

基于聚类的电网低压台区用电画像方法研究

时珊珊¹, 田英杰¹, 金瑞杨², 张灏宇², 袁志昌³, 宋洁², 苏运¹

- (1. 国网上海市电力公司电力科学研究院, 上海 200437;
2. 北京大学工学院, 北京 100871;
3. 清华大学电机系, 北京 100084)

摘要:对用电对象进行用电特征分析在电力系统的电力调度、负荷预测、安全性评估等方面具有重要意义。目前的用电特征分析多集中在对用户用电画像方法的研究。低压台区是电网用电中的重要维度,对低压台区进行用电特征分析同样不可或缺。台区用电画像可以帮助电网快速准确地把握台区的负荷特性和用电模式,对挖掘台区用电数据信息并对不同的业务场景进行指导具有重要意义。本文针对台区日冻结量和96点功率数据,提出了台区用电特征标签提取方法并形成标签系统,然后基于聚类技术对获得的台区标签进行聚类分析得到台区画像。最后,本文基于上海市181个台区的用电数据进行案例分析,得到台区用电标签和画像。

关键词:台区画像; 标签提取; 用电特征; 聚类分析

DOI: 10.12067/ATEEE2006020

文章编号: 1003-3076(2020)08-0066-09

中图分类号: TM714

1 引言

现代生产生活对供电量和供电质量要求的显著提高以及社会对供电服务的需求的多样化使得城市电网负荷快速增长,给电力系统的安全经济运行带来了巨大的挑战。因此,在配变在线监测设备^[1]采集到的大量用电数据基础上进行城市用电负荷特性分析,对负载模式分类并构建台区用电画像成为了电网管理者关注的重点之一。

负荷特性分析是进行电网负荷预测、过载预测等多个电网业务场景的基础,也是智能电网能源管理系统的重要组成部分^[2]。按照研究对象的不同,负荷特性分析可分为用户用电特性分析和台区用电特性分析和更大范围的区域用电特征分析。用户用电负荷特性因用户个体差异性和外部环境不同的敏感性,并不能很好地代表一定区域内用电的整体特征。而台区是电网管理的基础单元,因此,更具代表性的台区电网的用电特性分析对电力部门进行台区用电管理、新台区规划和电力调度等工作具有更重要的意义。对于台区的用电模式识别,台区的用电

模式数一般是未知的,有监督的机器学习方法难以处理。聚类是一种有效处理无监督学习问题的方法,聚类技术也已经在电力系统中的多个场景开展应用^[3-5]。本文在台区配变数据的基础上,通过多种方法提取台区特征标签,形成台区标签系统,在台区用电模式数量未知的情况下通过聚类构建台区画像。台区画像是真实台区用电情况的虚拟抽象,是根据台区真实用电数据挖掘出的标签化的用电特征模型,能够通过标签化的用电特征帮助电力部门快速掌握台区用电模式并辅助决策。

目前国内外已经有很多关注基于聚类的用户负荷特性分析的研究,主要是在电力大数据环境下对不同的用户用电指标进行关联性分析,利用数据分析算法挖掘潜在的用户用电行为习惯,以及利用聚类方法对用户进行用电模式分类。文献[6,7]利用模糊聚类方法和自组织映射神经网络对电力用户的负荷曲线进行聚类,得到并分析了用户用电的聚类结果,文献[8,9]提出了电力用户画像的概念,通过多维标签刻画了用户的用电特征,但提取的用户标签相对简单,未形成较全面的标签系统。文献[10]

收稿日期: 2020-06-10

基金项目: 国网上海市电力公司科技项目(52094018002K)

作者简介: 时珊珊(1985-),女,上海籍,高级工程师,博士,研究方向为分布式电源及储能;

金瑞杨(1997-),男,北京籍,博士研究生,研究方向为运筹优化与机器学习。

利用马尔科夫模型对用户用电进行建模并利用基于密度峰值的快速聚类算法进行聚类。目前针对基于聚类的台区负荷特性分析的研究还处在初步阶段^[11]。针对台区负荷与聚类研究多集中于台区线损^[12-14]、台区及用电识别^[15,16]和台区健康评估^[17]。文献[12]提出了一种基于聚类算法的台区线损分析方法,但并未对台区除线损外的用电特征进行分析;文献[15]根据同一台区电能采集设备的电压数据的相关性,通过计算待识别的采集设备归属于各个分类的概率判断其台区归属关系。文献[16]提出了通过对台区用电数据进行聚类,并分析台区静态信息与用电信息的关系来判断未监测台区的用电模式,但聚类的对象仅限于日用电负荷。文献[17]建立了台区指纹与台区健康体系,但其所提取的用于描述台区用电特性的标签系统并不完善。

总的来说,目前用电画像分析方面的研究集中关注用户级别用电特征分析,但针对基于聚类的低压台区用电特征分析研究相对较少,尤其是关于台区用电画像构建的研究还相对空白^[18]。关于台区画像的构建主要的挑战有两点,一是如何在海量数据中提取能够有效刻画台区用电特征的标签,二是如何在多维特征标签的基础上通过聚类技术形成完整有效的台区用电画像系统。

针对以上挑战,本文主要有以下创新与贡献。首先,基于大型城市台区用电的日冻结量和96点功率数据提取标签,相较已有研究获得更加全面的标签系统。通过直接获取、统计分析和数据挖掘的方法提取台区的用电特征标签,得到舒适敏感度、舒适度截止量、用电波动率、电量稳定度等特征标签。另外,本文提出对提取的多维特征标签进行聚类分析的台区用电画像构建方法,能够快速掌握台区用电模式,填补了台区画像研究的空白。最后,基于上海市181个台区的用电数据进行了标签提取和画像构建。台区用电画像方法的框架如图1所示。

本文接下来将在第2节介绍台区用电标签模型的构建,得到标签系统后提出利用聚类技术构建台区用电画像的方法;第3节基于上海市181个台区用电数据进行案例分析;最后在第4节总结本文。

2 台区用电画像模型

台区的日冻结量(日用电量)数据和96点功率数据是构成电力系统用电数据的主要部分,这两部分数据描述了台区的主要用电特征。本文选择台区

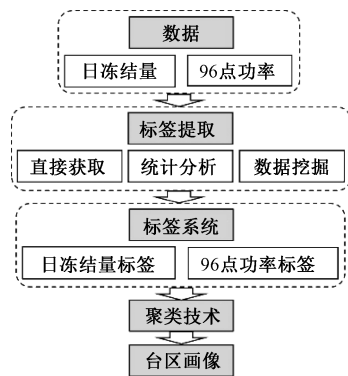


图1 台区用电画像方法框架图

Fig.1 Flow chart of power consumption portrait method of substation area

的日冻结量和96点功率数据作为研究对象进行标签提取和画像构建。日冻结量数据以天为单位记录台区的日用电量,可用于刻画台区在日级别上的用电量特征;96点功率数据每隔15 min采集一次,能描述台区在一天24 h内的负荷变化,可用于刻画台区负荷在较小时间尺度上的负荷特征。本节将结合日冻结量和功率两方面数据,挖掘台区用电的特征标签,构造完整的台区画像系统。

2.1 标签提取

台区用电特征标签是用来评价用电特性不同维度的指标,提取标签的方法一般有直接获取、统计分析、数据挖掘等^[19]。标签通常具有语义化和短文本化的特征,能够让人直观地理解标签的含义,并且标签本身一般不需要更多的数据分析工作,方便直接利用标签信息^[20]。

通过统计分析获得的基本统计量标签包括均值、标准差、最大最小值、负荷峰谷差等^[21]。但基本统计量标签难以对台区负荷特性进行较为全面的刻画。利用数据挖掘方法提取新的用电特征标签并与基本统计量标签相结合能够更加全面地描述台区用电特征。在本节将给出相关标签的具体设计方法。

2.1.1 平均负荷率

不同用电模式的台区负荷曲线的形态有显著差异,是刻画台区用电模式的重要特征。这里将平均负荷率作为表示负荷曲线整体形态的台区标签。平均负荷率 r 的定义如下:

$$r = \frac{P_{\text{avg}}}{P_{\text{max}}} \quad (1)$$

式中, P_{avg} 为一定时间跨度内负荷均值; P_{max} 为该时间跨度内的最大负荷。

平均负荷率标签反映了平均意义下台区的日内负荷波动程度^[22],平均负荷率越高,说明台区的平均负荷水平与负荷峰值的比值较高,日负荷曲线波动较小。

2.1.2 负荷率曲线聚类

为了综合考虑负荷曲线在一天内的特征信息并对台区用电模式进行分类,这里将聚类方法应用于台区负荷率数据,提出用电模式分类标签的挖掘方法。

数据中的异常值和噪声对聚类方法结果的有效性影响较大,在进行聚类分析之前对数据进行清洗,并利用高斯平滑滤波方法过滤掉原本曲线中的高频噪声成分,使得负荷曲线更加平滑。高斯滤波方法以高斯核函数形状选择权值对曲线进行平滑滤波。相比一般的局部平均平滑方法而言,高斯滤波能在平滑处理的同时尽可能保留曲线的边缘信息和细节信息,一定程度上减少后续聚类方法的误差。高斯核函数可以表示为:

$$gauss(x, \sigma) = e^{\frac{-x^2}{2\sigma^2}} \quad (2)$$

方差 σ^2 决定核函数的扁平程度,方差越大高斯函数开口越大,函数越扁平,处理后的负荷曲线越平滑,但会丢掉更多细节信息;反之方差越小函数越陡峭,能够保留更多细节信息,但平滑效果较差。高斯平滑滤波减少了高频随机成分的影响,曲线中真实的台区用电特征因素占比升高,提高了台区负荷信息的信噪比。

负荷率是负荷与监测时间跨度内最大负荷的比值,负荷率曲线在保留负荷曲线形态特征的同时,除去了负荷平均水平的影响。对负荷率曲线进行 K-means 聚类,算法步骤如图 2 所示。

聚类的对象为监测时间内的负荷率序列。

$$\mathbf{R}_i = [r_{i1}, r_{i2}, \dots, r_{in}] \quad (3)$$

式中, \mathbf{R}_i 为第 i 个台区的负荷率曲线; n 为采样点个数,本文选择欧式距离作为样本间距离的度量,其表达式如下:

$$dist(\mathbf{R}_i, \mathbf{R}_j) = \sqrt{\sum_{p=1}^n (r_{ip} - r_{jp})^2}, \quad \forall i \neq j \quad (4)$$

选择轮廓系数作为聚类结果的评价指标,寻找最优的分类簇数 K 作为最终的台区负荷率模式数。

2.1.3 用电量稳定度

台区负荷的日冻结量数据是具有较强周期性的时间序列数据,其中的周期性成分是可以合适的方法精确预测的,而高频的噪声无法预测,属于预

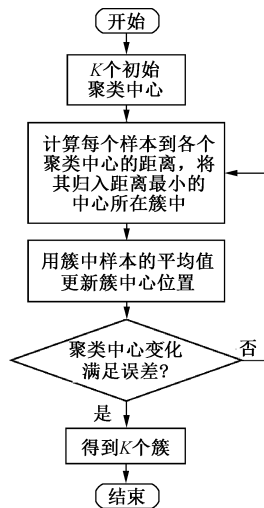


图 2 K-means 算法流程图

Fig.2 Flow chart of K-means algorithm

测系统误差。负荷样本的规律性决定了负荷预测所能达到的精确度上限,无法通过改进算法模型来超越。这里引入用电量稳定度作为评估预测精度上限和下限的标签^[23]。

对日冻结量时间序列做有限傅立叶分解不同的频率成分重新组合,按照不同日冻结量成分的周期性差异将曲线分解为如下 3 个分量:日冻结量曲线周分量是曲线中以周为周期重复变化的分量;日冻结量曲线的低频分量是曲线中体现长期缓慢变化趋势的分量;日冻结量曲线中的高频分量是曲线中体现短期波动的随机分量。

分解得到的 3 个成分中周分量具有明显的周期性,属于可预测成分;高频分量变化剧烈,且没有规律性,属于不可预测成分;低频分量的可预测性由所选预测模型决定。因此根据低频分量的可预测程度,得到用电量稳定度上限和下限的概念。

负荷预测精度的上限是能精确预测除了高频随机分量之外的成分,因此电量稳定度上限的表达式为:

$$L_U = \left[1 - \sqrt{\frac{\sum_{i=1}^N \left(\frac{|H(i)|}{S(i)} \right)^2}{N}} \right] \times 100\% \quad (5)$$

式中, $S(i)$ 为原始日冻结量序列; $H(i)$ 为上述傅立叶分解得到的日冻结量序列高频成分。

日冻结量曲线中的低频成分是长期缓慢变化的时间序列,一般情况下主要受气候、温湿度等外界因素的影响,并不总能精确预测,属于部分可预测分量。由此引出电量稳定度下限的表达式:

$$L_D = \left[1 - \sqrt{\frac{\sum_{i=1}^N \left(\frac{|H(i) + L(i)|}{S(i)} \right)^2}{N}} \right] \times 100\% \quad (6)$$

式中, $L(i)$ 为通过傅立叶分解得到的负荷序列低频成分。

2.1.4 用电波动率

与平均负荷率类似,用电波动率可以有效地表示台区日冻结量曲线的整体形态,给出用电波动率 r 的表达式为:

$$r = \frac{S_{\text{avg}}}{S_{\text{max}}} \quad (7)$$

式中, S_{avg} 为监测时间内日冻结量的均值; S_{max} 为监测时间内的日冻结量最大值。若监测时段时间跨度为 K 周,则监测时间内的平均用电波动率如下:

$$r = \frac{1}{K} \sum_{i=1}^K r_i \quad (8)$$

式中, r_i 为第 i 周的用电波动率。用电波动率反映了监测时间内日冻结量曲线的波动幅度,用电波动率越低,该监测时间内峰谷差越大,日冻结量曲线的波动幅度越大。

2.1.5 舒适敏感度和舒适截止量

台区用电行为与外部天气因素尤其是温度和湿度密切相关^[24],因此利用复合指标舒适度指数来反映二者综合作用的效果。人体舒适度是从气象学角度评价人类在不同气候条件下舒适感的一项生物气象指标,关于人体舒适度的量化方法有很多研究,本文选择其中应用最为广泛的温湿指数^[25]作为舒适度指数,其定义如下:

$$C = T - 0.55(1 - U)(T - 14.5) \quad (9)$$

式中, C 为舒适度指数; T 为日平均气温; U 为相对湿度。因为综合考虑了温湿度的影响,舒适度指数能更为全面地刻画台区日冻结量和外部天气因素的联系。

这里选取其中一个台区日冻结量与舒适度指数的关系图如图3所示。台区日冻结量与舒适度指数的拟合曲线为近似3段的分段线性函数,在高温区域和低温区域日冻结量和舒适度指数近似为线性关系,而在舒适温度区域日冻结量近似为常数,符合台区实际用电情况。这里将高温区和低温区的拐点分别定义为高温舒适截止量和低温舒适截止量,将高温和低温区域拟合曲线的斜率绝对值分别定义为高温舒适敏感度和低温舒适敏感度。

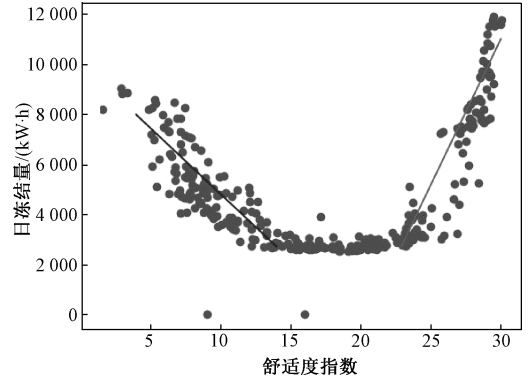


图3 台区日冻结量与舒适度指数的拟合曲线

Fig.3 Fitting curve of daily power consumption with comfort index

舒适截止量的确定通过以最小化损失函数为准则选取拐点作为舒适截止量。由于高低温区域的拐点是关注的重点,可以将高低温区间分开考虑选取高温和低温拐点。设舒适度范围为 $C \in [C_l, C_h]$, 高低温区域以 C_0 划分。以高温区域 $[C_0, C_h]$ 为例,设拐点为 C_p ,则得到高温区域日冻结量与舒适度的回归方程:

$$\hat{P} = \hat{\beta}C \quad (10)$$

式中, $\hat{\beta}$ 为舒适度的回归系数,拟合曲线相应的损失函数为:

$$L(C_p, \hat{\beta}) = \frac{1}{n_h} \sum_{i=1}^{n_h} (P_i - \hat{P}_i)^2 \quad (11)$$

式中, i 为舒适度大于 C_p 的对应日冻结量下标; n_h 为舒适度大于 C_p 的样本数量。则所求高温舒适截止量即为最小化损失函数对应的舒适度:

$$C_p^* = \arg \min_{C_p} L(C_p, \hat{\beta}) \quad (12)$$

其对应估计的参数的绝对值即为高温舒适敏感度。相同的方法应用于低温区域可获得低温舒适截止量和敏感度。舒适敏感度刻画了天气变化造成舒适度变化而引起台区日冻结量的变化,舒适敏感度越高表明台区日冻结量对舒适度变化越敏感。舒适截止量则刻画了台区的舒适区间,在舒适区间内台区的用电需求基本不变。舒适敏感度和舒适度截止量是描述台区日冻结量随气候外部因素的变化的重要标签。

2.1.6 日冻结量曲线聚类

在保留日冻结量曲线所有信息的情况下利用聚类算法对台区用电曲线进行聚类能够挖掘台区日冻结量模式。聚类的对象为监测时间内日冻结量序列

$$S_i = [s_{i1}, s_{i2}, \cdots, s_{in}] \tag{13}$$

式中, S_i 为第 i 个台区的日冻结量曲线; n 为日冻结量采样点个数, 这里对日冻结量选择一年 365 天的数据进行研究, 同样选择欧式距离作为样本间距离的度量, 轮廓系数作为聚类效果的评价指标, 并选择轮廓系数最大的聚类结果对应簇的数量作为最终日冻结量模式数。

2.2 图像构建

基于已得到的台区负荷的多维度特征标签, 本文选择 K -means++ 聚类算法对标签进行聚类分析, 建立完整的台区用电画像系统。初始聚类中心的选择对 K -means 算法的聚类结果有较大影响, 随机地初始化聚类中心有时不能得到很好的聚类结果, 改进的 K -means++ 算法来进行聚类分析可以避免上述问题。

K -means++ 算法在 K -means 算法的基础上优化了初始化聚类中心这一步骤。 K -means++ 算法首先从样本集中随机选择一个样本作为第一个聚类中心, 然后对样本集中的每个样本计算它与其最近的聚类中心的距离 $D(x)$, 然后按照 $D(x)$ 越大的点被选作新的聚类中心的概率越高的原则选出一个样本点作为新的聚类中心。重复该过程直到选出 K 个聚类中心, 完成聚类中心的初始化。之后再执行 K -means 算法得到聚类结果。 K -means++ 算法步骤如图 4 所示。

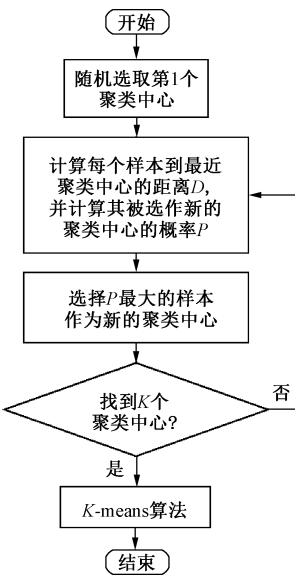


图 4 K -means++ 算法流程图

Fig.4 Flow chart of K -means++ algorithm

通过前文介绍的标签提取方法, 从台区用电数

据中挖掘出一套标签系统来描述台区用电的特征信息。将提取的特征标签信息作为台区的标签向量, 对其进行聚类分析, 就能构建台区用电画像系统, 结合多个维度的特征信息, 将台区用电模式进行分类。

3 案例分析

本文基于上海市 181 个台区的用电数据进行案例分析, 其中日冻结量包括 2017. 6. 1 ~ 2018. 5. 31 一年的数据, 96 点功率包括 2018. 5. 1 ~ 2018. 5. 31 共 31 天的数据。

通过前文提出的标签提取方法获得台区的电量稳定度、用电波动率、舒适敏感度、96 点功率平均值、负荷稳定度、平均负荷率、日冻结量曲线聚类结果和 96 点负荷率曲线的聚类结果作为特征标签。得到的标签和台区用电画像。

3.1 台区 96 点功率标签

负荷率曲线的聚类结果如图 5 所示, 得到的不同类别的台区的日用电模式存在明显差异。第一类的台区日用电曲线存在上午和晚间两个用电高峰, 晚高峰的峰值更高; 第二类台区存在早间和午间两个峰值较低的用电小高峰和峰值较高的用电晚高峰; 第三类台区存在早间和晚间两个用电高峰, 两个时段的用电峰值非常接近; 第四类台区存在早间、午间两个峰值较低的用电高峰和晚间一个峰值较高的用电高峰, 其早间和午间的用电高峰峰值比第二类台区更高。将以上 4 个簇按照顺序依次定义为类型 1, 2, 3, 4。

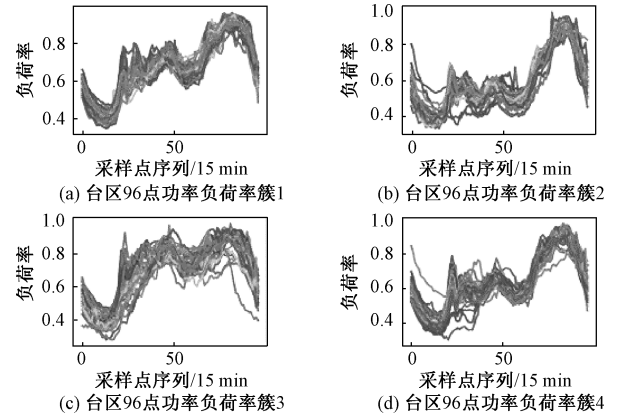


图 5 96 点功率负荷率曲线聚类结果

Fig.5 Results of clustering for 96-point load curve

根据 181 个台区的实际数据结果, 对于台区 96 点功率的平均值标签, 定义 50 ~ 100 为功率均值标签的低水平区间, 100 ~ 150 为中等水平区间, 150 以

上为高水平区间;对于功率的平均负荷率标签,定义0.5~0.6为负荷率标签的低水平区间,0.6~0.7为中等水平区间,0.7以上为高水平区间。

3.2 台区日冻结量标签

不同台区的用电波动率如图6所示。结果表明,大部分台区的用电波动率标签集中分布在0.2~0.4区间内。部分台区的波动率标签在0~0.2区间内,这部分台区长期以较低水平的用电量运行,有短期超高水平的高峰用电,用电模式变化剧烈,进行异常用电模式监测时需要重点关注;另有极少数量的台区波动率标签超过0.4,这类台区的负荷长时间较接近最大水平,高峰用电水平较平均用电水平增量相对很小,用电变化平缓,属于用电模式较为平稳的台区。据此,本文将用电波动率0~0.2定义为低波动率区间,0.2~0.4为中波动率区间,0.4~1为高波动率区间。

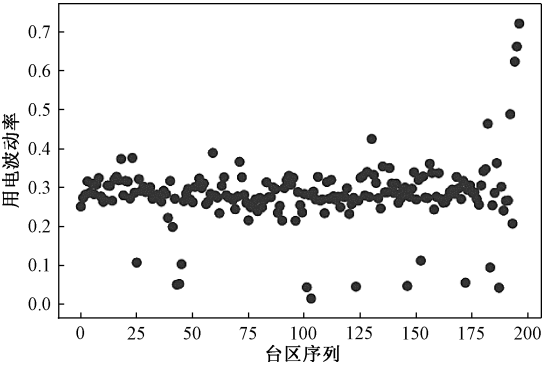


图6 台区用电波动率

Fig.6 Fluctuation rate of power consumption of each substation area

不同台区的用电量稳定度标签结果如图7所示,结果表明,大部分台区电量稳定度上限标签都能达到超过0.95的较高水平,且分布比较集中,极少数台区稳定度上限标签在0.9以下;大部分台区电量稳定度下限标签分布在0.5~0.8区间内,极少数在0.5以下,这类台区的日冻结量曲线中不可预测成分占比较高,对其进行负荷预测的难度较大。不同台区电量稳定度上限区别差异不大,因此本文根据电量稳定度下限的结果对台区进行区分。定义电量稳定度下限数据的0~0.5为低稳定度区间,0.5~0.8为中稳定度区间,0.8~1为高稳定度区间。

台区日冻结量曲线的聚类结果如图8所示。对目标台区进行聚类分析的结果表明,不同簇内台区的日冻结量曲线的曲线形状大体相似,都呈现双峰分布,在夏季和冬季分别出现高峰用电情况,且夏季

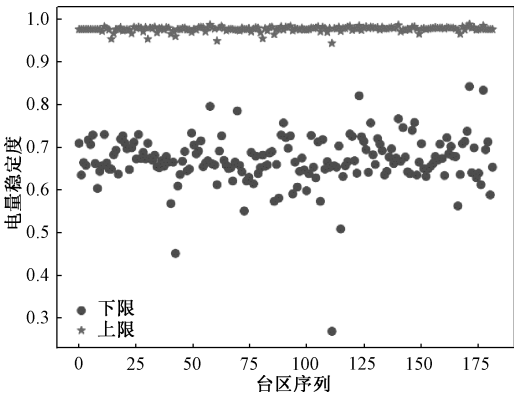


图7 台区用电量稳定度

Fig.7 Power consumption stability of each substation area

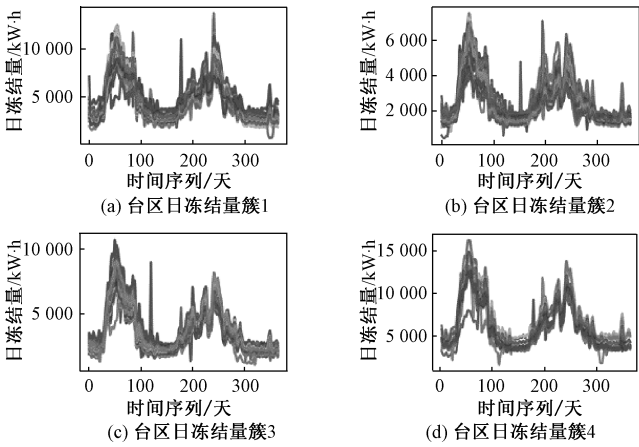


图8 台区日冻结量聚类结果

Fig.8 Result of clustering for daily power consumption of substation areas

峰值稍高于冬季峰值,曲线区别主要体现在数量级水平上,据此将台区日冻结量曲线的形态分为类型1、类型2、类型3、类型4四个类别。

对于舒适敏感度标签,根据181个台区的数据结果,将0~1000定义为低舒适敏感度,1000~1500为中舒适敏感度,1500以上为高舒适敏感度。

3.3 台区用电画像

以上得到的台区日冻结量和功率标签反映了台区负荷的多方面特性,基于标签数据进行聚类分析得到最终的台区用电画像系统。

连续型标签的聚类结果如图9所示,这里的标签进行了归一化处理便于比较,聚类结果图中展示了每类台区聚类中心的值。选择聚类簇数为5进行标签聚类分析时的轮廓系数最高,结果最优。在K-means++聚类算法的结果中,按照特征标签分为了5类台区。

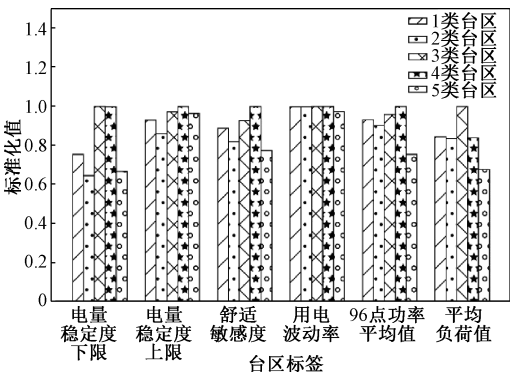


图9 台区连续型标签聚类结果

Fig.9 Result of clustering for continuous labels of substation areas

表 1 的台区画像结果是对标签聚类结果中属于每个类别的所有台区的各个标签取平均值,再对应前文定义的每个标签的离散分类区间得到的每个标签的离散变量结果。

表 1 台区画像结果

Tab.1 The result of the portraits of the substation areas

台区类别	电量稳定度	用电波动率	舒适敏感度	日冻结量类型	平均负荷率	功率均值	负荷率类型
1 类	中	中	中	类型 1	中	中	类型 4
2 类	中	低	中	类型 3	低	低	类型 2
3 类	高	中	高	类型 4	高	高	类型 1
4 类	高	高	中	类型 4	高	高	类型 3
5 类	低	低	低	类型 2	中	低	类型 1

对上海市真实台区用电数据的画像构建结果实现了本文提出的基于标签提取和聚类分析的画像构建方法,可以帮助电网管理者快速地把握台区用电特征,为后续开展台区负荷预测、新台区规划等工作提供了基础。

4 结论

本文将更具区域代表性的低压台区用电数据作为研究对象,提出一套用于台区用电特征分析的标签系统,并基于该标签系统利用聚类技术进行台区用电画像构建。该方法构建的台区画像能快速、准确的识别不同用电模式的台区,协助电力系统管理者进行多个业务场景下的决策。首先,文中基于台区日冻结量和 96 电功率数据,结合统计分析、数据挖掘以及机器学习方法提出了负荷稳定度、舒适敏感度等台区用电特征标签;其次,对得到的台区用电特征标签进行 K-means++ 聚类分析,构建了台区画

像系统;最后通过对上海市 181 个台区用电数据进行标签提取和台区画像构建,验证了台区画像构建方法的效果,对于后续开展台区负荷预测、新台区规划等工作提供了重要基础。

尽管如此,对于台区画像方法的研究仍然有很大的拓展空间和方向。未来还可以考虑将台区的社会信息数据如经济数据、地理数据等纳入标签系统,获取影响台区用电特征的更全面的标签。同时,本文台区均来自上海市某区,总体上具有相似的用电特征,未来可考虑对类型更多样化的台区进行画像构建。

参考文献 (References):

[1] 肖宏, 陆新, 乔飞, 等 (Xiao Hong, Lu Xin, Qiao Fei, et al.). 台区低压配电载波通信在线监测装置 (The area of low voltage distribution carrier wave communication online monitoring device) [J]. 信息通信 (Information&Communications), 2018 (3): 101-103.

[2] 朱斌, 姜宁, 霍雪松, 等 (Zhu Bin, Jiang Ning, Huo Xuesong, et al.). 南京城市电网核心区短期负荷特性分析及预测 (Forecasting and studies on load characteristics of Nanjing center area power network) [J]. 中国电力 (Electrical Engineering), 2016, 48 (2): 67-72, 89.

[3] 许晔, 郭谋发, 杨耿杰, 等 (Xu Hua, Guo Moufa, Yang Gengjie, et al.). 采用仿射不变矩的谐振接地系统故障选线方法研究 (New approach to detect fault line in resonant earthed system using affine moment invariants) [J]. 电工电能新技术 (Advanced Technology of Electrical Engineering and Energy), 2015, 34 (7): 37-44.

[4] 曲朝阳, 熊泽宇, 颜佳, 等 (Qu Chaoyang, Xiong Zeyu, Yan Jia, et al.). 基于 Spark 的电力设备在线监测数据可视化方法 (Visualization method of electrical equipment online monitoring data based on Spark) [J]. 电工电能新技术 (Advanced Technology of Electrical Engineering and Energy), 2016, 35 (11): 72-80.

[5] 宋人杰, 刘瑞英, 王林 (Song Renjie, Liu Ruiying, Wang Lin). 灰色定权聚类和变权模式在变压器状态评估中的应用研究 (Application of grey fixed weight clustering and variable weight model in transformer condition evaluation) [J]. 电工电能新技术 (Advanced Technology of Electrical Engineering and Energy), 2017, 36 (3): 75-80.

[6] 冯晓蒲, 张铁峰 (Feng Xiaopu, Zhang Tiefeng). 基于实际负荷曲线的电力用户分类技术研究 (Research on

- electricity users classification technology based on actual load curve) [J]. 电力科学与工程 (Electric Power Science and Engineering), 2010, 26 (9): 18-22.
- [7] 吴晶莹, 吴为麟, 宋保明 (Wu Jingying, Wu Weilin, Song Baoming). 基于自组织映射神经网络的电力用户负荷曲线聚类 (Power customers load profile clustering using the SOM neural network) [J]. 电力系统自动化 (Automation of Electric Power Systems), 2008, 32 (15): 66-70, 78.
- [8] 陆俊, 朱炎平, 彭文昊, 等 (Lu Jun, Zhu Yanping, Peng Wenhao, et al.). 智能用电用户行为分析特征优选策略 (Feature selection strategy for electricity consumption behavior analysis in smart grid) [J]. 电力系统自动 (Automation of Electric Power Systems), 2017, 41 (5): 58-63, 83.
- [9] 徐涛, 黄莉, 李敏蕾, 等 (Xu Tao, Huang Li, Li Minlei, et al.). 基于多维细粒度行为数据的居民用户画像方法研究 (Research on portrait method of residential users based on multi-dimensional fine-grained) [J]. 电力需求侧管理 (Power Demand Side Management), 2019, 21 (3): 47-52, 58.
- [10] Wang Yi, Chen Qixin, Kang Chongqing, et al. Clustering of electricity consumption behavior dynamics toward big data applications [J]. IEEE Transactions on Smart Grid, 2016, 7 (5): 2437-2447.
- [11] 蒋玮, 黄丽丽, 祁晖, 等 (Jiang Wei, Huang Lili, Qi Hui, et al.). 基于分布式图计算的台区负荷预测技术研究 (Research on load forecasting technology of transformer areas based on distributed graph) [J]. 中国电机工程学报 (Proceedings of the CSEE), 2018, 38 (12): 3419-3430.
- [12] 张洋瑞, 杨鹏, 申洪涛 (Zhang Yangrui, Yang Peng, Shen Hongtao). 基于聚类算法的低压台区线损分析方法研究 (Research of low-voltage distribution network loss based on cluster analysis) [J]. 河北电力技术 (Hebei Electric Power), 2017, 36 (2): 8-10.
- [13] 李亚, 刘丽平, 李柏青, 等 (Li Ya, Liu Liping, Li Baiqing, et al.). 基于改进 K-Means 聚类和 BP 神经网络的台区线损率计算方法 (Calculation of line loss rate in transformer district based on improved K-means clustering algorithm and BP neural network) [J]. 中国电机工程学报 (Proceedings of the CSEE), 2016, 36 (17): 4543-4551.
- [14] 王守相, 周凯, 苏运 (Wang Shouxiang, Zhou Kai, Su Yun). 基于随机森林算法的台区合理线损率估计方法 (Line loss rate estimation method of transformer district based on random forest algorithm) [J]. 电力自动化设备 (Electric Power Automation Equipment), 2017, 37 (11): 39-45.
- [15] 阿辽沙·叶, 顾君, 张小秋 (Ye Aliaosha, Gu Jun, Zhang Xiaoqiu). 基于数据时空相关性的智能台区识别技术 (Intelligent transformer room recognition technology based on temporal and spatial data correlation) [J]. 现代电子技术 (Modern Electronics Technique), 2018, 41 (4): 38-41.
- [16] Li Ran, Gu Chenghong, Li Furong, et al. Development of low voltage network templates-part I: Substation clustering and classification [J]. IEEE Transactions on Power Systems, 2015, 30 (6): 3036-3044.
- [17] 陈灵君, 张叶, 吴颖晖, 等 (Chen Lingjun, Zhang Ye, Wu Yinghui, et al.). 基于 K-means 算法的配电网台区健康评估 (Evaluation of health status of low-voltage distribution network based on K-means algorithm) [J]. 供用电 (Distribution & Utilization), 2019, 36 (2): 42-47.
- [18] 朱健峥 (Zhu Jianzheng). 基于群集智能手段的低压台区电网负荷预测手段研究 (Study on forecasting method for low voltage area load based on swarm intelligence) [D]. 上海: 上海交通大学 (Shanghai: Shanghai Jiao Tong University), 2011.
- [19] 余向前, 王林信, 李云冰, 等 (Yu Xiangqian, Wang Linxin, Li Yunbing, et al.). 电力营销信息化客户画像的应用研究 (Application research on the customer's portrait of electric power marketing informatization) [J]. 计算技术与自动化 (Computing Technology and Automation), 2017, 36 (4): 122-126.
- [20] 吕辉, 许道强, 仲春林, 等 (Lv Hui, Xu Daoqiang, Zhong Chunlin, et al.). 基于电力大数据的标签画像技术与应用研究 (Study on tag portrait technology based on electric power big data and its application) [J]. 电力信息与通信技术 (Electric Power Information and Communication Technology), 2017, 15 (2): 43-48.
- [21] 赵莉, 候兴哲, 胡君, 等 (Zhao Li, Hou Xingzhe, Hu Jun, et al.). 基于改进 K-means 算法的海量智能用电数据分析 (Improved K-means algorithm based analysis on massive data of intelligent power utilization) [J]. 电网技术 (Power System Technology), 2014, 38 (10): 2715-2720.
- [22] 郇嘉嘉, 左郑敏, 程鑫, 等 (Huan Jiajia, Zuo Zhengmin, Cheng Xin, et al.). 外部因素对广东省电网负荷特性的影响分析 (Influence analysis of external factors on load characteristics of Guangdong power grid) [J]. 水电能源科学 (International Journal Hydroelectric Energy), 2018, 36 (5): 210-213.
- [23] 康重庆, 苗键强, 韦仲康, 等 (Kang Chongqing, Miao Jianqiang, Wei Zhongkang, et al.). 短期负荷的标准

化定义模型 (Standard definition model for short-term load) [J]. 电力科学与技术学报 (Journal of Electric Power Science and Technology), 2016, 31 (1): 3-10.

[24] 谢宏, 陈志业, 牛东晓, 等 (Xie Hong, Chen Zhiye, Niu Dongxiao, et al.). 基于小波分解与气象因素影响的电力系统日负荷预测模型研究 (The research of dai-

ly load forecasting model based on wavelet decomposing and climatic influence) [J]. 中国电机工程学报 (Proceedings of the CSEE), 2001, 21 (5): 6-11.

[25] Thom E C. The discomfort index [J]. Weatherwise, 1959, 12 (2): 57-60.

Research on power consumption portrait method of low-voltage substation areas based on clustering

SHI Shan-shan¹, TIAN Ying-jie¹, JIN Rui-yang², ZHANG Hao-yu²,
YUAN Zhi-chang³, SONG Jie², SU Yun¹

- (1. State Grid Shanghai Electric Power Research Institute, Shanghai 200437, China;
2. College of Engineering, Peking University, Beijing 100871, China;
3. Department of Electrical Engineering, Tsinghua University, Beijing 100084, China)

Abstract: It is of great significance to analyze the characteristics of power consumption for power dispatching, load forecasting and safety assessment in the power system. The research about the characteristics analysis of power consumption mainly focuses on the user portrait construction method. The power consumption of low-voltage substation areas is an important part and the corresponding analysis is also significant. The portrait method of low-voltage substation area can help the grid manager to get the load characteristics and power consumption mode quickly and accurately. In this paper, the label system is formulated by label extraction based on the data of daily power consumption and 96-point daily load. Then the portrait of the substation area is constructed by clustering method. Moreover, the label system and portraits of 181 substation areas in Shanghai are derived based on real data and our portrait method.

Key words: portrait of substation area;label extraction;power consumption characteristics;clustering analysis