

基于密度空间聚类 and 引力搜索算法的居民负荷用电模式分类模型

苏 适¹, 李康平², 严玉廷¹, 陆 海¹, 汪新康², 刘力铭², 王 飞^{2,3}, 董 凌⁴

(1. 云南电网有限责任公司电力科学研究院, 云南 昆明 650217;

2. 华北电力大学 新能源电力系统国家重点实验室, 河北 保定 071003; 3. 美国伊利诺伊大学厄巴纳-香槟分校

电气与计算机工程系, 美国 厄巴纳 61802; 4. 国网青海省电力公司, 青海 西宁 810008)

摘要: 居民用户用电模式分类研究可为需求侧响应方案设计、负荷特性分析及其高精度预测提供支撑。首先, 利用基于密度的空间聚类算法提取得到用户的典型用电模式; 然后, 考虑每天不同时段及季节变换对用户用电行为的影响, 提取能够描述用户在不同时间尺度下用电行为的6个特征; 在此基础上, 提出了一种基于引力搜索算法的用户用电模式分类模型; 最后, 对实测居民用电数据进行聚类, 并对各类用户的用电模式及其参与需求侧响应的潜力进行了分析。

关键词: 用电模式; 聚类算法; 特征提取; 分类; 引力搜索算法; 密度空间聚类

中图分类号: TM 761

文献标识码: A

DOI: 10.16081/j.issn.1006-6047.2018.01.019

0 引言

随着智能电网的不断发展, 智能电表的普及率越来越高。作为智能电网高级量测体系的重要组成部分, 智能电表是连接用户和电网的关键设备, 是了解用户用电情况的重要接口^[1]。如何利用智能电表采集的大量用户用电数据并对其进行数据挖掘以获得有助于提高运行可靠性与经济社会效益的有效信息是传统电力公司向综合能源服务商转型过程中迫切需要解决的重要问题之一^[2]。

负荷曲线聚类是指依据用户的用电特性对用户进行划分, 将具有相似用电模式的用户归类, 这有助于电力公司了解不同用户的用电行为, 从而为制定需求侧响应政策以及高精度负荷预测提供支撑。此外, 负荷模式聚类还可用于异常用电检测^[3]、负荷控制^[4]、配电网规划等。因此, 研究电力用户负荷模式分类技术对于提升电力公司服务水平、改善电网资产利用效率、提高企业经济效益、节能增效具有重要意义^[5]。

近年来, 受电网运营商提高供电服务质量的驱动, 针对居民用户用电行为分类的研究愈来愈多^[6-9]。

文献[6]提出了一种基于云计算平台和并行 K -means 聚类算法的居民用电行为分析模型, 仿真结果表明基于云计算的 K -means 聚类算法有更高的聚类效率。文献[7]对常见的几种聚类算法(改进的 leader 聚类算法、层次聚类算法、 K -means 聚类算法、模糊聚类算法)在用电模式分类中的应用进行了对比分析, 结果表明改进的 leader 聚类算法和层次聚类算法可以更好地处理孤立样本, 在异常用电模式存在的情况下仍然具有较好的聚类效果。文献[8]根据居民用电需求行为差异, 利用高斯混合模型进行聚类分析, 定义了对称的广义 Kullback-Leibler 散度对聚类结果进行评价, 并据此选取最佳聚类数, 仿真结果表明该方法能够有效提取用户的典型用电模式。文献[9]将一种新型的密度峰值点快速搜索聚类算法 CFSFDP (Clustering by Fast Search and Find of Density Peaks) 应用到负荷曲线聚类中, 该算法计算复杂度小, 能够适应用户用电数据“量大面广”的特性。

在现有的聚类算法中, K -means 聚类算法因其原理简单、编程容易实现、运行速度较快等优点而被广泛采用^[10]。近年来也有一些文献将智能群体算法引入聚类领域, 并取得了不错的应用效果^[11-12]。然而无论是经典的 K -means 聚类算法还是如粒子群优化 PSO (Particle Swarm Optimization) 算法等智能优化算法, 都存在对初始聚类中心敏感、易陷入局部最优的缺点, 导致聚类质量不高、聚类结果稳定性差等问题。

针对上述问题, 本文利用引力搜索算法 GSA (Gravitational Search Algorithm) 对居民用户用电模式分类开展研究。GSA 由 E. Rashedi 教授等人于 2009 年提出^[13], 是一种基于万有引力定律的智能优

收稿日期: 2017-05-10; 修回日期: 2017-09-06

基金项目: 云南省新能源重大科技专项(2013ZB005); 云南电网有限责任公司科技项目(YNKJQQ00000280); 华北电力大学新能源电力系统国家重点实验室开放课题资助项目(LAPS15009, LAPS16007, LAPS16015)

Project supported by the Major Key Science and Technology Project in Renewable Energy Area of Yunnan Province(2013-ZB005), the Science and Technology Project of Yunnan Power Grid Limited Liability Company(YNKJQQ00000280) and the Open Foundation of State Key Lab of Alternate Electrical Power System with Renewable Energy Sources of North China Electric Power University(LAPS15009, LAPS16007, LAPS16015)

化算法,相比于 PSO 算法,GSA 的全局搜索能力更强,收敛速度更快,不易陷入局部最优解。用户用电模式的分类是一个聚类问题,而聚类本质上是一个非线性优化问题,因此理论上将 GSA 应用于用电模式的分类是可行的。

本文首先剔除了每个用户的异常用电曲线,提取得到各个用户的典型用电曲线;其次,在对居民用电实测数据进行详细分析的基础上提取能够刻画其用电行为的主要特征;然后,利用 GSA 对提取得到的特征进行聚类,采用综合考虑类内紧凑性和类间分散性的聚类评价指标对聚类结果进行评价;最后,基于实测用电数据的聚类结果,对各类用户的用电模式及其参与需求侧响应的潜力进行了分析。

1 特征提取

本文所提方法主要分为 5 个部分:数据预处理、典型负荷曲线提取、特征提取、用电模式聚类以及聚类结果分析,流程图如图 1 所示。下文将对每个部分进行逐一介绍。

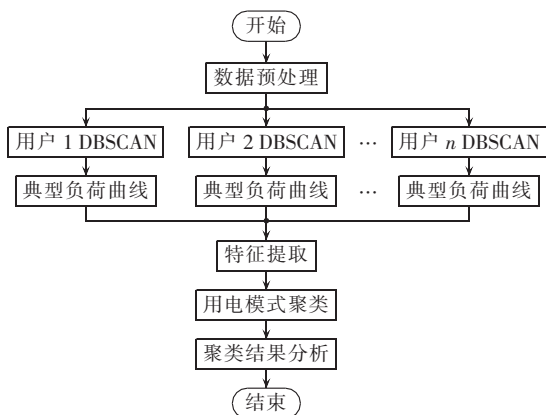


图 1 分类模型流程图

Fig.1 Flowchart of classification model

1.1 数据预处理

本文对某地区居民用户的实测数据进行分析,该数据包含了 220 个家庭在 600 多天时间内的有功功率值,采样时间间隔为 30 min。

本文经过预处理剔除了其中部分数据不全的用户,最终选择了 208 个用户,并对其中第一年的数据(从 2013 年 6 月 1 日开始)进行分析。

为了消除负荷曲线幅值差异对聚类结果的影响,本文在进行聚类前对负荷曲线进行归一化处理,每个用户的每一条日负荷曲线归一化方法为:

$$p'(i)=\frac{p(i)}{\sum_{i=1}^{48} p(i)} \tag{1}$$

其中, $p(i)$ 为归一化前用户第 i 个时段的功率; $p'(i)$ 为归一化后用户第 i 个时段的平均功率。

1.2 典型负荷曲线提取

观察用户的用电曲线可以发现,受极端天气、节假日等因素的影响,用户在某些天的用电曲线与平常相比存在较大差别,主要表现为负荷高峰的峰值以及出现的时间段有所不同,将这些用电曲线记为异常用电曲线。典型负荷曲线提取是指从某一用户一定时间段内的负荷时间序列中剔除这些异常用电曲线,从而提取最能代表用户正常用电形态的负荷曲线^[14]。对同一用户而言,其典型负荷形态与异常负荷形态在各时段的用电负荷具有较大差异,在向量空间中的分布存在较明显的密度差异。DBSCAN (Density Based Spatial Clustering of Applications with Noise) 是一种典型的基于密度的聚类算法,其不需要预先指定聚类数目,能根据聚类对象在空间中的密度分布差异自动聚类成不同的簇,并能够自动滤去偏离较大的噪声点,因此适用于异常用电曲线的识别。该算法主要涉及 2 个参数:搜索半径 ε 和最小对象数目 N_{minpts} 。其基本原理是:将搜索空间中每个对象在搜索半径 ε 范围内覆盖的其他对象的数量与预先设置的 N_{minpts} 进行比较,将数量大于 N_{minpts} 的对象划分为核心点,将数量等于 N_{minpts} 的对象划分为边界点,数量小于 N_{minpts} 的对象归为噪声点。这样,聚类得到的噪声点即为异常用电曲线,剔除这些异常用电曲线后,对剩余曲线取平均即为该用户的典型负荷曲线。

DBSCAN 中 2 个重要参数 ε 及 N_{minpts} 设置得合理与否,直接决定了异常负荷形态能否被识别以及典型负荷曲线提取的效果,本文参考文献[14]提出的经验公式,并在大量实验的基础上,提出了设置密度算法参数的方法。

假设共有 N_{days} 个正常工作日的负荷中异常用电曲线一般不超过 $N_{days}/5$ 条,为了保证异常用电曲线不分为一类,取:

$$N_{minpts}=N_{days}/5 \tag{2}$$

而对同一用户而言,考虑正常工作情况下的负荷波动不超过该用户最大负荷的 $\pm 10\%$,因此搜索半径取为:

$$\varepsilon=0.1\times\sqrt{T}\times P'_{max} \tag{3}$$

其中, T 为一天中采样点的个数,本文取为 48; P'_{max} 为日最大负荷。

本文利用上述方法对各个用户进行了典型负荷曲线提取,其中一个用户 2014 年 3 月的提取结果如图 2 所示。从图 2 中可以看出,DBSCAN 能够有效地检测出异常用电曲线,实现用户典型负荷曲线的提取。

1.3 特征提取

聚类算法按照聚类对象的不同可分为直接聚类和间接聚类,直接聚类不对负荷序列做处理而直接

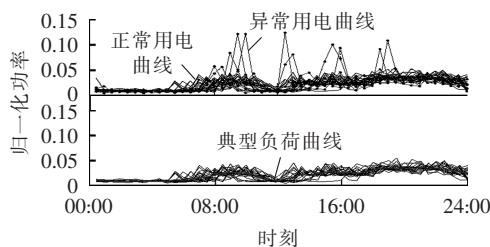


图 2 典型用电模式提取示例

Fig.2 Example of extracting typical power consumption mode

进行聚类;间接聚类则是在对负荷序列进行特征提取后,再对提取得到的特征进行聚类。直接聚类特征维数高、计算复杂度大,考虑到工程实际应用,本文选择间接聚类方法对用户的用电模式进行分析。

特征选择得合理与否对聚类结果的准确性有着重要影响。针对居民用电行为的分析和聚类是为了能够更好地理解用户的用电模式,从而发掘适合参与需求侧响应的用户。需求侧响应的一个主要目的是削峰填谷,因此本文关注的是通过分析处理用户的用电数据,得到数个能够较好地描述用户用电高峰及其季节性变化的特征量,在此基础上进行用电行为的分析和分类。

为了更好地探索每天不同阶段及一年不同季节对用户用电行为的影响,本文将全年分为 4 个季节,按照北半球的标准进行四季划分:春季(3—5 月),夏季(6—8 月),秋季(9—11 月),冬季(12 月至来年 2 月)。此外,根据居民的日常作息习惯,将 1 d 分为 4 个阶段:阶段 1(00:00—06:30, 22:00—24:00),阶段 2(06:30—09:00),阶段 3(09:00—16:00),阶段 4(16:00—22:00)。

为了方便叙述,给出如下变量定义:对于特定的用户,本文定义 $P_i (i=1, 2, 3, 4)$ 为 1 d 各个阶段的年平均功率, σ_i 为对应阶段的平均功率标准差;定义 $\bar{P}$ 为该用户全年的平均功率; P^{sp} 、 P^{su} 、 P^{au} 、 P^{wi} 分别为春、夏、秋、冬 4 个季节的典型负荷曲线(维数为 48 的向量)。利用上述定义,本文提取的特征如下。

a. 特征 1—4。

特征 1—4 选择为用户各阶段全年相对平均功率^[15],定义如下:

$$P_i^R = \frac{P_i}{\bar{P}} \quad i=1, 2, 3, 4 \quad (4)$$

这 4 个特征描述了每天各个阶段负荷高峰的出现情况。

图 3 显示了全年各阶段平均功率(标么值)分布直方图。观察图 3(a)可以看出,在阶段 1(深夜),大部分用户的平均功率都分布在 1 p.u. 的左侧,这表明大部分用户在该阶段都处于负荷低谷期,考虑到用户日常作息习惯,易想到大部分用户在该阶段都处于

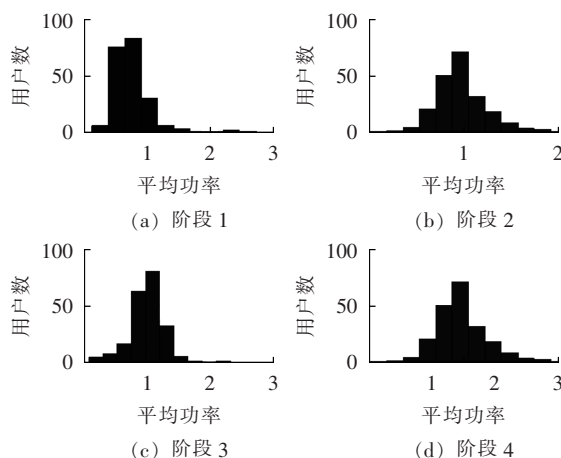


图 3 全年各阶段相对平均功率分布直方图

Fig.3 Histogram of relative average power in each time period of whole year

睡眠状态,因此在该阶段用电较少;相反地,由图 3(d)可以看出,在阶段 4(夜晚),大部分用户的平均功率都分布在 1 p.u. 的右侧,这表明大部分用户的负荷高峰都出现在阶段 4,联系用户的日常作息习惯可以想到用户在该阶段会有做饭、洗衣、休闲娱乐等各种耗能行为,因此图 3(d)所示功率分布与用户的作息习惯是相符的;由图 3(b)、(c)可以看出,与阶段 1 和 4 不同,用户平均功率在阶段 2、3 的分布比较分散,没有出现集中分布的情况,说明在这 2 个阶段不同用户的用电行为存在较大差异,这 2 个特征的区别性较大。

b. 特征 5。

特征 5 选择为用户全年相对平均功率标准差^[15],定义如下:

$$\hat{\sigma} = \frac{\sum_{i=1}^4 \sigma_i / P_i}{4} \quad (5)$$

特征 5 描述了用户全年负荷的平均波动情况,该特征越大表明用户每天负荷的波动越大。用户全年相对平均功率标准差(标么值)分布直方图如图 4 所示。由图 4 可知,各用户全年相对平均功率标准差分布较为分散,表明特征 5 对用户的区分度较强。

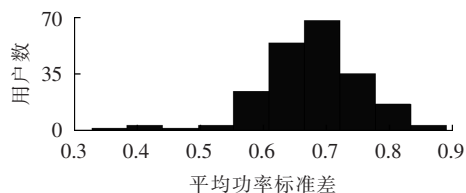


图 4 全年相对平均功率标准差分布直方图

Fig.4 Histogram of relative average power standard deviation of whole year

c. 特征 6。

特征 6 选择为负荷曲线的四季平均相关系数,定义如下:

$$\bar{r} = \text{mean}\{r_{h,l}\} = \text{mean}\left\{\frac{\sum_{t=1}^{48}(P_t^h - \bar{P}^h)(P_t^l - \bar{P}^l)}{\sqrt{\sum_{t=1}^{48}(P_t^h - \bar{P}^h)^2 \sum_{t=1}^{48}(P_t^l - \bar{P}^l)^2}}\right\}$$

$$h, l = \{\text{sp, su, au, wi}\} \quad (6)$$

其中, P_t^h 为 h 季节典型负荷曲线在第 t 个时段的归一化负荷值; $\bar{P}^h$ 为 h 季节典型负荷曲线的平均值; $r_{h,l}$ 为 h 季节与 l 季节典型负荷曲线的相关系数。特征 6 考虑了季节变换对用户用电行为的影响, $\bar{r}$ 越小表示用户四季典型负荷曲线间的相关程度越小, 即用户用电行为随季节变换的变化越明显。

图 5 给出了 2 个用户(用户 50 和用户 203)各个季节的典型负荷曲线, 从图 5(a)可以看出用户 50 一年四季的用电都较为规律, 用电行为随季节变换的变化很小; 相反地, 由图 5(b)可观察到用户 203 四季典型负荷曲线间相似性很小, 主要表现在负荷高峰出现的时间不同, 冬季负荷高峰出现在 08:00—12:00, 而其他 3 个季节则出现在 00:00—04:00。分别计算 2 个用户负荷曲线的四季平均相关系数, 得到用户 50 的 $\bar{r}$ 为 0.868, 而用户 203 的 $\bar{r}$ 则仅为 0.119, 这表明特征 6 能够较好地刻画季节因素对用户用电行为的影响。

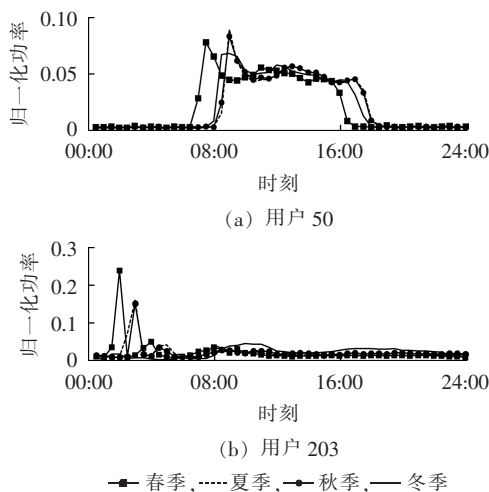


图 5 2 个用户各个季节典型负荷曲线对比
Fig.5 Comparison of typical load curves of two customers in different seasons

综上所述, 共提取出 6 个特征来描述用户的用电行为: 特征 1—4 为用户各阶段全年相对平均功率; 特征 5 为用户全年相对平均功率标准差; 特征 6 为负荷曲线的四季平均相关系数。分别计算各个用户的这 6 个特征, 然后将得到的特征进行聚类。

2 聚类算法

2.1 基于引力搜索的聚类算法

GSA 是一种基于万有引力定律进行寻优的智能优化方法。在 GSA 中, 待优化问题的解被看作是一

组在空间中运动的粒子, 粒子的质量与粒子的适应度大小有关, 适应度越大的粒子其质量也越大。在万有引力的作用下, 质量小的粒子朝着质量较大的粒子移动, 最终使算法收敛到最优解。本文采用 GSA 进行用户用电模式分类模型的研究, 其流程如图 6 所示, 具体的更新公式可参考文献[13]和[16], 这里不再赘述。

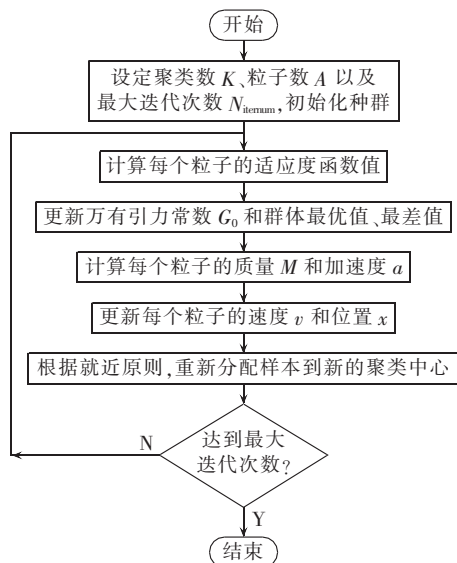


图 6 基于 GSA 的负荷模式聚类流程
Fig.6 Flowchart of load mode clustering based on GSA

每个粒子通过适应度函数来评价其好坏, 适应度函数如式(7)所示。

$$f_{\text{fit}} = \sum_{k=1}^K \sum_{f_j \in C_k} \|f_j - C_k\| \quad (7)$$

其中, K 为聚类数; C_k 为第 k 个聚类中心; f_j 为用户 j 所对应的特征向量, $f_j \in C_k$ 表示用户 j 被归到第 k 个聚类。适应度函数值越小, 则粒子越优。

2.2 聚类评价指标

可以从类内紧凑度和类间分散性 2 个角度评价聚类结果的优劣, 一个好的聚类结果具备 2 个特点: 类内尽可能紧凑, 类间尽可能分散。常见的聚类评价指标有: 均值适宜度(MIA)、戴维森堡丁指数(DBI)、聚类离散度(CDI)、簇内平方和与簇间差异之比(WCBCR)。其中, MIA 指标只考虑了类内紧凑度, 而未考虑类间分散性, 该指标具有一定的片面性; DBI 指标容易出现多个膝点, 不利于最佳聚类数的确定^[10]; CDI 和 WBCRCR 这 2 个指标既考虑了每一类中各个样本点到聚类中心的距离, 同时也将聚类中心间的距离作为评价因素。因此本文选用 CDI 和 WBCRCR 作为评价聚类质量的指标, 其计算公式如下。

a. CDI 指标。

$$\text{CDI} = \frac{1}{\hat{d}(C)} \sqrt{\sum_{i=1}^K \hat{d}(\Omega_k)} \quad (8)$$

其中, $\hat{d}(C)$ 为各个聚类中心间的平均距离, C 为聚类中心向量; $\hat{d}(\Omega_k)$ 为第 k 类内的平均距离, Ω_k 为第 k 类内所有对象。CDI 值越小, 则聚类效果越好。

b. WCBCR 指标。

$$WCBCR = \frac{\sum_{k=1}^K \sum_{f_i \in C_k} d(f_i, C_k)}{\sum_{q=1, q \neq p}^K d(C_p, C_q)}$$

(9)

WCBCR 综合考虑了类内紧凑度和类间分散性, 其值越小表明聚类效果越好。

3 算例分析

3.1 参数分析

3.1.1 粒子数目对聚类结果的影响

为了研究粒子数 A 对聚类结果的影响, 固定其他参数不变 (其中最大迭代次数 $N_{itermax}=300$, 万有引力常数 $G_0=100$, 衰减系数 $\alpha=20$), 改变粒子的数量, 观察聚类结果的变化情况。取 $A \in \{10, 30, 50\}$, 在各粒子数下, 分别对数据集聚类数从 4~15 进行聚类, 聚类评价指标采用 WCBCR, 结果如图 7 所示。

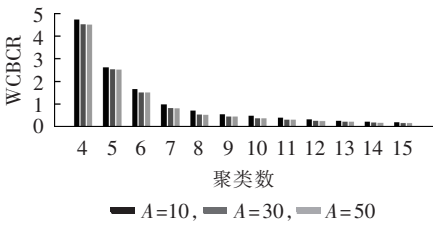


图 7 粒子数对聚类结果的影响
Fig.7 Effect of particle quantity on clustering result

由图 7 可知, 当粒子数从 10 变化到 30 时, WCBCR 指标有较明显的改善, 但当粒子数从 30 变化到 50 时, 指标改善的幅度大幅减小, 几乎不再变化。这是因为: 增加粒子数目能够丰富群体解的多样性, 这就使得 GSA 能够以较大的概率收敛到最优解, 但当粒子数增加到一定数量后, 已有的粒子数已经能够满足寻优的要求, 因此最优解的精度不再发生较大变化; 此外, 随着粒子数的增加, 算法的计算速度会减慢。因此, 在兼顾聚类质量及计算速度的情况下本文选择 30 作为最佳粒子数。

3.1.2 最大迭代次数对聚类结果的影响

为了研究 $N_{itermax}$ 对聚类结果的影响, 固定其他参数不变 (其中 $A=30, G_0=100, \alpha=20$), 改变最大迭代次数, 观察聚类结果的变化情况。取 $N_{itermax} \in \{100, 300, 500\}$, 在各最大迭代次数下, 分别对数据集聚类数从 4~15 进行聚类, 结果如图 8 所示。

由图 8 可知, 与粒子数对聚类结果的影响类似, 当最大迭代次数较小时, 增加其数值, 能给聚类质量

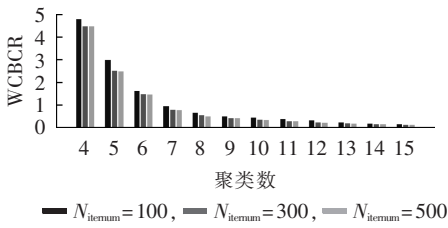


图 8 最大迭代次数对聚类结果的影响
Fig.8 Effect of maximum iteration on clustering result

带来较大的改善, 当最大迭代次数达到一定数值后, 这种改善效果就不再明显。这是因为: 当最大迭代次数为 100 时, 算法还未收敛到最优解就已经停止了, 聚类质量得不到进一步的改善; 而当最大迭代次数达到一定数值后, 算法已经收敛, 再增加迭代次数已经不能改善聚类质量, 反而会增大计算量。因此在后文的分析中, 将最大迭代次数设置为 300。

3.1.3 最佳聚类数的确定

最佳聚类数的确定是聚类算法的一个重要环节。聚类数太少, 则聚类结果过于粗糙, 聚类效果差; 聚类数太多虽然能改善聚类质量但同时会增加计算的复杂度, 不利于实际工程应用。因此最佳聚类数的确定需综合考虑聚类质量与计算复杂度。为了确定最佳聚类数, 设定聚类数从 2~15 变化, 采用 GSA 进行聚类, 其中 GSA 按照前文优化得到的参数进行设置, 采用 CDI 指标对聚类结果进行评价。为了减弱随机性影响, 在每个聚类数下 GSA 运行 20 次, 取这 20 次的指标平均值作为该聚类数下的聚类结果, 结果如图 9 所示。

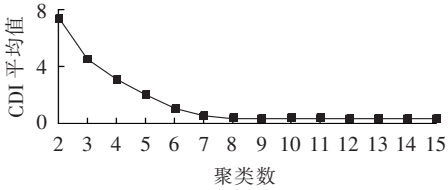


图 9 不同聚类数下 CDI 指标值
Fig.9 CDI values with different cluster numbers

由图 9 可以看出, 随着聚类数的增加, CDI 指标值呈单调下降趋势; 当聚类数在 2~7 变化时, 指标值下降较快; 当聚类数超过 7 后, CDI 指标值的下降趋势明显变缓, 呈现收敛趋势。同时观察图 7 和图 8 可知, WCBCR 指标与 CDI 指标呈现相同的变化趋势。因此本文选择 7 作为最佳聚类数。

3.1.4 万有引力常数 G_0 和衰减系数 α 对聚类结果的影响

万有引力常数 G_0 和衰减系数 α 是 GSA 的 2 个重要参数, 算法提出者 E. Rashedi 给出了这 2 个参数的参考值: $G_0=100, \alpha=20$ 。本文通过固定其他参数

(其中粒子数 $A=30$, 最大迭代次数 $N_{\text{itermax}}=300$, 聚类数 $K=7$), 改变这 2 个参数的取值来探究它们对聚类结果的影响。参考算法提出者给出的取值, 取 $G_0 \in \{40, 60, 80, 100, 120, 140, 160\}$, $\alpha \in \{10, 20, 30, 40, 50, 60, 70\}$, 观察每种参数组合下聚类评价指标 WCBCR 值的变化情况, 结果如图 10 所示。

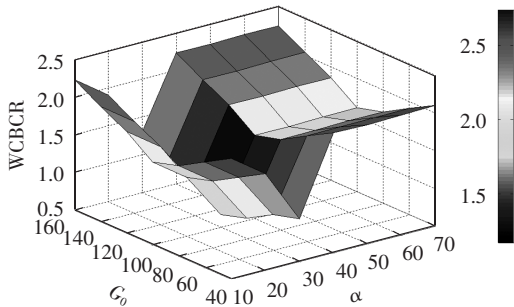


图 10 G_0 和 α 对聚类结果的影响

Fig.10 Effect of G_0 and α on clustering result

由图 10 可知, 随着万有引力常数 G_0 的增加, 指标 WCBCR 先减小后增大, 当 G_0 约为 80 时, WCBCR 的值最小。同理, 当衰减系数 α 较小时, 随着 α 的增大 WCBCR 指标逐渐减小, 当 α 增加到 30 后, 聚类质量不再改善, 反而逐渐变差。

参数 α 起着控制算法收敛速度的作用, 当 α 较小时, 算法的收敛速度较慢, 算法需要更多次迭代才能收敛到最优解, 因此聚类质量不高; 而当 α 的值过大时, 算法收敛速度过快, 容易陷入局部最优从而降低聚类质量。根据上述分析, 在后文的分析中, 将 G_0 设置为 80、 α 设置为 30 以获得最佳的聚类结果。

3.2 聚类结果的分析

利用 GSA 将用户聚成 7 类, 图 11 展示了用户数最多的一类(类别 4)下各个用户的春季典型负荷曲线。表 1 给出了当聚类数为 7 时各类用户各个特征的平均值。图 12 给出了聚类数为 7 时, 各类用户的四季平均负荷曲线。

用电模式的聚类结果有助于电网公司更好地理解与掌握用户的用电行为, 从而根据每类用户的用电特点科学地制定相应的需求侧响应项目。一般而言, 若用户用电行为不规律, 用电模式调节灵活, 则适合参与基于电价的需求侧响应项目; 反之, 若用户

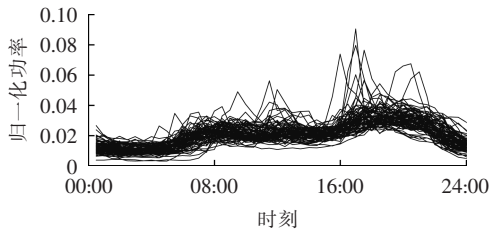


图 11 类别 4 各个用户的春季典型负荷曲线

Fig.11 Typical load curves of customers in cluster 4 in spring

表 1 聚类数为 7 时每类特征的平均值

Table 1 Mean value of each feature when cluster number is 7

用户类别	特征的平均值					
	特征 1	特征 2	特征 3	特征 4	特征 5	特征 6
类别 1	0.64	1.31	1.03	1.33	0.76	0.68
类别 2	1.22	1.06	0.86	0.87	0.67	0.24
类别 3	0.71	0.99	0.99	1.44	0.69	0.52
类别 4	0.65	0.91	1.12	1.43	0.68	0.78
类别 5	2.02	0.52	0.27	0.40	0.43	0.65
类别 6	0.33	1.28	1.85	0.82	0.64	0.82
类别 7	0.82	0.79	0.83	1.66	0.62	0.80

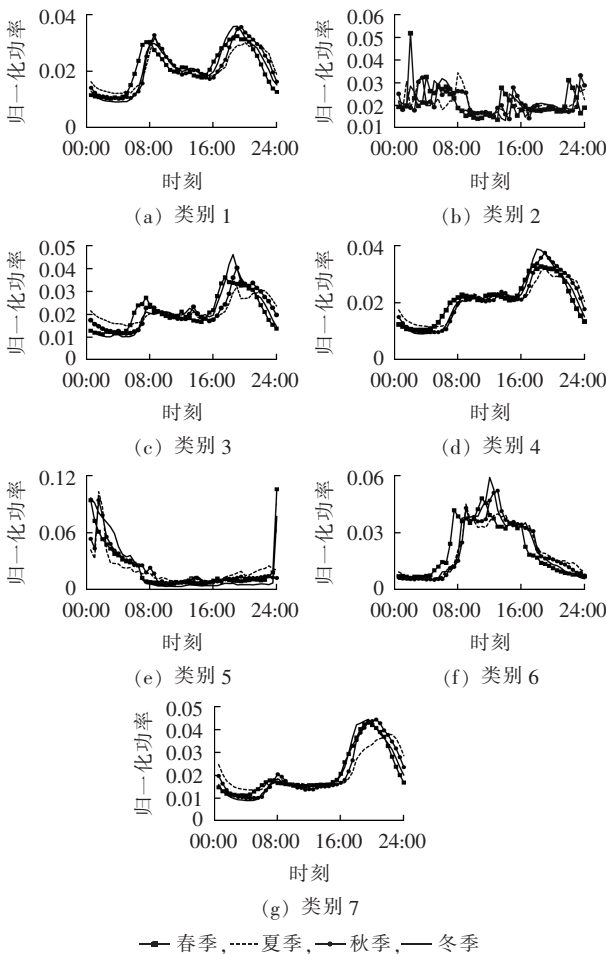


图 12 7 类用户四季平均负荷曲线

Fig.12 Mean load curves of seven types of customers in four seasons

拥有较为固定的用电模式, 用电行为易于预测, 则适合参与基于激励的需求侧响应项目^[9]。这里对各类用户分析如下。

对于类别 1 用户而言, 由图 12(a)可以看出, 其有 2 个明显的负荷高峰, 一个出现在 08:00 左右, 另一个则出现在 19:00 左右, 且这 2 个高峰的幅值相差很小。此类用户的特征 6(负荷曲线的四季平均相关系数)为 0.68, 说明该类用户用电模式较为固定, 受季节变换的影响不大, 适合参与基于激励的需求侧响应项目。

对于类别 2 用户而言,从图 12(b)中可以看出,该类用户一天内存在多个负荷高峰和低谷,负荷曲线波动频繁。其四季平均相关系数仅为 0.24,说明其用电模式受季节变换的影响很大。因此,该类用户适合参与基于电价的需求侧响应项目。

对于类别 3 用户而言,该类用户的用电模式与类别 1 用户相似,同样表现出双峰用电模式。但相比于类别 1 用户,其 2 个用电高峰的幅值相差较大,且四季用电模式更为不规律,该类用户适合参与基于激励的需求侧响应项目。

对于类别 4 用户而言,由图 12(d)可知,该类用户是典型的“谷、平、峰”式用户,即深夜阶段是用电低谷阶段,早上和白天用电比较平稳,在 19:00 左右出现负荷高峰。该类用户的特征 6 为 0.65,说明其用电模式受季节变换的影响较小。因此,该类用户适合参与基于激励的需求侧响应项目。

对于类别 5 用户而言,由图 12(e)可以看出,其负荷高峰出现在阶段 1(00:00—06:30,22:00—24:00),其余阶段用电很少且较为平稳。该类用户的平均功率标准差为 0.43 p.u.,说明其一天用电量变化较小,且其高峰出现在电力系统的谷时段,因此总体而言该类用户参与需求侧响应的潜力较小。

对于类别 6 用户而言,由图 12(f)可知,其负荷高峰出现在阶段 2 和阶段 3(08:00—16:00),其余阶段用电很少。该类用户四季用电模式非常相似(特征 6 为 0.82),受季节变换影响很小,适合参与基于激励的需求侧响应项目。

对于类别 7 用户而言,该类用户的用电行为与类别 4 用户较为相似,同样呈现出“谷、平、峰”的用电模式。但是相比于类别 4 用户,该类用户的峰谷差更大,四季用电更为规律,该类用户适合参与基于激励的需求侧响应项目且其削峰潜力较大。

4 结论

本文提出了一种基于 GSA 的居民用户用电模式分类模型,首先利用 DBSCAN 生成各个用户的典型负荷曲线,在此基础上提取得到了 6 个能够有效反映用户用电高峰变化情况的特征,最后对 208 个居民用户的实测用电数据进行了聚类分析。结果表明本文提取的 6 个特征能够有效刻画用户在不同时间尺度下的用电行为,所提 GSA 能够有效实现居民用户用电模式的分类,分类结果对于电网公司更好地理解与掌握用户用电行为规律具有重要的参考价值。

由于条件限制,本文方法没有在更大规模的数据集上进行实验,探究本文方法对于大规模数据集的适用性或者研究在大规模数据集下 GSA 的改进方法是笔者下一步的研究方向。

参考文献:

- [1] 葛磊蛟,王守相,张明,等. 智能用电条件下用户用能管理与服务平台[J]. 电力自动化设备,2015,35(3):152-156.
GE Leijiao,WANG Shouxiang,ZHANG Ming,et al. Power usage management & service platform in smart electricity utilization condition[J]. Electric Power Automation Equipment,2015,35(3):152-156.
- [2] 栾文鹏,余贻鑫,王兵. AMI 数据分析方法[J]. 中国电机工程学报,2015,35(1):29-36.
LUAN Wenpeng,YU Yixin,WANG Bing. AMI data analytics[J]. Proceedings of the CSEE,2015,35(1):29-36.
- [3] NIZAR A H,DONG Z Y,WANG Y. Power utility nontechnical loss analysis with extreme learning machine method[J]. IEEE Transactions on Power Systems,2008,23(3):946-955.
- [4] YANG H T,CHEN S C,PENG P C. Genetic K-means-algorithm-based classification of direct load-control curves[J]. IEE Proceedings-Generation,Transmission and Distribution,2005,152(4):489-495.
- [5] 张铁峰,顾明迪. 电力用户负荷模式提取技术及应用综述[J]. 电网技术,2016,40(3):804-811.
ZHANG Tiefeng,GU Mingdi. Overview of electricity customer load pattern extraction technology and its application[J]. Power System Technology,2016,40(3):804-811.
- [6] 张素香,刘建明,赵丙镇,等. 基于云计算的居民用电行为分析模型研究[J]. 电网技术,2013,37(6):1542-1546.
ZHANG Suxiang,LIU Jianming,ZHAO Bingzhen,et al. Cloud computing-based analysis on residential electricity consumption behavior[J]. Power System Technology,2013,37(6):1542-1546.
- [7] CHICCO G,NAPOLI R,PIGLIONE F. Comparisons among clustering techniques for electricity customer classification[J]. IEEE Transactions on Power Systems,2006,21(2):933-940.
- [8] HINO H,SHEN H,MURATA N,et al. A versatile clustering method for electricity consumption pattern analysis in households[J]. IEEE Transactions on Smart Grid,2013,4(2):1048-1057.
- [9] WANG Y,CHEN Q,KANG C,et al. Clustering of electricity consumption behavior dynamics toward big data applications[J]. IEEE Transactions on Smart Grid,2016,7(5):2437-2447.
- [10] 白雪峰,蒋国栋. 基于改进 K-means 聚类算法的负荷建模及应用[J]. 电力自动化设备,2010,30(7):80-83.
BAI Xuefeng,JIANG Guodong. Load modeling based on improved K-means clustering algorithm and its application[J]. Electric Power Automation Equipment,2010,30(7):80-83.
- [11] CHICCO G,IONEL O M,PORUMB R. Electrical load pattern grouping based on centroid model with ant colony clustering[J]. IEEE Transactions on Power Systems,2013,28(2):1706-1715.
- [12] MUNSHI A A,MOHAMED Y A R I. Photovoltaic power pattern clustering based on conventional and swarm clustering methods[J]. Solar Energy,2016,124:39-56.
- [13] RASHEDI E,NEZAMABADI-POUR H,SARYAZDI S. GSA:a gravitational search algorithm[J]. Intelligent Information Management,2012,4(6):390-395.
- [14] 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.

- [15] HABEN S, SINGLETON C, GRINDROD P. Analysis and clustering of residential customers energy behavioral demand using smart meter data[J]. IEEE Transactions on Smart Grid, 2015, 7(1):136-144.
- [16] 李鹏, 徐伟娜, 周泽远, 等. 基于改进万有引力搜索算法的微网优化运行[J]. 中国电机工程学报, 2014, 34(19):3073-3079.
- LI Peng, XU Weina, ZHOU Zeyuan, et al. Optimal operation of microgrid based on improved gravitational search algorithm[J]. Proceedings of the CSEE, 2014, 34(19):3073-3079.

作者简介:



苏 适

苏 适(1972—),男,云南昆明人,高级工程师,硕士,研究方向为新能源与智能电网(E-mail:sushil@sina.com);

李康平(1992—),男,浙江海宁人,硕士研究生,研究方向为电力系统负荷特性分析、需求侧响应等(E-mail:lkpncepu@sina.com)。

Classification model of residential power consumption mode based on DBSCAN and gravitational search algorithm

SU Shi¹, LI Kangping², YAN Yuting¹, LU Hai¹, WANG Xinkang²,
LIU Liming², WANG Fei^{2,3}, DONG Ling⁴

(1. Electric Power Research Institute of Yunnan Power Grid Co., Ltd., Kunming 650217, China;

2. State Key Laboratory of Alternate Electrical Power System with Renewable Energy Source, North China Electric Power University, Baoding 071003, China;

3. Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana 61802, USA;

4. State Grid Qinghai Province Electric Power Company, Xining 810008, China)

Abstract: The research on the classification of residential power consumption modes can provide support for the design of demand side response scheme, the load characteristic analysis and its high precision prediction. Firstly, the typical power consumption modes of customers are extracted by DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Then, considering the effect of different periods in a day and season change on the power consumption behaviors of customers, six features are extracted, which can describe the power consumption behaviors of customers in different time scales. Based on this, a classification model of residential power consumption mode based on gravitational search algorithm is proposed. Finally, measured residential power consumption data are clustered and the power consumption modes of customers of each cluster and as well as their potentials for participation in demand side response are analyzed.

Key words: power consumption mode; clustering algorithms; feature extraction; classification; gravitational search algorithm; density-based spatial clustering

(上接第 128 页 continued from page 128)

Voltage fluctuation characteristics of traction power supply system considering AC-DC-AC electric locomotives accessed

ZHANG Guinan, LIU Zhigang, XIANG Chuan, YAO Shulong

(School of Electrical Engineering, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: Aiming at the phenomenon of voltage instability and locomotive blockade in traction power supply system considering the AC-DC-AC electric locomotives accessed, it is proved that this phenomenon originates from a power quality problem in theory instead of a low frequency oscillation. The modulation theory is employed to analyze the mechanism of the symmetrical frequency component as a result of the low frequency signal transmitting in the rectifier. The voltage instability characteristics are analyzed by means of DFT (Discrete Fourier Transform), and the dominant modes are identified by the TLS-ESPRIT (Total Least Square-Estimation of Signal Parameters via Rotational Invariance Techniques) algorithm. On this basis, this phenomenon is defined as a voltage fluctuation problem. The signal is reconstructed by the mathematical expression of voltage fluctuation, and the effectiveness of the proposed methodology is further verified by analyzing the performance evaluation index of the reconstructed signal.

Key words: AC-DC-AC electric locomotive; voltage fluctuation; DFT; symmetrical frequency; TLS-ESPRIT algorithm; modal analysis; signal reconstruction