

DOI: 10.7667/PSPC151714

## 智能小区居民用电负荷特征权重分析

傅军栋, 杨姚, 罗善江

(华东交通大学电气与电子工程学院, 江西 南昌 330013)

**摘要:** 以往对智能小区居民用电行为聚类分析时, 存在着负荷特征选择与权重计算描述不足的问题。为了提高居民用电行为聚类分析的准确率, 降低聚类分析运行时间, 提出一种基于 ReliefF 算法建立的以峰时耗电率、日负荷峰值时刻、谷时耗电率、日负荷周期数、日最小负荷率等特征的数据模型。该模型可以对海量居民用电行为数据进行处理, 并通过  $k$ -means 算法对其进行聚类分析。实验数据来源为已建成的智能小区, 结果准确率达 94.61%, 证明了基于 ReliefF 算法建立的特征数据模型在居民用电行为类分析中是有效的。

**关键词:** 用电行为; 聚类分析; 负荷特征; 数据模型

### Residential electricity load features weighting analysis in smart community

FU Jundong, YANG Yao, LUO Shanjiang

(School of Electrical Engineering, East China Jiaotong University, Nanchang 330013, China)

**Abstract:** In order to solve the described insufficient problem of load feature selection and weight calculation in the past clustering analysis of residential electricity behavior, enhance the accuracy of clustering analysis in residential electricity behavior and reduce the time of clustering analysis operation, a data model based on ReliefF algorithm is proposed. The data model is characterized by electricity consumption rate during peak hour, the peak load time, the valley of the power, daily load cycles, the minimum load rate feature, and so on. The massive data of residential electricity behavior can be processed by the model, and clustering analysis of the model is made through  $k$ -means algorithm. Experimental data is obtained from a built-up smart community, and the result accuracy reaches to 94.61%, showing the proposed model based on ReliefF algorithm in clustering analysis of residential electricity behavior is effective.

**Key words:** electricity consumption behavior; clustering analysis; load characteristic; data model

## 0 引言

随着我国智能小区的不断建设和发展, 人们积累了大量居民用电数据<sup>[1]</sup>。虽然用电数据看似十分杂乱, 但实际上却隐藏着居民用电行为方式, 且数据之间还存在一定的关联性。若对散乱的海量数据进行挖掘并研究居民用电行为类型, 将同类居民用电负荷平均分配给 A、B、C 三相, 则可以使低压配电网三相电流接近, 从而降低三相不平衡产生的零序电流对线路、变压器等造成的不利影响<sup>[2]</sup>。同时, 也可以简单、有效地针对不同类型用户制定不同服务<sup>[3]</sup>, 帮助电网削峰填谷<sup>[4]</sup>、节能减排。

聚类分析凭借处理数据在全局范围内分布特征的高效能力, 正逐渐应用在电力用户分类领域<sup>[5]</sup>。文献[6]以电力用户负荷、行业及现行电价为基础, 对用户进行聚类分析。文献[7]根据工业、农业、市

政的用电行业特征进行聚类分析。上述文献以及其他用电行为分类的文献[8-11]均未涉及居民用电行为分类。文献[12]基于聚类算法首次对智能小区用户用电行为进行研究, 直接选择了峰时耗电率、负荷率、谷电系数、平段的用电量百分比等 4 个负荷特征进行聚类分析, 得出 5 类典型用户类型, 但作者没有对各种负荷特征进行分析选择及权重计算, 忽略了不同负荷特征对聚类产生的影响。

智能小区居民用电数据集中负荷特征向量多, 存在冗余特征, 冗余特征会增加特征模型的复杂度, 影响算法的运算效率<sup>[13]</sup>, 如果先通过特征分析及权重计算, 便可以降低特征维度, 从而提高模型的精确度和降低算法运行时间。本文首次基于 ReliefF 算法对居民用电负荷进行特征选择及权重计算, 建立以峰时耗电率、日负荷峰值时刻、谷时耗电率、日负荷周期数<sup>[14]</sup>、日最小负荷率等 5 种特

征为特征的数据模型,并结合  $k$ -means 算法结果的准确率来反应模型的有效性。

本文的主要贡献为:第一,提出在多种负荷特征情况下,对居民用电行为进行聚类分析时需对负荷特征进行分析选择及权重计算;第二,首次将 ReliefF 算法应用到居民用电行为聚类分析中;第三,建立了峰时耗电率、日负荷峰值时刻、谷时耗电率、日负荷周期数、日最小负荷率等 5 种特征为特征的负荷特征数据模型,并对每一种特征进行分类准确率测试。

## 1 负荷特征

智能小区用智能电表及智能插座采集小区居民用电数据<sup>[15]</sup>,数据既可存放在家庭智能网关给家庭用电进行自我调整,又可通过智能网关传送到小区主站供值班人员分析决策。基于用电数据集,提出以下可能影响聚类分析的特征:

- (1) 日最小负荷率,最小负荷/日最大负荷。
- (2) 日峰谷差,日最大负荷-日最小负荷。
- (3) 日峰谷差率,(日最大负荷-日最小负荷)/日最大负荷。
- (4) 日负荷率,用户平均负荷/最大负荷。
- (5) 峰时耗电率,高峰时段用电量/总的用电量。
- (6) 谷时耗电率,低谷时段用电量/总的用电量。
- (7) 平段的用电量百分比,平段用电量/总的用电量。

- (8) 日负荷峰值时刻  $T_{\max}^{[14]}$ 。

$$P_{d,T_{\max}} = P_{d,\max} \quad (1)$$

式中,  $P_{d,\max}$  表征了一天中最大的运行负荷。

- (9) 日负荷周期数  $n_T^{[14]}$ 。

$P > P_{d,\max} - 0.3(P_{d,\max} - P_{d,\min})$  为高负荷  $P_H$ ,  $P < P_{d,\max} + 0.3(P_{d,\max} - P_{d,\min})$  为低负荷  $P_L$ ,一日中负荷在  $P_H$  和  $P_L$  之间变换一次记为 0.5 周期,一日总周期数为  $n_T$ ,日负荷周期数表征了负荷波动性。

## 2 模型建立与模型评价

### 2.1 基于 ReliefF 算法建立模型

本文基于 ReliefF 算法建立负荷特征数据模型。ReliefF 算法是通过 Kira 提出的 Relief 算法改进而来。Relief 算法运行效率高,对数据类型没有要求,对特征间的关系不敏感,但只局限于两类数据的分类问题,而 ReliefF 算法是可以处理多类问题和回归问题,并补充了数据缺失情况下的处理办法,是公认的效果最好的 filter 特征评估算法<sup>[13]</sup>。因此本文选用 ReliefF 算法建立特征数据模型,具体步骤如下。

输入:从实验数据源中随机选取 4 类典型特征

的家庭用户的正常用电数据共 4 800 条作为输入数据集  $D$ ,根据样本数量和特征数设定迭代次数  $m=80$  和最近邻近样本个数  $k=8$ ,特征的类别数  $s=9$ 。

步骤 1 置用电数据集  $D$  所有样本的各个特征权重为 0,即:  $W(a)=0, a=1, 2, \dots, s$ 。

步骤 2 从用电数据集  $D$  中随机选择一个样本  $R$ 。

步骤 3 利用编写的 GetRandSamples 函数从  $D$  中找出与  $R$  同类的  $k=8$  个最邻近  $H_j(j=1, 2, \dots, k)$ ;从  $R$  不同类样本集中找出  $k=8$  个最邻近  $M_j(c)$ 。

步骤 4 根据式(2)更新每一个特征  $a$  的权值  $W(a)$ 。

$$W(a) = W(a) \sum_{j=1}^k d(a, R, H_j) / (mk) + \sum_{c \neq \text{class}(R)} \left[ \frac{p(c)}{1 - p(\text{class}(R))} \sum_{j=1}^k d(a, R, M_j(c)) \right] / (mk) \quad (2)$$

步骤 5 步骤 4 重复  $N=40$  次,求出  $W(a)$  的平均值;

步骤 6 步骤 2~步骤 5 重复  $m=80$  次,输出每个特征的平均权重  $W(a)$ 。

输出:9 个特征的平均权重  $W(a)$ ,降序排列。

其中:  $d(a, R_1, R_2)$  表示样本  $R_1$  和  $R_2$  在特征  $a$  的欧式距离;  $p(c)$  表示类  $c$  目标的概率;  $M_j(c)$  表示第  $c$  类目标的第  $j$  个样本。

### 2.2 模型评价

$k$ -means 算法是最常用且最著名的聚类算法之一,最主要的优点是计算简单、聚类结果稳定,因此本文基于  $k$ -means 聚类分析结果的准确率判断模型的有效性。具体步骤如下所述。

步骤 1 对数据集  $D$  中 4 类典型用户的实际用户类型进行编号: A 类型编号为 1, B 类型编号为 2, C 类型编号为 3, D 类型编号为 4。

步骤 2 从用电数据集  $D$  中选择需要聚类分析的特征向量构成数据集  $X$ ,如单独对特征 2 聚类分析,数据集  $D$  中特征 2 的列向量就构成了数据集  $X$ ,再随机从数据集  $X$  中选择  $k$  个样本点作为初始的聚类中心,由于本文选取的是 4 类典型用户,因此  $k=4$ 。

步骤 3 调用 Matlab 中  $k$ -means 函数聚类分析。数据类为 4 类,距离计算方式“Distance”选择“city”,“Options”选项选择“Opts”,聚类分析得出居民用电行为类型。

步骤 4 将步骤 3 聚类分析得出的居民用电行为类型与步骤 1 居民用电实际类型比较判断,得出对数据集  $X$  的特征进行聚类分析结果的准确率。

步骤 5 输出每次聚类分析结果的准确率,并通过准确率来评价模型是否有效。

### 3 实验设计及结果分析

#### 3.1 数据来源

数据来源于南昌市某智能小区 2013 年 4~6 月份用户每天的用电数据。每个家庭用户每天的数据包括日最小负荷率、日峰谷差、日峰谷差率、日负荷率、峰时耗电率、谷时耗电率、平段用电百分比、日负荷峰值时刻、日负荷周期数等 9 个负荷特征列向量并将列向量数据进行标准归一化处理, 以及 24 个(由于现场原因每 1 h 采集一次)行向量, 以此建立最初用电特征数据维度模型。

建立初始特征数据维度模型, 特征名称及简要说明如表 1。

表 1 特征说明  
Table 1 feature instruction

特征名称	数据范围	特征标号
日最小负荷率	[0,1]	1
日峰谷差	[0,1]	2
日峰谷差率	[0,1]	3
日负荷率	[0,1]	4
峰时耗电率	[0,1]	5
谷时耗电率	[0,1]	6
平段用电百分比	[0,1]	7
日负荷峰值时刻	[0,1]	8
日负荷周期数	[0,1]	9

从实验数据源中随机选取 4 类典型特征<sup>[16]</sup>的家庭用户用电数据作为测试数据源, 4 类典型特征的家庭为: A 类: 空置房用户; B 类: 老人家庭用户; C 类: 上班族家庭用户; D 类: 老人+上班族家庭用户。将 4 类用户的日负荷曲线标准化后的日负荷曲线如图 1。

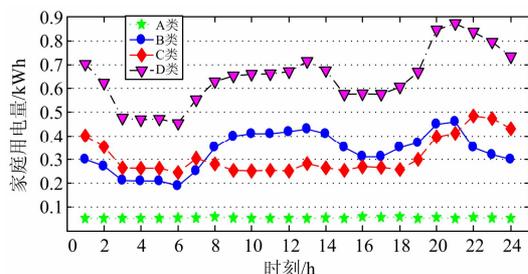


图 1 4 类典型特征家庭用户日负荷曲线

Fig. 1 Four kinds of typical characteristics of home users daily load curve

#### 3.2 实验结果验证与分析

实验一: 基于 ReliefF 算法对初始负荷特征权重计算。

由于算法在运行过程中会选择随机样本  $R$ , 随

机数的不同将导致结果权重有一定的出入, 因此, 本实验采用平均方法, 将主程序运行 40 次, 然后将结果汇总求出每种权重的平均值。9 个特征权重的大小分布如图 2 所示。

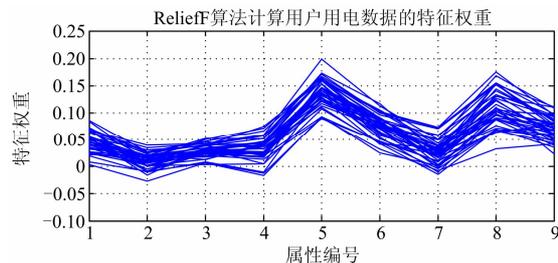


图 2 各个特征权重分布图

Fig. 2 Each feature weight distribution

每个特征权重的平均值从大到小排序如表 2。

表 2 特征权重平均值

Table 2 Feature of average weight

特征标号	平均权重值
5	0.138 095
8	0.109 097
6	0.074 306
9	0.069 983
1	0.045 859
4	0.033 325
3	0.028 216
7	0.026 74
2	0.010 621

由表 2 得: 特征 5 > 特征 8 > 特征 6 > 特征 9 > 特征 1 > 特征 4 > 特征 3 > 特征 7 > 特征 2。

从上面的特征权重可以看出, 特征 5 是最主要的特征, 说明峰时耗电率是影响不同居民用电日负荷曲线相似度的关键特征, 其次是负荷峰值时刻、谷时耗电率、日负荷周期数、日最小负荷率, 以下特征日负荷率、日峰谷差率、平段用电百分比、日峰谷差权重大小接近。

实验二: 基于  $k$ -means 算法测试每一个特征聚类结果的准确率。

聚类  $k$  值选 4, 反复聚类 5 次, 取 5 次聚类结果的平均准确率作为评价  $k$ -means 算法的准确率。分析的结果如表 3。

实验三: 基于 ReliefF 和  $k$ -means 对特征权重分析。

单从分类正确率来看,  $k$ -means 算法已经可以对居民用电行为进行分类做出较高准确的判断, 但考虑 ReliefF 算法对权重的影响及海量数据对  $k$ -means 算法运行时间的影响, 故实验三将结合 ReliefF 算法和  $k$ -means 算法对数据集进行分析。

表 3 *k*-means 聚类结果

Table 3 *k*-means clustering results

用户类型	数据量/条	<i>k</i> -means 准确率
A	1 200	94.7%
B	1 200	94.9%
C	1 200	94.3%
D	1 200	94.4%

(1) 单独对各个特征的数据进行聚类分析, 详细结果如表 4。

表 4 单独对各个特征分析结果

Table 4 Analysis result of each feature

特征标号	总的分类准确率	A 类分类准确率	B 类分类准确率	C 类分类准确率	D 类分类准确率
1	80.56%	82.38%	84.33%	87.94%	84.73%
2	71.04%	73.43%	80.63%	82.69%	81.68%
3	80.41%	81.99%	83.64%	84.74%	82.72%
4	78.36%	82.03%	84.12%	87.41%	84.41%
5	84.80%	87.42%	87.93%	89.78%	88.15%
6	82.90%	86.07%	86.55%	88.45%	86.78%
7	79.83%	80.46%	82.90%	87.05%	84.02%
8	83.78%	86.78%	86.67%	88.88%	87.03%
9	82.17%	84.92%	86.17%	88.29%	86.40%

将表 4 用柱状图 3 的形式显示出来, 更直观。

由图 3 可直观地得出: 各类准确率中, 特征 5 最高, 特征 2 最低, 这与 ReliefF 权重分析的结果一致, 但由于 ReliefF 算法有部分特征权重接近, 所以也区分不明显, 这说明特征权重对分类是有影响的, 并且从表 4 可以看出, 单独从一个特征判定其类型是不可靠的。

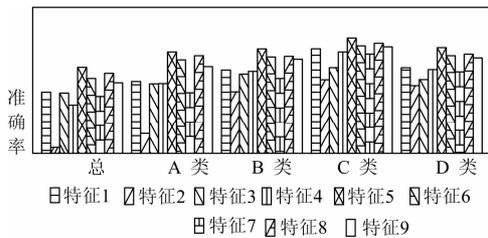


图 3 各个特征分析结果

Fig. 3 Each feature analysis results

(2) 将特征种类按照权重从大到小的顺序排列并结合相应特征聚类分析, 记录每次运行的时间, 结果如表 5。

从表 5 可以看出: 第一, 选择权重排名前 5 的特征聚类分析, 其分析结果准确率几乎达到选择所有特征的准确率; 第二, 选择权重排名前 8 特征聚类分析, 虽然其分析结果准确率比选择权重排名前 5 的高, 但对聚类算法而言, 选择 5 个特征向量时

表 5 ReliefF 和 *k*-means 结合分析结果

Table 5 Result of the analysis of ReliefF and *k*-means

编号	选择特征种类	总分类准确率	运行时间 <i>t</i> /s
1	5、8	91.36%	6.5
2	5、8、6、9	93.23%	7.1
3	5、8、6、9、1	94.61%	8.4
4	5、8、6、9、1、4、3、7	94.82%	11.2
5	全部特征	94.67%	12.6

的运算速率比选择 8 个特征向量运行速率快。当面对海量数据时, 上述优势会更明显。第三: 当选择所有特征时, 准确率却低于选择 8 种特征的准确值, 说明基于数据集的所有特征聚类分析通常得不到最优结果。因此可以得出结论: 居民用电行为聚类分析时, 选择以峰时耗电量、日负荷峰值时刻、谷时耗电量、日负荷周期数、日最小负荷率等 5 个特征建立的最终特征模型更有效。

## 4 结论

本文首先基于 ReliefF 算法对居民用电负荷特征进行特征选择和权重分析, 再结合 *k*-means 算法对居民用电行为进行聚类分析, 同时以峰时耗电量、日负荷峰值时刻、谷时耗电量、日负荷周期数、日最小负荷率为特征建立特征模型。实验结果表明, 模型的建立在居民用电行为聚类分析时是有效的。

## 参考文献

- [1] 林弘宇, 田世明. 智能电网条件下的智能小区关键技术[J]. 电网技术, 2011, 35(12): 1-7.  
LIN Hongyu, TIAN Shiming. Research on key technologies for smart residential community[J]. Power System Technology, 2011, 35(12): 1-7.
- [2] 傅军栋, 喻勇, 黄来福. 不平衡负载的一种更加经济的补偿方法[J]. 电力系统保护与控制, 2015, 43(2): 126-132.  
FU Jundong, YU Yong, HUANG Laifu. A more economical compensation method for unbalanced load[J]. Power System Protection and Control, 2015, 43(2): 126-132.
- [3] 张素香. 智能小区的商业智能[J]. 北京邮电大学学报, 2012, 35(5): 94-97.  
ZHANG Suxiang. Business intelligence in the smart community[J]. Journal of Beijing University of Posts and Telecommunications, 2012, 35(5): 94-97.
- [4] 汤庆峰, 刘念, 张建华. 计及广义需求侧资源的用户侧自动响应机理与关键问题[J]. 电力系统保护与控制, 2014, 42(24): 138-147.  
TANG Qingfeng, LIU Nian, ZHANG Jianhua. Theory and key problems for automated demand response of user

- side considering generalized demand side resources[J]. Power System Protection and Control, 2014, 42(24): 138-147.
- [5] 张东霞, 苗新, 刘丽平, 等. 智能电网大数据技术发展研究[J]. 中国电机工程学报, 2015, 35(1): 2-12.  
ZHANG Dongxia, MIAO Xin, LIU Liping, et al. Research on development strategy for smart grid big data[J]. Proceedings of the CSEE, 2015, 35(1): 2-12.
- [6] 冯晓蒲, 张铁峰. 基于实际负荷曲线的电力用户分类技术研究[J]. 电力科学与工程, 2010, 26(9): 18-22.  
FENG Xiaopu, ZHANG Tiefeng. Research on electricity users classification technology based on actual load curve[J]. Electric Power Science and Engineering, 2010, 26(9): 18-22.
- [7] 李培强, 李欣然, 陈辉华, 等. 基于模糊聚类的电力负荷特性的分类与综合[J]. 中国电机工程学报, 2005, 25(24): 73-78.  
LI Peiqiang, LI Xinran, CHEN Huihua, et al. The characteristics classification and synthesis of power load based on fuzzy clustering[J]. Proceedings of the CSEE, 2005, 25(24): 73-78.
- [8] 黎祚, 周步祥, 林楠. 基于模糊聚类与改进 BP 算法的日负荷特性曲线分类与短期负荷预测[J]. 电力系统保护与控制, 2012, 40(3): 56-60.  
LI Zuo, ZHOU Buxiang, LIN Nan. Classification of daily load characteristics curve and forecasting of short-term load based on fuzzy clustering and improved BP algorithm[J]. Power System Protection and Control, 2012, 40(3): 56-60.
- [9] 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, ZHAI Denghui. Application of k-means clustering algorithm in load curve classification[J]. Power System Protection and Control, 2011, 39(23): 65-68.
- [10] 朱晓清. 电力负荷的分类方法及其应用[D]. 广州: 华南理工大学, 2012.  
ZHU Xiaoqing. Load classification method and its application[D]. Guangzhou: South China University of Technology, 2012.
- [11] 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.
- [12] 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.
- [13] 陈晓琳. 采用 ReliefF 特征加权的 NIC 算法研究[D]. 郑州: 郑州大学, 2014.  
CHEN Xiaolin. A feature weighted NIC algorithm based on ReliefF[D]. Zhengzhou: Zhengzhou University, 2014.
- [14] 孙璞玉, 李家睿, 王承民, 等. 基于负荷特征向量的负荷分类与预测方法研究及其应用[J]. 电气应用, 2013(增刊 1): 234-238.  
SUN Puyu, LI Jiarui, WANG Chengmin, et al. Study and its application of the load classification and prediction method based on load characteristic vector[J]. Electrotechnical Application, 2013(S1): 234-238.
- [15] 谢江宏, 王进平, 上官明霞. 一种智能小区用户用电模式的探索与研究[J]. 华东电力, 2013, 41(1): 157-159.  
XIE Jianghong, WANG Jinping, SHANGGUAN Mingxia. The study of a power consumption model directing at the intelligent community[J]. East China Electric Power, 2013, 41(1): 157-159.

收稿日期: 2015-09-24

作者简介:

傅军栋(1972-), 男, 副教授, 研究生导师, 主要从事电力系统、建筑电气及智能化研究。E-mail: 8755915@qq.com

(编辑 周金梅)