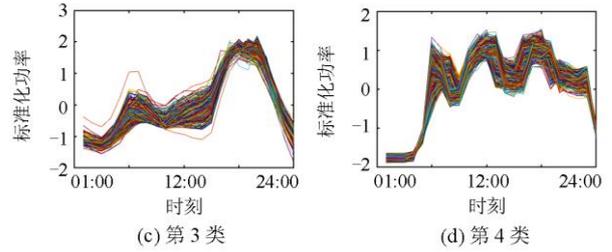
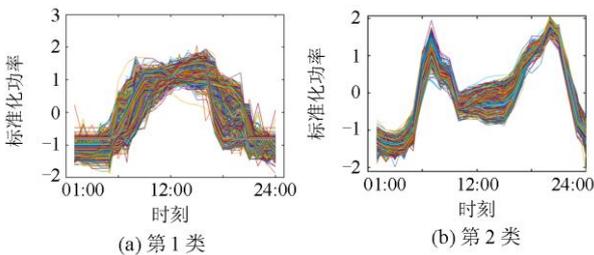


图 8 负荷曲线簇间粗粒度聚类结果

Fig. 8 Coarse-grained clustering results among load curve clusters

表 1 簇间粗粒度聚类的用户信息

Table 1 User information of coarse-grained clustering among clusters

行业类型	第 1 类	第 2 类	第 3 类	第 4 类
全服务餐厅	0	0	0	850
医院	850	0	0	0
大型酒店	0	850	0	0
大型办公室	850	0	0	0
中型办公室	850	0	0	0
中层公寓	0	0	850	0
门诊部	850	0	0	0
小学	850	0	0	0
快速服务餐厅	0	0	0	850
中学	850	0	0	0
小型酒店	0	850	0	0
小型办公楼	850	0	0	0
独立零售店	850	0	0	0
沿公路商业区	850	0	0	0
超级市场	850	0	0	0
仓库	850	0	0	0

3.1.2 负荷曲线簇内细粒度聚类

分别观察粗粒度聚类形成各个簇的决策图, 判断可作为聚类中心的样本点的数量是否大于 1, 最终发现只有第 1 类簇可进行簇内精细聚类。选取第 1 类负荷曲线的 BLIMF1 和 BLIMF2 模态分量进行簇内时间细颗粒度聚类, 通过图 9 的决策图, 选取分布突出的 3 个样本作为聚类中心, 聚类中心形态如图 10 所示, 最终的聚类结果如图 11 所示, 各类簇的用户信息如表 2 所示。

第 1-1 类主要包括办公楼、学校和超市, 峰值期间主要在 09:00—17:00, 峰值共持续约 8 h。第 1-2 类主要包括医院和门诊, 峰值期间主要在 07:00—

17:00, 峰值共持续约 10 h。第 1-3 类仅包括仓库, 峰值期间主要在 09:00—16:00, 峰值共持续约 7 h。

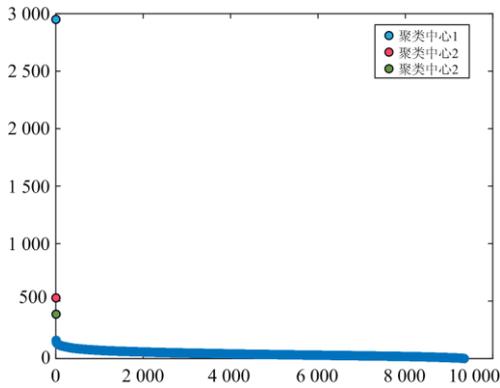


图 9 第 1 类的决策图

Fig. 9 Decision diagram of cluster 1

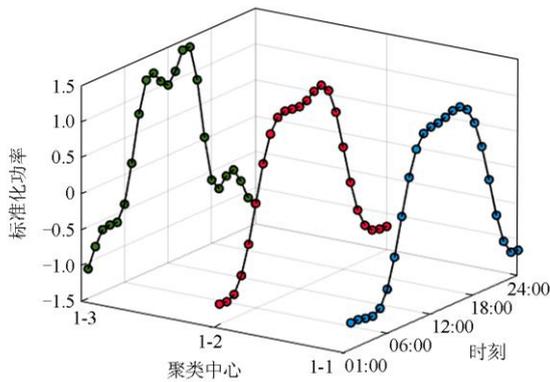


图 10 第 1 类细粒度聚类中心

Fig. 10 Fine-grained cluster centers of cluster 1

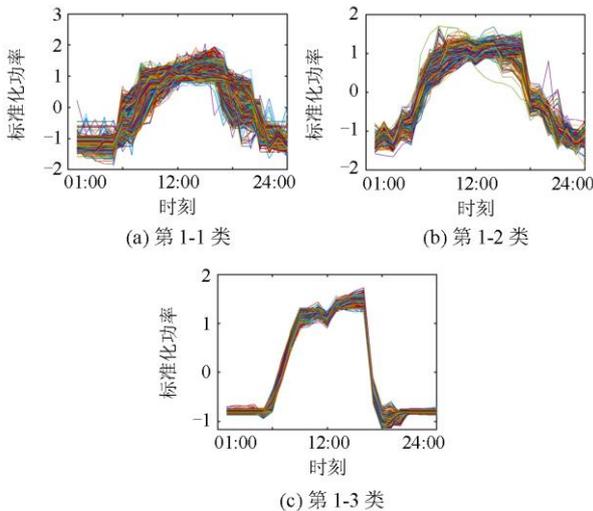


图 11 第 1 类的细粒度聚类结果

Fig. 11 Fine-grained clustering results of cluster 1

表 2 第 1 类细粒度聚类的用户信息

Table 2 User information of cluster 1 fine-grained clustering

行业类型	第 1-1 类	第 1-2 类	第 1-3 类
医院	0	850	0
大型办公室	850	0	0
中型办公室	824	26	0
门诊部	20	830	0
小学	850	0	0
中学	846	4	0
小型办公楼	850	0	0
独立零售店	850	0	0
沿公路商业区	850	0	0
超级市场	850	0	0
仓库	0	0	850

3.2 负荷曲线聚类对比

高斯混合模型(GMM)适用于多种灵活的类簇形状聚类, 是一种高效的聚类方法^[8]。为了验证本文方法的有效性, 采用 CFSFDP 算法和 GMM 算法对原始样本(z-score 标准化)进行直接聚类。CFSFDP 通过图 12 的决策图, 挑选 6 个样本作为聚类中心, 聚类中心形态如图 13 所示, 最终的聚类结果如图 14 所示。GMM 算法由于需要人工设置聚类个数, 为了方便对比设置与 CFSFDP 相同的聚类个数, 最终的聚类结果如图 15 所示。CFSFDP 和 GMM 的聚类结果的各类簇的用户信息如表 3 所示。

通过观察图 6、图 9 和图 12 的 γ 分布可知, VMD-CFSFDP 和 CFSFDP 均发现了 6 个聚类中心, 说明通过不同时间尺度的层次可控聚类不会对数据潜在的聚类数目产生影响。通过观察图 7、图 10 和图 13 的聚类中心的形态, 由于两种不同形态的双峰型负荷曲线在低频分量的形态和原始形态差异度较大, 因此 VMD-CFSFDP 算法能够有效地区分这两种负荷, 而 CFSFDP 由于直接对样本进行聚类, 导致密度峰值点不清晰, 而无法提取出这两种形态不同的双峰型聚类中心。

CFSFDP 算法的聚类结果中, 酒店和公寓被划分为第 2 类, 两种不同形态特征的双峰型用户存在混叠。办公楼被划分为为了第 6 类, 该类的负荷形态为单峰, 峰值期间大概为 08:00—16:00, 峰值共持续约 8 h。与 VMD-CFSFDP 第 1-1 类户仅存在峰期起始和结束时段的不同, 它们的峰期持续时间都大概为 8 h。GMM 算法的聚类结果中, 大型酒店和小型酒店都属于双峰 A 型负荷形态, 却被划分到了不同的簇中, 中层公寓和小型酒店两种不同负荷形态的双峰型被划分混叠在第 3 类。在单峰型的负荷分类中, 不同峰期时段和峰期持续时间的用户也发生

表 3 CFSFDP 和 GMM 聚类的各类簇的用户信息

Table 3 User information of various clusters clustered by CFSFDP and GMM

用户类别	CFSFDP 算法聚类簇用户信息						GMM 算法聚类簇用户信息					
	第 1 类	第 2 类	第 3 类	第 4 类	第 5 类	第 6 类	第 1 类	第 2 类	第 3 类	第 4 类	第 5 类	第 6 类
全服务餐厅	0	0	0	0	850	0	0	850	0	0	0	0
医院	0	0	850	0	0	0	850	0	0	0	0	0
大型酒店	0	850	0	0	0	0	0	0	2	0	848	0
大型办公室	186	0	0	0	0	664	118	0	0	732	0	0
中型办公室	0	0	0	0	0	850	0	0	0	850	0	0
中层公寓	0	850	0	0	0	0	0	0	850	0	0	0
门诊部	2	0	840	0	0	8	840	0	0	10	0	0
小学	850	0	0	0	0	0	0	0	0	0	0	850
快速服务餐厅	0	0	0	0	850	0	0	850	0	0	0	0
中学	848	0	0	0	0	2	364	0	0	0	0	486
小型酒店	0	850	0	0	0	0	0	0	850	0	0	0
小型办公楼	0	0	0	0	0	850	0	0	0	850	0	0
独立零售店	850	0	0	0	0	0	0	0	0	0	0	850
沿公路商业区	850	0	0	0	0	0	0	0	0	0	0	850
超级市场	850	0	0	0	0	0	0	0	0	0	0	850
仓库	0	0	0	850	0	0	0	0	0	850	0	0

了分类混叠, 办公楼被分到仓库所在簇, 部分大型办公室和中学用户被分为了医院和门诊所在簇。

为了客观评价上述 3 种算法对负荷曲线的聚类质量, 通过 1.3 节中的 3 个聚类质量评价指标进行对比, 其结果如表 4 所示。从聚类指标的反映来看, VMD-CFSFDP 的 SSE 指标的值最小, 说明聚类结果各个簇的凝聚度最好; 它的 DBI 指标的值最小, 说明聚类结果簇内最为紧密, 簇间最为分散; 它的 SC 指标的值最大, 说明聚类结果的簇间疏密程度最

表 4 聚类质量评价指标的对比

Table 4 Comparison of cluster quality evaluation indexes

聚类评价指标	VMD-CFSFDP	CFSFDP	GMM
SEE	6 605	6 690	7 319
DBI	0.715 7	0.922 8	1.215 8
SC	0.522 2	0.433 4	0.342 8

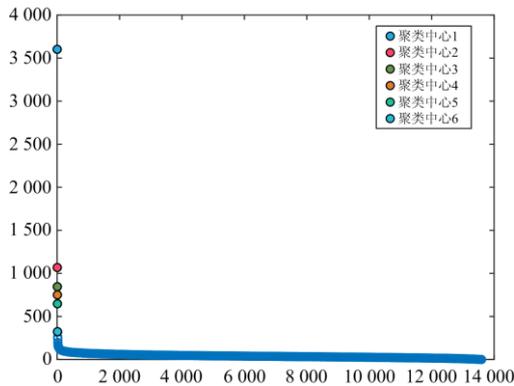


图 12 原始数据样本的决策图

Fig. 12 Decision diagram of original samples

大。因此, VMD-CFSFDP 在这 3 种聚类模型中的聚类质量表现最好。

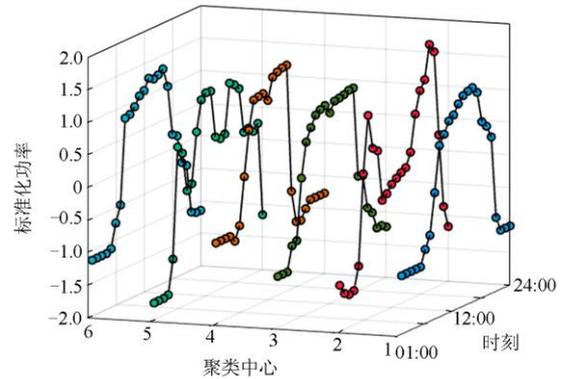
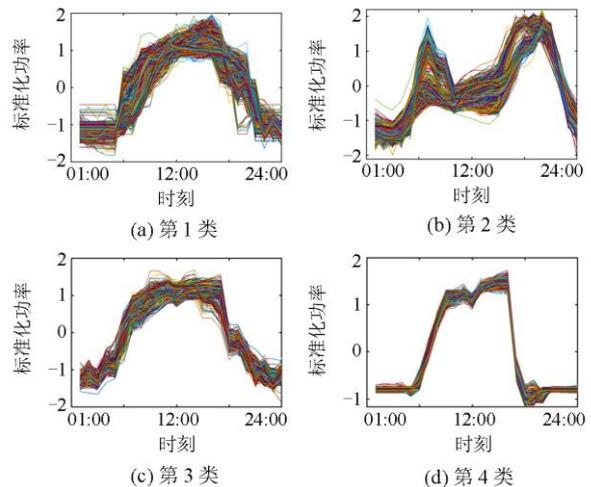


图 13 原始样本的聚类中心

Fig. 13 Original samples cluster centers



(c) 第 3 类

(d) 第 4 类

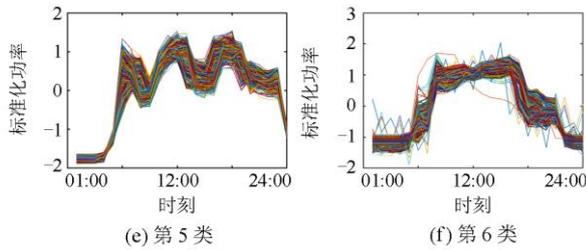


图 14 CFSFDP 聚类结果

Fig. 14 CFSFDP clustering results

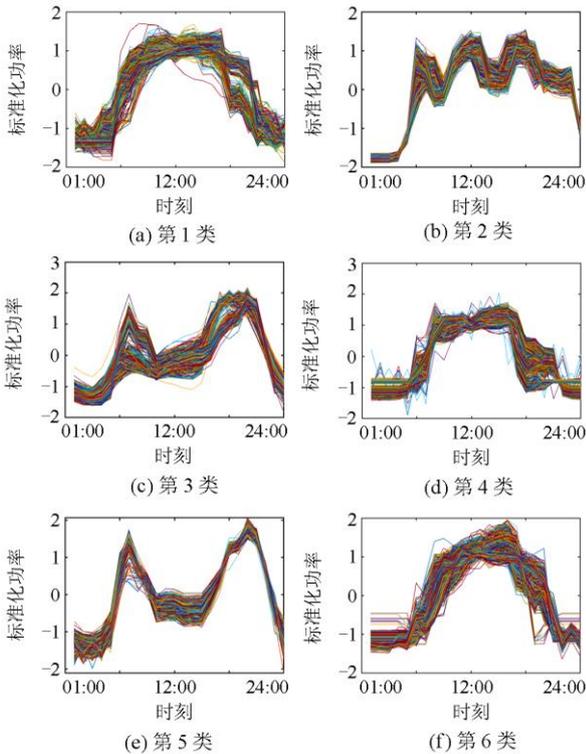


图 15 GMM 聚类结果

Fig. 15 GMM clustering results

4 结语

本文所提模型通过变分模态分解算法获取原始信号的不同频率分量，通过密度峰值快速搜索算法选取不同时间尺度的模态分量实现负荷的可控聚类。负荷数据的低频分量反映了负荷的主要特征，以低频分量进行聚类可忽略原始负荷之间的细节特征信息，并实现聚类数目较少的粗粒度聚类。在低频分量聚类结果的基础上添加中频细节分量，实现簇内细粒度聚类。通过案例分析，表明本文模型可以实现不同时间颗粒度的可控聚类分析，并且能够得到较高的聚类质量结果。

运用该模型进行负荷可控分类，有助于供电部门根据需求响应掌握用户不同时间尺度的负荷特

性，从而制定合理的政策，最终能够为消费者提供更加优质的服务。

参考文献

[1] 肖振锋, 辛培哲, 刘志刚, 等. 泛在电力物联网形势下的主动配电网规划技术综述[J]. 电力系统保护与控制, 2020, 48(3): 43-48.
XIAO Zhenfeng, XIN Peizhe, LIU Zhigang, et al. An overview of planning technology for active distribution network under the situation of ubiquitous power internet of things[J]. Power System Protection and Control, 2020, 48(3): 43-48.

[2] HUANG Wujing, ZHANG Ning, KANG Chongqing, et al. From demand response to integrated demand response: review and prospect of research and application[J]. Protection and Control of Modern Power Systems, 2019, 4(12): 148-160. DOI: <https://doi.org/10.1186/s41601-019-0126-4>.

[3] 何奉禄, 陈佳琦, 李钦豪, 等. 智能电网中的物联网技术应用与发展[J]. 电力系统保护与控制, 2020, 48(3): 58-69.
HE Fenglu, CHEN Jiaqi, LI Qin hao, et al. Application and development of internet of things in smart grid[J]. Power System Protection and Control, 2020, 48(3): 58-69.

[4] 彭小圣, 邓迪元, 程时杰, 等. 面向智能电网应用的电力大数据关键技术[J]. 中国电机工程学报, 2015, 35(3): 503-511.
PENG Xiaosheng, DENG Diyuan, CHENG Shijie, et al. Key technologies of electric power big data and its application prospects in smart grid[J]. Proceedings of the CSEE, 2015, 35(3): 503-511.

[5] 韩帅, 孙乐平, 杨艺云, 等. 基于改进 K-Means 聚类 and 误差反馈的数据清洗方法[J]. 电网与清洁能源, 2020, 36(7): 9-15.
HAN Shuai, SUN Leping, YANG Yiyun, et al. A data cleaning method based on improved K-Means clustering and error feedback[J]. Power System and Clean Energy, 2020, 36(7): 9-15.

[6] 林顺富, 田二伟, 符杨, 等. 基于信息熵分段聚合近似和谱聚类的负荷分类方法[J]. 中国电机工程学报, 2017, 37(8): 2242-2252.
LIN Shunfu, TIAN Erwei, FU Yang, et al. Power load classification method based on information entropy piecewise aggregate approximation and spectral clustering[J]. Proceedings of the CSEE, 2017, 37(8): 2242-2252.

[7] 吴亚雄, 高崇, 曹华珍, 等. 基于灰狼优化聚类算法的日负荷曲线聚类分析[J]. 电力系统保护与控制, 2020, 48(6): 68-76.
WU Yaxiong, GAO Chong, CAO Huazhen, et al. Clustering analysis of daily load curves based on GWO algorithm[J]. Power System Protection and Control, 2020, 48(6): 68-76.

[8] LI Kehua, MA Zhenjun, ROBINSON D, et al. Identification of typical building daily electricity usage profiles using Gaussian mixture model-based clustering and hierarchical clustering[J]. Applied Energy, 2018, 231(1): 331-342.

- [9] AHMADI H, MARTI J R. Load decomposition at smart meters level using eigenloads approach[J]. IEEE Transactions on Power Systems, 2015, 30(6): 3425-3436.
- [10] XIA Yang, GOU Bin, XU Yan. A new ensemble-based classifier for IGBT open-circuit fault diagnosis in three-phase PWM converter[J]. Protection and Control of Modern Power Systems, 2018, 3(33): 364-372. DOI: <https://doi.org/10.1186/s41601-018-0109-x>.
- [11] ZHAO Wenjing, SHANG Liqun, SUN Jinfan. Power quality disturbance classification based on time-frequency domain multi-feature and decision tree[J]. Protection and Control of Modern Power Systems, 2019, 4(27): 337-342. DOI: <https://doi.org/10.1186/s41601-019-0139-z>.
- [12] PATEL U J, CHOTHANI N G, BHATT P J. Sequence-space-aided SVM classifier for disturbance detection in series compensated transmission line[J]. IET Science, Measurement & Technology, 2018, 12(8): 983-993.
- [13] 谢敏, 邓佳梁, 吉祥, 等. 基于信息熵和变精度粗糙集优化的支持向量机降温负荷预测方法[J]. 电网技术, 2017, 41(1): 210-214.
XIE Min, DENG Jialiang, JI Xiang, et al. Cooling load forecasting method based on support vector machine optimized with entropy and variable accuracy roughness set[J]. Power System Technology, 2017, 41(1): 210-214.
- [14] 王德文, 周昉昉. 基于无监督极限学习机的用电负荷模式提取[J]. 电网技术, 2018, 42(10): 3393-3400.
WANG Dewen, ZHOU Fangfang. Extraction of electricity consumption load pattern based on unsupervised extreme learning machine[J]. Power System Technology, 2018, 42(10): 3393-3400.
- [15] 张斌, 庄池杰, 胡军, 等. 结合降维技术的电力负荷曲线集成聚类算法[J]. 中国电机工程学报, 2015, 35(15): 3741-3749.
ZHANG Bin, ZHUANG Chijie, HU Jun, et al. Ensemble clustering algorithm combined with dimension reduction techniques for power load profiles[J]. Proceedings of the CSEE, 2015, 35(15): 3741-3749.
- [16] RODRIGUEZ A, LAIO A. Clustering by fast search and find of density peaks[J]. Science, 2014, 344: 1492-1496.
- [17] HOU Jian, ZHANG Aihua. Enhancing density peak clustering via density normalization[J]. IEEE Transactions on Industrial Informatics, 2020, 16(4): 2477-2485.
- [18] 谢娟英, 高红超, 谢维信. K近邻优化的密度峰值快速搜索聚类算法[J]. 中国科学: 信息科学, 2016, 46(2): 258-280.
XIE Juanying, GAO Hongchao, XIE Weixin. K-nearest neighbors optimized clustering algorithm by fast search and finding the density peaks of a dataset[J]. Science China: Information Sciences, 2016, 46(2): 258-280.
- [19] LIN Shunfu, LI Fangxing, TIAN Erwei, et al. Clustering load profiles for demand response applications[J]. IEEE Transactions on Smart Grid, 2017, 10(2): 1599-1607.
- [20] 卜凡鹏, 陈俊艺, 张琪祁, 等. 一种基于双层迭代聚类分析的负荷模式可控精细化识别方法[J]. 电网技术, 2018, 42(3): 903-913.
BU Fanpeng, CHEN Junyi, ZHANG Qiqi, et al. A controllable refined recognition method of electrical load pattern based on bilayer iterative clustering analysis[J]. Power System Technology, 2018, 42(3): 903-913.
- [21] 陈俊艺, 丁坚勇, 田世明, 等. 基于改进快速密度峰值算法的电力负荷曲线聚类分析[J]. 电力系统保护与控制, 2018, 46(20): 85-93.
CHEN Junyi, DING Jianyong, TIAN Shiming, et al. An improved density peaks clustering algorithm for power load profiles clustering analysis[J]. Power System Protection and Control, 2018, 46(20): 85-93.
- [22] DRAGOMIRETSKIY K, ZOSSO D. Variational mode decomposition[J]. IEEE Transactions on Signal Processing, 2014, 62(3): 531-544.
- [23] HUANG N E, SHEN Z, LONG S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. Proceedings A, 1998, 454: 983-995.
- [24] SMITH J S. The local mean decomposition and its application to EEG perception data[J]. Journal of the Royal Society Interface, 2005, 2(5): 443-454.
- [25] 李卫国, 焦盘龙, 刘新宇, 等. 基于变分模态分解的储能辅助传统机组调频的容量优化配置[J]. 电力系统保护与控制, 2020, 48(6): 43-52.
LI Weiguo, JIAO Panlong, LIU Xinyu, et al. Capacity optimization configuration of energy storage auxiliary traditional unit frequency modulation based on variational mode decomposition[J]. Power System Protection and Control, 2020, 48(6): 43-52.
- [26] WANG Shuliang, WANG Dakui, LI Caoyuan, et al. Comment on "clustering by fast search and find of density peaks"[J]. arXiv preprint arXiv: 1501.04267, 2015.
- [27] DAVIES D L, BOULDIN D W. A cluster separation measure[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1979, 1(2): 224-227.
- [28] KWAC J, FLORA J, RAJAGOPAL R. Household energy consumption segmentation using hourly data[J]. IEEE Transactions on Smart Grid, 2014, 5(1): 420-430.

收稿日期: 2020-06-23; 修回日期: 2020-08-17

作者简介:

谷紫文(1995—), 男, 硕士研究生, 主要研究方向为电力大数据分析建模、智能信息处理; E-mail: gzw@mail.ynu.edu.cn

李鹏(1976—), 男, 通信作者, 博士, 副教授, 主要研究方向为电力物联网与能源互联网、电力大数据分析建模; E-mail: lipeng@ynu.edu.cn

郎恂(1994—), 男, 博士后, 主要研究方向为工业大数据分析建模。E-mail: langxun@ynu.edu.cn

(编辑 姜新丽)