

基于特征优选和加权聚类的商场用电模式分析

张美霞¹, 李泰杰¹, 杨秀¹, 蔡鹏飞¹, 张勇², 方陈²

(1. 上海电力大学电气工程学院, 上海 200090; 2. 上海市电力公司电力科学研究院, 上海 200437)

摘要:随着用电信息的采集完善, 准确的用户用电模式分析将为电力智能化建设提供重要依据, 在此背景下, 针对用电模式分析中考虑聚类特征类型单一的问题, 提出一种考虑多类型特征优选的加权聚类分析方法。首先, 将负荷类特征和气象类特征归一化建立待选择特征集合; 然后, 结合互信息和灰色关联度优选出聚类特征集; 最后, 采用权重分配的 k -means 方法对优选特征集合进行聚类, 结合负荷曲线分析各用电模式的典型用电行为。通过对上海市某商场用电负荷数据的分析, 证明该方法能够减少数据冗余信息的干扰并提升聚类质量。

关键词:特征优选; 加权聚类; 商场负荷; 用电模式

DOI:10.19781/j.issn.1673-9140.2021.05.017 中图分类号:TM73 文章编号:1673-9140(2021)05-0137-07

Analysis of power consumption mode for shopping malls based on feature selection and weighted clustering

ZHANG Meixia¹, LI Taijie¹, YANG Xiu¹, CAI Pengfei¹, ZHANG Yong², FANG Chen²

(1. College of Electric Engineering, Shanghai University of Electric Power, Shanghai 200090, China;

2. State Grid Shanghai Electric Power Research Institute, Shanghai 200437, China)

Abstract: With the eventually improvement of power consumption information collection, the accurate analysis for user power consumption mode will provide an important basis for power intelligent construction. When analyzing power consumption modes, the load is the only clustering feature to be taken into account. Therefore, a weighted clustering analysis method considering multi-type feature selection is proposed. Firstly, the load and meteorological features are normalized to establish a feature set to be selected. Then, the clustering feature set is selected by combining mutual information and grey correlation degree. Finally, the weighted k -means algorithm is utilized to cluster the selected feature sets, and the typical behavior of each power consumption mode is analyzed with the load curve. Through the analysis of the electrical load data of a shopping mall in Shanghai, it is proved that this method can reduce the interference of redundancy information and improve the clustering quality.

Key words: feature selection; weighted clustering; load of shopping mall; power consumption mode

随着智能电网的快速发展和用电信息采集系统的逐渐完善, 电力公司积累了大量楼宇用电负荷数据^[1-2]。通过数据分析, 探究用户的用电模式和特性, 帮助电力公司了解用户的用电习惯与个性化需求, 为需求侧响应方案的制定与电力智能化建设提

供数据支撑^[3]。对用户而言, 可以根据电力公司提供的用电分析报告, 调整用电方式, 并达到节能用电的目的^[4]。

目前, 国内外对用电模式分析的研究主要侧重于聚类方法的改进和应用, 包括初始聚类中心的选

取、聚类算法的融合和新算法的应用。文献[5]采用数据密度参数方法确定初始聚类中心,有效地提高了对工业用户的聚类分析质量;文献[6]提出了一种余弦距离和欧式距离融合的双层聚类方法,进行地区负荷识别时,能够兼顾负荷曲线的形态和数值差异;文献[7]采用无监督深度学习的方法提取香港某教学楼宇的用电模式特征,并在此基础上进行负荷预测,取得了良好的预测效果。少部分学者对聚类特征的分析 and 选择进行了研究,文献[8]建立了以日峰谷差、日负荷率、日最小负荷率等为特征的数据模型,并采用 ReliefF 算法进行权重计算和特征选择;文献[9]分析了负荷类特征对聚类分析的有效性,并提出了一种自适应的特征优选策略,提高了用户行为分析的效率。

大部分研究以负荷曲线为聚类对象,或只对负荷类特征进行了分析选择,所考虑的聚类特征类型较为单一,其聚类对象的有效性有待验证,聚类质量可以进一步提升。该文在负荷类特征的基础上,结合体感温度、气温等气象类特征,建立待选择特征集合;考虑互信息与灰色关联度选择结果的差异性,提出一种加权关联度的方法对聚类特征集合进行分析和优选,并通过加权 k -means 算法提高聚类质量。以上海市某商场楼宇为例,对比分析该方法与传统 k -means 算法的聚类效果,验证提出方法在用电模式识别方面的准确性更好,能够为楼宇用户的个性化用能管理提供决策依据。

1 待选择聚类特征集合的建立

用电模式分类的结果与选用聚类样本数据密切相关,不同样本数据分析得到的结果差异较大,因此,需要对聚类对象进行有效性评估和优选,以提升聚类质量。

从用电负荷数据的角度出发,日最大负荷、日负荷率、日峰谷差等日负荷类特征能够反映用户的用电行为特点,可以作为聚类对象进行聚类分析^[10]。同时,气象类因素如体感温度、气温、相对湿度等可能会造成负荷的短期变化^[11],在进行聚类分析时应将其影响考虑在内。该文基于商场负荷数据,综合考虑气象类因素的影响,选取了日最大负荷、日最小

负荷、最高气温等 14 个指标建立的待选择聚类特征集合如图 1 所示。

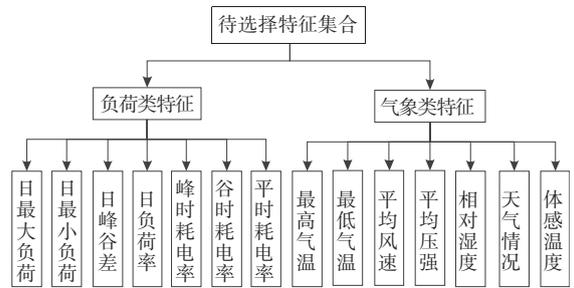


图 1 待选特征集合

Figure 1 Feature set to be selected

各特征的量纲不同会影响聚类结果的准确性,需要对数据进行归一化处理。

1) 日特征向量。原始数据中提取出的全年每日特征值所构成的日特征向量在数值上是连续的,可采用最小-最大归一化的方法将各特征转化为 $[0, 1]$ 区间上的值,即

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

式中 x_{\max} 为样本数据中的最大值, x_{\min} 为样本数据中的最小值。

2) 天气情况。上海地区天气情况主要分为晴、多云、阴、雨等情况,根据文献[12]中定义的区别系数,将晴、多云、阴分别量化为 1、0.9 和 0.8,雨天分为小雨、中雨、大雨和雨夹雪,依次量化为 0.6、0.5、0.4 和 0.3。

2 聚类特征优选方法

考虑到数据信息间的耦合和交差干扰,选取合适的方法对聚类特征进行评价,可以提取出具有代表性的特征并剔除关联性差的特征。采用加权关联度的方法进行聚类特征优选,既从互信息的角度考虑了特征的有效性,又结合灰色关联度反映了特征与负荷之间变化趋势的一致性。

2.1 互信息

信息论中的互信息表示一个系统中所包含另一个系统中信息量的多少^[13],应用在特征选择中表示特征对负荷的影响作用程度,互信息值越大,其影响作用越大。设负荷序列为

$$A = [a_1, a_2, \dots, a_t] \quad (2)$$

特征序列为

$$B_i = [b_1, b_2, \dots, b_l] \quad (3)$$

基于熵估计的互信息为

$$I(A, B_i) = H(A) + H(B_i) - H(A, B_i) \quad (4)$$

式中 $H(A)$ 为负荷的信息熵; $H(B_i)$ 为特征的信息熵; $H(A, B_i)$ 为特征与负荷序列的联合熵。具体计算为

$$\begin{cases} H(A) = - \sum_{a_p \in A} p(a_p) \log p(a_p) \\ H(B_i) = - \sum_{b_q \in B} p(b_q) \log p(b_q) \\ H(A, B_i) = - \sum_{a_p, b_q} p(a_p, b_q) \log p(a_p, b_q) \end{cases} \quad (5)$$

式中 $p(a_p)$ 为负荷的边缘概率分布; $p(b_q)$ 为特征的边缘概率分布; $p(a_p, b_q)$ 为负荷与特征的联合概率分布。

2.2 灰色关联分析

灰色关联分析根据因素之间发展趋势的相似或相异程度判断待识别对象与研究对象的关联程度^[14],灰色关联度数值越大,关联程度越高。

定义负荷序列 A 为参考序列,特征序列 B_i 为比较序列,负荷序列与特征序列的灰色关联度为

$$\gamma_i = \frac{1}{t} \sum_{p=1}^t \zeta_i(p) \quad (6)$$

式中 ζ_i 为比较序列与参考序列的相关系数,即

$$\zeta_i(p) = \frac{\min_i \min_p \Delta_i + \rho \max_i \max_p \Delta_i}{\Delta_i + \rho \max_i \max_p \Delta_i} \quad (7)$$

式中 ρ 为分辨系数; Δ_i 为比较序列与参考序列的绝对差值。

2.3 基于加权关联度的聚类特征优选

采用互信息和灰色关联分析法对聚类特征进行选择,部分特征选择结果如表 1 所示。由表 1 可知,互信息和灰色关联分析法对特征选择的结果存在偏差,需要对选择结果进行综合评估。

表 1 部分特征选择结果

Table 1 Partial feature selection results

标号	特征名称	互信息	灰色关联度
1	日最大负荷	1.107 4	0.673 2
2	日峰谷差	1.022 5	0.684 8

互信息值度量了特征对负荷的影响作用,其权重反映了特征的有效性。灰色关联度量化计算特征与负荷的贴近程度,揭示了特征与负荷之间变化趋势的一致性。将互信息权重与灰色关联度相结合,得到相应特征的加权关联度能够同时反映特征对负荷的影响作用和共有的变化趋势。加权关联度值越大,表示该特征越重要,其计算为

$$J_i = \frac{I(A, B_i)}{\sum_{i=1}^n I(A, B_i)} \gamma_i \quad (8)$$

式中 n 为待选择特征的总数。

基于加权关联度的特征优选步骤如下。

1) 计算出所有待选择特征的加权关联度,按其值大小降序排列。

2) 选择加权关联度值最大的特征,加入至优选聚类特征集合 S 中,并从待选择特征集合中剔除。

3) 从余下待选择特征中取出加权关联度值最大的特征并进行特征优选,即

$$\omega_j = \frac{J_j}{J_s + J_j} < \alpha \quad (9)$$

式中 J_s 为优选聚类特征集合 S 中各特征加权关联度之和; ω_j 为第 j 个优选特征权重,其值小于阈值 α 时,聚类特征选择完成。

3 加权聚类分析方法

3.1 基于权重分配的 k -means 算法

传统的 k -means 算法以欧式距离为聚类相似度量,计算简单方便,但只考虑了每个对象之间距离的相近程度,忽略了内在特征的重要性。基于各优选聚类特征的重要程度,将特征权重分配与欧式距离相结合,则优选聚类特征集合 S 中聚类对象之间的特征加权欧式距离为

$$d(s_i, s_k) = \left[\sum_{j=1}^m \omega'_j (s_{lj} - s_{kj})^2 \right]^{1/2} \quad (10)$$

式中 ω'_j 为由式(9)得到的各优选特征权重,所有 ω'_j 之和为 1; s_i, s_k 为不同的聚类对象。

基于特征加权的 k -means 聚类步骤如下。

- 1) 获取优选特征的权重 ω'_j 。其中, $j=1, 2, \dots, m$ 。
- 2) 对优选聚类特征组成的数据样本集,选取 K 个对象作为初始聚类中心。

3)计算数据集样本到聚类中心的加权距离,按照加权距离最小的原则分配最邻近类别。

4)根据聚类结果重新计算 K 个聚类中心作为新的聚类中心。

5)重复步骤 3、4,直到聚类中心不再变化,或满足迭代次数要求。

3.2 聚类有效性评判指标

Davies-Bouldin 指数是一种适用于评估电力负荷聚类有效性的指标^[15],其结果可作为确定最优聚类划分数目和特征选择有效性验证的依据,其计算为

$$\begin{cases} D_{BI} = \frac{1}{K} \sum_{K=1}^K R_K \\ R_K = \max_{l \neq k} \frac{d(s_l) + d(s_k)}{d(c_l, c_k)} \end{cases} \quad (11)$$

式中 $d(s_l)$ 、 $d(s_k)$ 为矩阵内部距离; c_l 、 c_k 分别为 s_l 和 s_k 的中心; $d(c_l, c_k)$ 为向量间的距离。 D_{BI} 数值越小,表明类内距离越小,类间距离越大,即聚类效果越好。

4 算例分析

该文以上海市某商场建筑 2016 年 1 月 1 日至 2016 年 12 月 31 日的负荷数据及气象数据为分析样本,其中负荷数据包括总负荷数据和分项负荷数据(照明插座、空调、动力和特殊其它),采样频率为 15 min/点,每日记录 96 个数据点。从原始数据中提取对应的日特征向量,建立待选择特征集合,并对该特征集合进行特征优选和聚类分析,具体流程如图 2 所示。

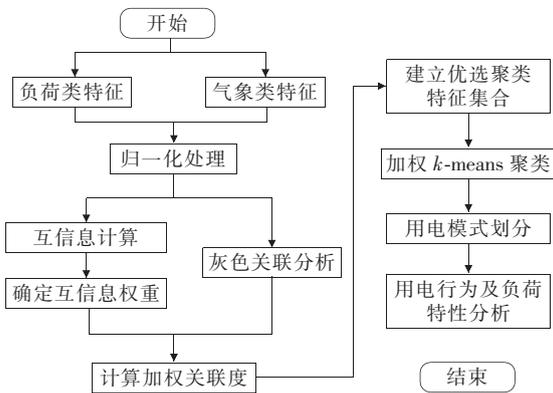


图 2 用电模式分析流程

Figure 2 Flow chart of power consumption mode analysis

4.1 聚类特征优选

以归一化后的日总负荷用电数据为参考序列,分别计算出各特征关于负荷的互信息值和灰色关联度,并通过式(8)得到各特征的加权关联度。对每一个特征单独进行聚类,其聚类效果按 D_{BI} 数值大小升序排列,各特征的加权关联度和聚类效果如表 2 所示。

加权关联度的计算结果与聚类效果基本一致,加权关联度数值越大的特征,其单独聚类的效果也越好,加权关联度能够正确反映样本单一特征的聚类效果。

考虑到所选特征的加权关联度是递减的,取阈值 α 为 0.15,即所选特征权重小于 0.15 时,可以认为余下特征对负荷的影响作用不是很大,此时特征选择完毕^[8]。采用该文提出的特征优选方法得到的优选特征集合 S 由日最大负荷、日峰谷差、体感温度和最高气温共 4 个特征组成。为验证优选集合的有效性,将待选择集合中各特征按其加权关联度由大到小的顺序依次添加到聚类特征集中,聚类结果如表 3 所示。

聚类效果随着特征数的增加先好后坏,聚类特征数为 4 的时候聚类效果最好,与优选结果一致,所提出的特征优选方法是有效的。

表 2 特征集合的评价和聚类结果

Table 2 Evaluation and clustering results of feature set

标号	特征名称	加权关联度	D_{BI}
1	日最大负荷	0.109 6	0.618 7
2	日峰谷差	0.103 0	0.635 6
3	体感温度	0.082 1	0.649 9
4	最高气温	0.072 8	0.676 1
5	最低气温	0.062 7	0.687 1
6	平均压强	0.047 1	0.815 2
7	峰时耗电量	0.043 1	0.846 8
8	日最小负荷	0.043 0	0.881 8
9	平时耗电量	0.019 6	1.422 2
10	谷时耗电量	0.018 2	1.485 6
11	日负荷率	0.013 4	2.099 5
12	相对湿度	0.005 0	5.610 8
13	平均风速	0.003 2	12.177 1
14	天气情况	0.002 9	14.267 2

表 3 不同特征集合对应的聚类结果

Table 3 Clustering results of different feature sets

聚类特征数	D_{BI} 指数	运行计时/s
2	0.593 8	0.239
4	0.571 5	0.251
6	0.631 5	0.292
8	0.637 5	0.309
10	0.635 1	0.339
12	0.631 6	0.341
14	0.635 1	0.343

4.2 用电模式分析

对优选特征集合进行加权聚类,图 3 为不同聚类数目对应的 D_{BI} 指数,当聚类数目为 3 时, D_{BI} 数值最小,即应将实验商场的用电模式划分为 3 类。图 4、5 分别为各用电模式的划分情况及其分项负荷构成。

结合负荷的日期分布及其分项负荷构成对各用电模式进行分析可知,模式 1 主要分布于夏季的 7、8 月,为高峰用电模式,为保证顾客的舒适度,需要空调制冷,该模式下空调负荷和照明与插座为主要负荷,占比均为 42%。模式 2 主要分布于 4、5、6、9、10 月,为次高峰用电模式,照明与插座和空调负荷

为主要负荷,占比分别为 52%和 29%。模式 3 主要分布于 1、2、3、11、12 月,为低谷用电模式,实验商场有别的供暖方式,空调负荷主要发挥通风的作用,负荷较低,照明与插座为主要负荷,占比为 68%。

同一商场不同用电模式的负荷形态相似,在负荷值大小和分项负荷构成上存在差异。以模式 1 为例,分析其用电负荷的变化情况,模式 1 分项负荷及其占比变化情况如图 6、7 所示。

商场负荷可按其营业情况进行分析。07:00—09:00 为商场营业准备时段,空调负荷先投入使用,其占比迅速增大至 70%,其他分项负荷不变,对应

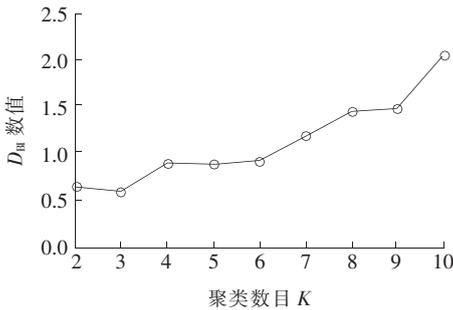


图 3 最优聚类数目选择

Figure 3 Selection of optimal number of clustering

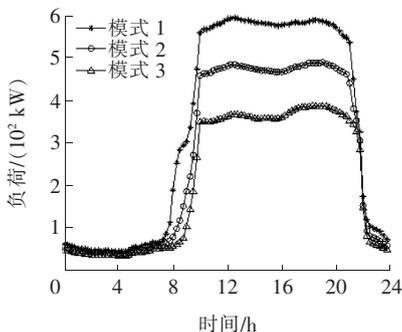


图 4 用电模式划分结果

Figure 4 Results of power mode clustering

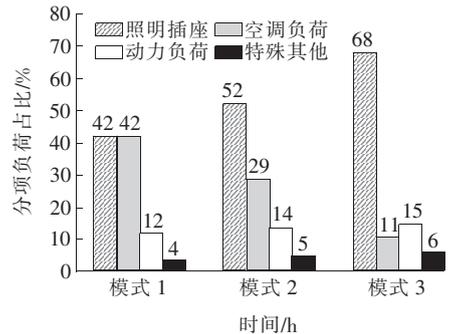


图 5 用电模式的分项负荷比构成

Figure 5 Classified load ratio of the power consumption mode

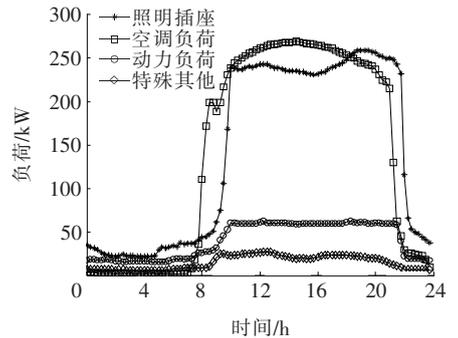


图 6 模式 1 分项负荷变化情况

Figure 6 Component load changes of Mode 1

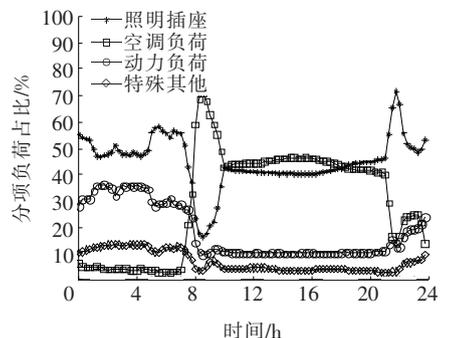


图 7 模式 1 分项负荷占比变化情况

Figure 7 Component load ratio changes of Mode 1

占比降低。09:00—10:00 商场开始营业,各分项负荷均增大,照明和插座占比升至 42%,空调负荷占比下降。10:00—21:00 商场正常营业,各分项负荷波动不大,照明与插座与空调负荷占比维持在 42% 和 47% 左右。21:00—22:00 商场准备打烊,空调负荷占比迅速降低至 17%,其他分项负荷不变,对应占比升高。22:00—06:00 为非营业时间,各分项负荷减小至一个稳定的低水平,照明与插座和动力负荷为该时段的主要用能负荷。

对于模式 1,夏季空调负荷占比高,日间空调用电集中,负荷值较大,适合推行节能技术改造,或调整空调控制策略,如采用温度控制,适当提高中央空调温度设定值等。对于模式 2、3,照明插座为主要用电负荷,应考虑更换节能灯具,减少无人区域照明及插座电源的使用,并减少夜间投入使用的电梯等动力设备。

4.3 实验方法的有效性分析

为检验方法的聚类效果,将传统 k -means 方法直接对商场负荷曲线进行聚类作为对比方法,分别计算出 2 种方法的 D_{BI} 数值并记录程序运行时间,其聚类性能对比结果如表 4 所示。

表 4 实验方法的聚类性能对比

Table 4 Clustering performance comparison of methods

实验方法	D_{BI} 指数	运行计时/s
对比	0.714 0	0.433
该文	0.571 5	0.439

从原始数据中提取特征集合并进行特征优选的程序运行时间为 0.195 s,对优选后的特征集进行加权聚类的程序运行时间为 0.244 s,总计时为 0.439 s,与传统 k -means 方法用时相差不大,但该文方法的 D_{BI} 数值更小,证明了该方法的聚类质量优于传统 k -means 聚类方法。

5 结语

该文利用加权关联度优选聚类特征集合,并通过加权 k -means 算法进行聚类分析,取得了较好的结果。通过该方法提取分析了商场楼宇的 3 种用电模式,可以帮助电力公司更好的了解用户用电特性

以提供个性化服务,如提供节能改造或调整空调运行策略建议等。但该文仅对单个商场一年的数据进行了用电模式分析。在面对更大的数据量时,可以考虑结合降维技术等,进一步改善聚类效果,提高聚类效率。

参考文献:

- [1] 胡江溢,祝恩国,杜新纲,等.用电信息采集系统应用现状及发展趋势[J].电力系统自动化,2014(2):131-135.
HU Jiangyi, ZHU Enguo, DU Xingang, et al. Application status and development trend of electric power information acquisition system[J]. Automation of electric power system, 2014(2):131-135.
- [2] 刘思贤,康军胜,刘素芳.基于用电信息采集系统的高压电能表失压故障监测和损失评估[J].供用电,2019,36(8):49-53.
LIU Sixian, KANG Junsheng, LIU Sufang. Loss-of-voltage fault monitoring and loss assessment of high-voltage energy meter based on electricity information acquisition system [J]. Distribution & Utilization, 2019, 36(8): 49-53.
- [3] 纪永新,王承民,张玉林,等.智能配电网一二次协同规划方法研究[J].智慧电力,2020,48(1):69-73.
JI Yongxin, WANG Chengmin, ZHANG Yulin, et al. Primary and secondary collaborative planning method for smart distribution network[J]. Smart Power, 2020, 48(1):69-73.
- [4] 邬晓韬.基于大数据技术的电网能源智能调配应用研究[J].电网与清洁能源,2019,35(10):23-27.
WU Xiaotao. Research on the application of the intelligent allocation of grid energy based on big data technology[J]. Power System and Clean Energy, 2019, 35(10): 23-27.
- [5] 徐磊,杨秀,张美霞.基于数据挖掘的工业用户用电行为分析[J].电测与仪表,2017,54(16):68-74.
XU Lei, YANG Xiu, ZHANG Meixia. Industrial users of electricity behavior analysis based on data mining[J]. Electrical Measurement & Instrumentation, 2017, 54(16):68-74.
- [6] 王星华,陈卓优,彭显刚.一种基于双层聚类分析的负荷形态组合识别方法[J].电网技术,2016,40(5):1495-1501.
WANG Xinghua, CHEN Zhuoyou, PENG Xiangang.

- A new combinational electrical load analysis method based on bilayer clustering analysis[J]. *Power System Technology*, 2016, 40(5):1495-1501.
- [7] Cheng F, Fu X, Yang Z. A short-term building cooling load prediction method using deep learning algorithms[J]. *Applied Energy*, 2017, 195:222-233.
- [8] 傅军栋, 杨姚, 罗善江. 智能小区居民用电负荷特征权重分析[J]. *电力系统保护与控制*, 2016, 44(18):41-45.
FU Jundong, YANG Yao, LUO Shanjiang. Residential electricity load features weighting analysis in smart community[J]. *Power System Protection and Control*, 2016, 44(18):41-45.
- [9] 陆俊, 朱炎平, 彭文昊, 等. 智能用电用户行为分析特征优选策略[J]. *电力系统自动化*, 2017, 41(5):58-63.
LU Jun, ZHU Yanping, PENG Wenhao, et al. Feature selection strategy for electricity consumption behavior analysis in smart grid[J]. *Automation of Electric Power Systems*, 2017, 41(5):58-63.
- [10] 孙毅, 冯云, 崔灿, 等. 基于动态自适应 K 均值聚类的电力用户负荷编码与行为分析[J]. *电力科学与技术学报*, 2017, 32(3):3-8.
SUN Yi, FENG Yun, CUI Can, et al. Power user load code and behavior analysis based on dynamic adaptive k -means clustering[J]. *Journal of Electric Power Science and Technology*, 2017, 32(3):3-8.
- [11] 刘庆时, 赵贺, 刘迪, 等. 基于支持向量机的家庭负荷控制策略研究[J]. *电力科学与技术学报*, 2016, 31(4):96-101.
LIU Qingshi, ZHAO He, LIU Di, et al. Research on family load control strategy based on support vector machine[J]. *Journal of Electric Power Science and Technology*, 2016, 31(4):96-101.
- [12] 张志, 杜延菱, 崔慧军, 等. 考虑关联因素的智能化中长期电力负荷预测方法[J]. *电力系统保护与控制*, 2019, 47(2):24-30.
ZHANG Zhi, DU Yanling, CUI Huijun, et al. Intelligent mid-long electricity load forecast method considering associated factors[J]. *Power System Protection and Control*, 2019, 47(2):24-30.
- [13] 谢真桢, 杨秀, 张鹏, 等. 基于信息论与混合聚类分析的短期负荷预测方法研究[J]. *电测与仪表*, 2017, 54(19):67-72.
XIE Zhenzhen, YANG Xiu, ZHANG Peng, et al. Study on short-term load forecasting method based on information theory and mixed cluster analysis[J]. *Electrical Measurement & Instrumentation*, 2017, 54(19):67-72.
- [14] 高亚静, 孙永健, 杨文海, 等. 基于新型人体舒适度的气象敏感负荷短期预测研究[J]. *中国电机工程学报*, 2017, 37(7):1946-1954.
GAO Yajing, SUN Yongjian, YANG Wenhai, et al. Weather-sensitive load's short-term forecasting research based on new human body amenity indicator[J]. *Proceedings of the CSEE*, 2017, 37(7):1946-1954.
- [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.