

# 基于优化多类 Adaboost 的非侵入式负荷监测

张顺森,陈铭龙,洪茂雄

(福建工程学院 信息科学与工程学院,福建 福州 350118)

**摘要:** 针对非侵入式负荷监测技术在多状态设备的工作状态辨识研究较少及精度不高的问题,提出了一种基于遗传算法优化的多类 Adaboost 的非侵入式负荷监测技术。首先提取原始数据集有效特征(电流有效值及其变化量、有功功率及其变化量、无功功率)。其次利用遗传算法优化多类 Adaboost 中的五个参数,得到最优强分类器。最后通过第六届“泰迪杯”数据挖掘挑战赛 A 题数据对同时运行两个设备(九阳热水壶、激光打印机)所有状态进行识别。实验结果表明,该算法识别能力优于决策树算法和 SVM 算法。

**关键词:** 非侵入式负荷监测技术;多类 Adaboost;多状态辨识;遗传算法

中图分类号: TM714

文献标志码: A

文章编号: 1672-4348(2019)04-0352-07

## Non-intrusive load monitoring based on optimized multi-class Adaboost

ZHANG Shunmiao, CHEN Minglong, HUNG Maosiung

(School of Information Science and Engineering, Fujian University of Technology, Fuzhou 350118, China)

**Abstract:** Aiming at the lack of research and low accuracy of non-intrusive load monitoring technology in working status identification of multi-state equipment, non-intrusive load monitoring technology based on genetic algorithm optimization for multi-class Adaboost is proposed. Firstly, the effective features (effective value of current and its variation, active power and its variation, reactive power) of the original data set are extracted. Secondly, the genetic algorithm is used to optimize the five parameters of multi-class Adaboost to obtain the optimal strong classifier. Finally, all the states of two devices (Joyoung kettle and FUJI laser printer) running simultaneously are identified by the data set of item A of the 6th “Teddy Cup” Data Mining Race. The experimental results show that the recognition ability of this algorithm is better than that of decision tree algorithm and SVM algorithm.

**Keywords:** non-intrusive load monitoring; multi-class Adaboost; multi-state identification; genetic algorithm

非侵入式负荷监测(NILM)技术<sup>[1]</sup>,是在用户入口处安装传感器,采集终端的总电流、总功率等用电数据,从而获得用户家中每台用电器的用电情况。与传统的侵入式负荷监测(ILM)技术相比,NILM 具有安装方便,经济可靠的优点。目前,在大数据和智能用电背景的推动下,NILM 再次引起了研究人员和专家的广泛关注。NILM 系

统的概念在 1982 年由麻省理工学院的 Hart 教授提出,它旨在为公用事业公司提供一种有效便捷的能耗数据收集方法<sup>[2]</sup>。通过 NILM 技术,居民可以实时观察家中各用电器的用电情况,从而提高居民的节能意识,达到节能减排的目的。电力公司可以通过 NILM 了解每个负荷的特性,从而起到保护电网的作用。因此,研究 NILM 技术具

收稿日期: 2019-06-01

基金项目: 福建省教育厅青年基金项目(JA14217);福建省自然科学基金资助项目(2017J01727)

第一作者简介: 张顺森(1974-),男,福建永泰人,副教授,博士研究生,研究方向:电力大数据,智能计算等。

通信作者: 洪茂雄(1973-),男,台湾高雄人,副教授,博士,研究方向:图像处理,机器学习等。

有重要意义。

国内外学者对 NILM 技术进行了大量的研究。文献[3]使用有功功率,无功功率和功率因数等特征,利用支持向量机和 k-NN( $k=5$ )算法进行分类,比仅使用有功功率特征实现了更精确的设备识别。文献[4]提出了一种利用谐波和功率多重特征的改进遗传算法(GA),其性能明显优于仅使用单功率特征的遗传算法。文献[5]提出了一种基于高阶统计量(HOS)对电器进行分类的 NILM 方法。采用 HOS 算法结合 Fisher 的判别式和遗传算法(GA),对每个负载的 50 个电流暂态信号进行降维,提取具有代表性的特征,适用于实时监测。文献[6]提出了一种负荷分解和决策融合的分类方法,融合了多种算法的结果,提高了设备状态识别的精度。

文献[7]中提出了一种 NILM 方法,该方法每 10 秒采样一次功率变化数据,利用 SVM 算法对电器事件进行检测。该方法对高能耗设备的精度可达 80% 以上。Baets L D 等<sup>[8]</sup>提出了将电流电压轨迹通过加权像素化作为一个新特征,采用卷积神经网络(CNN)来提取分类的关键特征,并在大量的电器分类中取得了良好的效果。

王守相等<sup>[9]</sup>提出一种基于用电模式的居民负荷梯度提升树分类识别方法,采用 PCA 降维和 k-均值算法,从特征的主成分中获得各类负荷用电模式,训练 GBDT 并进行超参数优化。实验结果表明,该模型具有较好的分类性能。文献[10]提出了一种基于 Adaboost 的训练集的筛选方法,并采用 k-NN 和核 Fisher 判别分析(FDA)相结合的算法,提高了对特征相近的电气设备的识别精度。文献[11]提出将基于 Adaboost 改进的 BP 神经网络用于负荷识别,将 BP 神经网络作为弱分类器,负荷识别准确率高。

上述论文主要研究负荷的识别和开关两种状态,缺乏对设备多状态识别的研究。为了解决多状态用电设备下的状态识别问题,本文提出了一种基于遗传算法优化的多类 Adaboost 的非侵入式负荷监测技术,该方法利用电流有效值,有功功率,电流有效值变化量和有功功率变化量、无功功率等特征,通过测试集数据验证了算法的有效性。

## 1 特征提取

一般 NILM 系统框图如图 1 所示,分别为数

据采集、预处理、特征提取和模式识别四步。本文先对原始数据集进行特征提取,然后使用遗传算法优化的多类 Adaboost 执行模式识别分类。

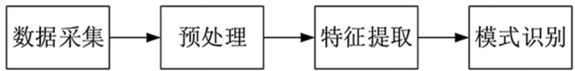


图 1 系统框图  
Fig.1 Basic structure of the system

对原始数据集中单个设备的单一态的电流有效值和有功功率,以及九阳热水壶和 FUJI 激光打印机两个设备共同运行的叠加态的电流有效值、有功功率数据进行分析。

如图 2、图 3 所示,九阳热水壶启动后,有功功率和电流有效值从一个低水平上升到一个高水平,随后保持平稳状态直到关闭。FUJI 激光打印机在工作状态下(扫描、打印、复印)的工作时间短。如图 4 所示,九阳热水壶和 FUJI 激光打印机共同运行时电流的有效值和有功功率呈现出可叠加性。

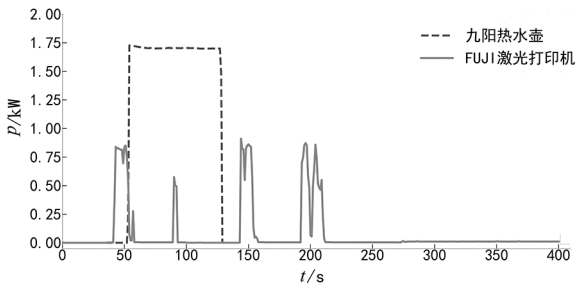


图 2 九阳热水壶和 FUJI 激光打印机有功功率  
Fig.2 Active powers of Joyoung kettle and FUJI laser printer

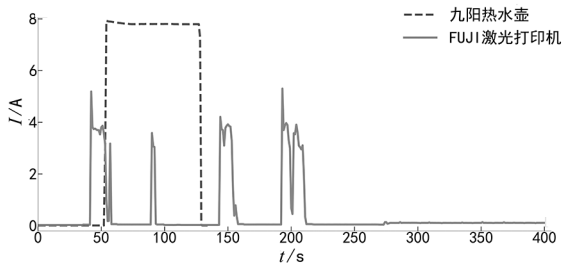


图 3 九阳热水壶和 FUJI 激光打印机电流有效值  
Fig.3 Effective current values of Joyoung kettle and FUJI laser printer

根据叠加原理,通过式(1)表示几种设备的总有功功率:

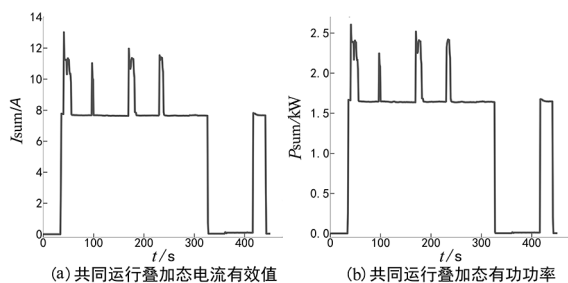


图 4 九阳热水壶和 FUJI 激光打印机叠加态电流有效值(左)和有功率(右)

Fig.4 Effective current values (left) and active powers (right) of the superposition state of Joyoung kettle and FUJI laser printer

$$P(t) = \sum_{i=1}^n P_i(t) \quad (1)$$

式中,  $P_i(t)$  为第  $i$  个设备在  $t$  时刻的有功功率,  $n$  为设备总数。通过计算两个连续有功功率之间的差, 得到其变化量, 并将所有变化量相加, 如式(2)所示:

$$\begin{cases} \Delta P_i(t) = P_i(t) - P_i(t-1) \\ \Delta P(t) = \sum_{i=1}^n \Delta P_i(t) \end{cases} \quad (2)$$

相似地, 对于电流有效值, 总的电流有效值如式(3)所示, 变化量以及变化量之和如式(4)所示:

$$I(t) = \sum_{i=1}^n I_i(t) \quad (3)$$

$$\begin{cases} \Delta I_i(t) = I_i(t) - I_i(t-1) \\ \Delta I(t) = \sum_{i=1}^n \Delta I_i(t) \end{cases} \quad (4)$$

式中,  $I_i(t)$  为第  $i$  个设备在  $t$  时刻的电流有效值。 $\Delta P_i(t)$  和  $\Delta I_i(t)$  为提取设备负载工作状态变化量的特征。结合有功功率、电流有效值、有功功率变化量、电流有效值变化量、无功功率判断此时设备的状态。

## 2 优化多类 Adaboost 算法状态识别

SAMME<sup>[12]</sup> 算法是一种重要的多类 Adaboost 算法, 通过扩展指数损失函数直接把二类的 AdaBoost 算法扩展为多类<sup>[13]</sup>, 并降低对弱分类器分类正确率的要求, 很容易获得足够的分类器, 对多类问题具有很好的泛化效果。通过建立多个 CART 分类树弱分类器, 对各个 CART 分类树弱

分类器进行训练, 根据每次迭代结果, 增加识别用电设备工作状态精确率低的样本的权重, 降低识别用电设备工作状态精确率高的样本的权重。直到满足迭代次数, 退出循环。最后, 将多个 CART 分类树弱分类器进行组合, 得到强分类器, 减少决策树容易发生的过拟合问题, 提高识别精度。

本文利用遗传算法具有搜索全局最优解特点<sup>[14]</sup>, 优化多类 Adaboost 算法参数, 并用 GA-Adaboost 算法对训练集进行 3 折交叉验证的准确率均值作为适应度函数, 选择出最佳参数组合, 提高多类 Adaboost 对用电设备工作状态的识别准确率。

基于遗传算法优化的多类 Adaboost 算法将 CART 作为弱分类器, 利用 GA 对 CART 中的 max\_depth、min\_samples\_split、min\_samples\_leaf 以及 Adaboost 中 n\_estimators、learning\_rate 等参数进行优化, 得到最优强分类器, 最终识别用电设备工作状态。具体实现步骤如下:

输入:  $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$  有效特征训练集, 其中  $x_i$  是由有功功率  $P$ 、电流有效值  $I$ 、有功功率变化量  $\Delta P$ 、电流有效值变化量  $\Delta I$ 、无功功率  $Q$  组成的任一向量,  $y_i$  为标签状态编号, 其中  $y_i \in Y = \{1, 2, 3, \dots, k\}$ ,  $k$  为状态类别数。

输出: 各用电设备工作状态。

步骤 1: 通过式(5)初始化训练集  $S$  的权重  $w_i^{(0)}$ ,  $i = 1, 2, \dots, m$ ,  $m$  为最大样本数

$$w_i^{(0)} = \frac{1}{m} \quad (5)$$

步骤 2: 利用遗传算法, 寻找最优参数。

步骤 3: 根据训练集和样本权重, 训练 CART 分类树弱分类器  $h_t(x)$ ,  $t = 1, 2, \dots, n$ ,  $n$  为最大迭代次数。

通过式(6)计算分类误差率:

$$\varepsilon_t = \sum_{i=1}^m w_i^{(t)} I(y_i \neq h_t(x_i)) \quad (6)$$

其中  $I(\cdot)$  是一个指示符函数, 如果条件为真, 则返回 1, 否则返回 0。

通过式(7)计算 CART 分类树弱分类器  $h_t(x)$  系数:

$$\alpha_t = \frac{1}{2} \log \left( \frac{1 - \varepsilon_t}{\varepsilon_t} \right) + \log(k - 1) \quad (7)$$

通过式(8)更新下一次迭代的样本权重:

$$w_i^{(t+1)} = \frac{w_i^{(t)} \exp(-\alpha_t y_i h_t(x_i))}{\sum_{i=1}^m w_i^{(t)}} \quad (8)$$

判断是否满足迭代次数,若是,执行步骤 4,否则,执行步骤 3。

步骤 4:输出用电设备工作状态

通过式(9)得到的强分类器表示为:

$$G(X) = \operatorname{argmax}_{y \in \{1,2,\cdots,k\}} \left( \sum_{t=1}^n \alpha_t I(h_t(x) = y) \right) \quad (9)$$

将含有  $P$ 、 $I$ 、 $\Delta P$ 、 $\Delta I$ 、 $Q$  等特征的待测样本  $x$ ,利用训练好的强分类器  $G(X)$  进行状态识别。

输出用电设备工作状态。

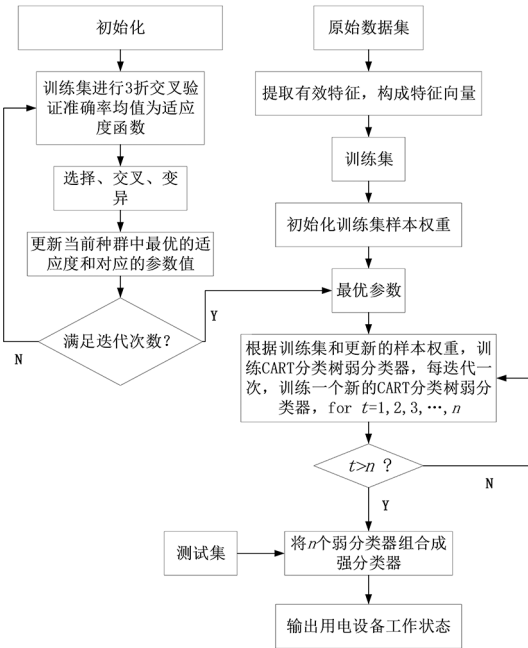


图 5 基于遗传算法优化的多类 Adaboost 的 NILM 流程  
Fig.5 Flow chart of non-intrusive load monitoring based on multi-class Adaboost optimized by GA

3 实验

实验采用的是 Windows 7 64 位操作系统的台式计算机,主频为 3.30 GHz,8G 内存,结合 python3 试验平台,以第六届泰迪杯 A 题数据为例,选择两种常见用电设备,九阳热水壶和 FUJI 激光打印机作为我们的测试目标,其中九阳热水壶有 2 个状态(关闭,运行),FUJI 激光打印机有 5 个状态(关闭,待机,打印,复印,扫描),采用九阳热水壶,FUJI 激光打印机叠加态数据样本为本文数

据集。

表 1 用电设备工作情况

Tab.1 Working conditions of electrical equipments

设备类型	工作状态	标签
九阳热水壶	关闭	0
	运行	1
FUJI 激光打印机	关闭	0
	待机(打开)	1
	打印	2
	待机(结束)	1
	复印	3
	扫描	4
	关闭/关闭	0
两设备共同运行	关闭/待机	1
	关闭/打印	2
	关闭/复印	3
	关闭/扫描	4
	运行/关闭	5
	运行/待机	6
	运行/打印	7
	运行/复印	8
	运行/扫描	9

3.1 评价指标

为了评估本文方法,使用了四个常见的评估指标,包括准确率 (acc)、精确率 (  $P$  )、召回率 (  $R$  )和 F-Score 值(  $F1$  )。其中,准确率提供实验结果的全局信息,F-Score 值综合了精确率和召回率信息。采用下列公式计算四个评价指标。

$$\text{acc} = \frac{1}{N} \sum_{i=0}^{N-1} I(e_i = z_i) \quad (10)$$

在式(10)中,  $e_i$  和  $z_i$  分别为第  $i$  个样本的预测值和真实值,  $I(.)$  是一个指示函数,如果  $e_i = z_i$  返回 1,否则返回 0。

$$p = \frac{\#(\text{TP})}{\#(\text{TP}) + \#(\text{FP})} \times 100\% \quad (11)$$

在式(11)中,TP(true positive)表示样本被标记为  $c_i$  并被预测为  $c_i$ ,而 FP(false positive)表示样本被标记为  $c_j$  但被错误地预测为  $c_i$ 。 $c_i$  和  $c_j$  为两个类,  $c_i \neq c_j$ 。 $\#(.)$ 返回全部预测结果。

$$R = \frac{\#(TP)}{\#(TP) + \#(FN)} \times 100\% \quad (12)$$

在式(12)中,FN(false negative)表示样本被标记为 $c_i$ 但被错误地预测为 $c_j$ 。

$$F1 = \frac{2 \times P \times R}{P + R} \quad (13)$$

3.2 实验结论与分析

本文应用3折交叉验证,对模型进行训练,并对超参数进行优化,最后用测试集测试。从数据样本中随机选择70%和80%作为训练集样本,剩余的30%和20%分别为其测试集样本,分别定义为样本1、样本2。将GA-Adaboost算法与DT和SVM算法进行比较(如表2)。针对样本1、样本2,GA-Adaboost算法对识别九阳热水壶和FUJI激光打印机共同运行状态的整体识别率高于决策树和SVM。针对样本1、样本2的训练时间,决策树算法训练时间最短,SVM次之,GA-Adaboost算法训练时间较长,该算法超参数较多,因此寻找最优超参数增加了训练时间;从测试时间上看,GA-Adaboost算法测试时间仍多于决策树和SVM算

法。本文方法的训练时间和测试时间较决策树和SVM算法长,但其准确率比其余方法高。

利用3种算法对九阳热水壶和FUJI激光打印机共同运行状态的精确率、召回率和F1值进行测试(如表3、表4)。最后给出了两种不同样本比例下设备状态分类的混淆矩阵图(如图6、图7)。

如表3和图6所示,针对样本1,对于识别九阳电热水壶与FUJI激光打印机共同运行的七种工作状态(数据集中仅含0、1、4、6、7、8、9状态),GA-Adaboost对于7状态的识别精确率较低,P值为50%,对8状态的召回率低,对其余状态的识别能力较好,决策树算法未能识别出8和9状态,SVM算法未能识别出8状态。

如表4和图7所示,针对样本2,相对于GA-Adaboost算法,决策树未能识别9状态,SVM算法未能识别出7和8状态。

在样本1以及样本2中,基于遗传算法优化的Adaboost算法分类效果最好,SVM算法在3种算法中效果较差。

表 2 不同划分比例样本下的 DT、SVM 和 GA-Adaboost 算法的状态分类结果

Tab.2 State classification results of DT, SVM and GA-Adaboost algorithms under different scaled samples

算法	样本 1			样本 2		
	准确率/%	训练时间/s	测试时间/s	准确率/%	训练时间/s	测试时间/s
DT	93.33	25.757	0.003 6	94.44	26.71 2	0.003 0
SVM	91.85	83.246	0.004 2	90.00	98.55 8	0.004 6
GA-Adaboost	94.81	134	0.056 0	95.56	278	0.069 2

表 3 样本 1 下的 DT、SVM 和 GA-Adaboost 算法的状态分类精确率、召回率、F1 值

Tab.3 Precision, recall rate and F1 value of state classification by DT, SVM and GA-Adaboost algorithms under Sample 1

工作状态	DT			SVM			GA-Adaboost			support
	P	R	F1	P	R	F1	P	R	F1	
0	100	100	100	100	100	100	100	100	100	14
1	100	94	97	93	88	90	100	88	93	16
4	100	100	100	100	94	97	100	100	100	17
6	96	97	97	91	97	94	95	97	96	79
7	33	100	50	50	33	40	50	100	67	3
8	0	0	0	0	0	0	100	25	40	4
9	0	0	0	67	100	80	100	100	100	2

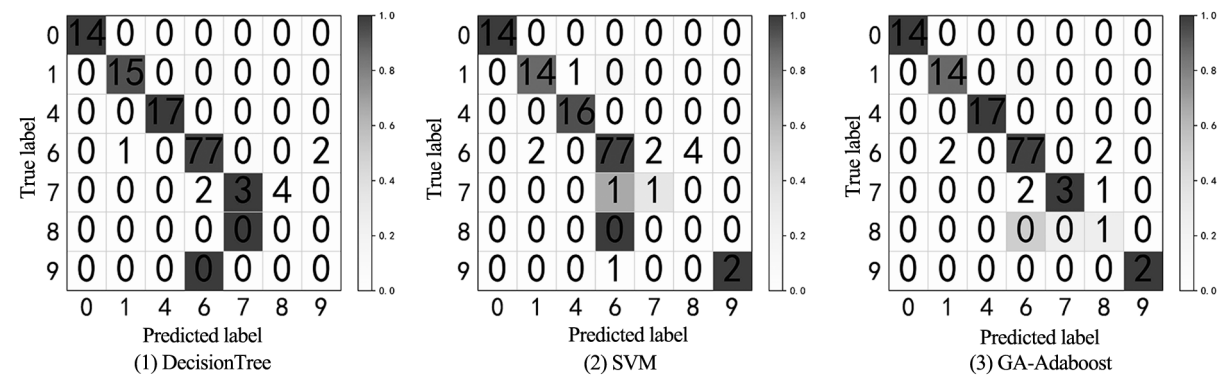


图 6 在样本 1 下,基于 DecisionTree(左)、SVM(中)、GA-Adaboost(右)对九阳热水壶和 FUJI 激光打印机共同运行的混淆矩阵图

Fig.6 Confusion matrix diagram of common operation between Jiuyang Kettle and FUJI laser printer based on DecisionTree (left) , SVM (middle) and GA-Adaboost (right) under Sample 1

表 4 样本 2 下的 DT、SVM 和 GA 优化的多类 Adaboost 算法的状态分类精确率、召回率、F1 值

Tab.4 Precision, recall rate and F1 value of state classification by DT, SVM and GA-Adaboost algorithms under Sample 2

工作状态	DT			SVM			GA-Adaboost			support
	P	R	F1	P	R	F1	P	R	F1	
0	100	100	100	100	100	100	100	100	100	6
1	100	92	96	91	83	87	100	92	96	12
4	100	100	100	100	91	95	100	100	100	11
6	95	98	96	87	100	93	93	100	96	53
7	75	100	86	0	0	0	100	67	80	3
8	67	67	67	0	0	0	100	33	50	3
9	0	0	0	100	100	100	100	100	100	2

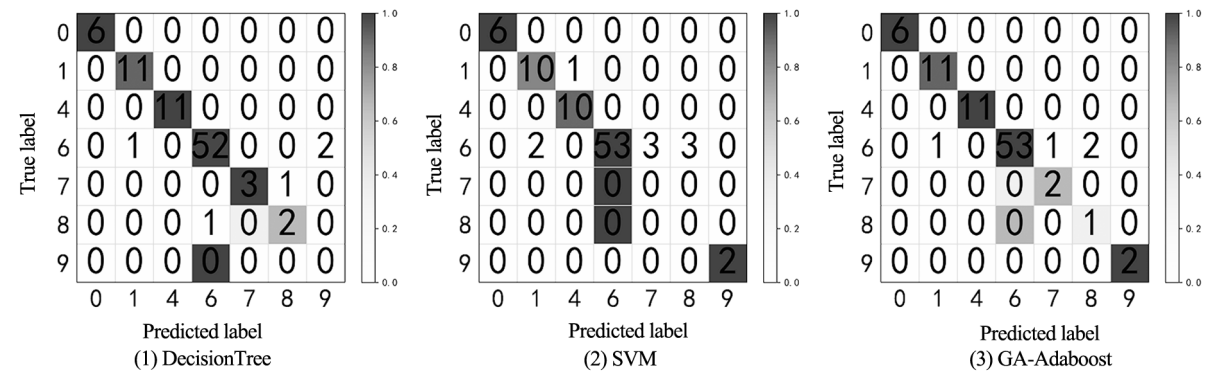


图 7 在样本 2 下,基于 DecisionTree(左)、SVM(中)、GA-Adaboost(右)对九阳热水壶和 FUJI 激光打印机共同运行的混淆矩阵图

Fig.7 Confusion matrix diagram of common operation between Jiuyang kettle and FUJI laser printer based on DecisionTree (left) , SVM (middle) and GA-Adaboost (right) under Sample 2

## 4 结论

提出一种基于遗传算法优化的多类 Adaboost 算法的非侵入式负荷监测,采用电流有效值、有功功率、有功功率变化量、电流有效值变化量、无功功率等多特征,实验结果表明该方法对于多

状态设备的状态辨识,较决策树算法和 SVM 算法精度高。但是实际生活中,居民家庭中不仅仅只有两种用电器共同运行,因此对于多种用电设备共同运行的状态辨识,特别是多状态设备的状态辨识,仍然需要进一步研究。

## 参考文献:

- [1] 余贻鑫, 刘博, 栾文鹏. 非侵入式居民电力负荷监测与分解技术[J]. 南方电网技术, 2013, 7(4): 1-5.
- [2] HART G W. Nonintrusive appliance load monitoring[J]. Proceedings of the IEEE, 1992, 80(12): 1870-1891.
- [3] FIGUEIREDO M, DE ALMEIDA A, RIBEIRO B. Home electrical signal disaggregation for non-intrusive load monitoring (NILM) systems[J]. Neurocomputing, 2012, 96(3): 66-73.
- [4] 孙毅, 崔灿, 陆俊, 等. 基于遗传优化的非侵入式家居负荷分解方法[J]. 电网技术, 2016, 40(12): 3912-3917.
- [5] GUEDES J D S, FERREIRA D D, BARBOSA B H G. A non-intrusive approach to classify electrical appliances based on higher-order statistics and genetic algorithm: a smart grid perspective[J]. Electric Power Systems Research, 2016, 140: 65-69.
- [6] 王晓换, 李如意, 周东国, 等. 基于决策融合的非侵入式电力负荷辨识方法及应用[J]. 电力系统保护与控制, 2016, 44(7): 115-121.
- [7] BASU K, DEBUSSCHERE V, BACHA S, et al. A generic data driven approach for low sampling load disaggregation[J]. Sustainable Energy, Grids and Networks, 2017, 9: 118-127.
- [8] DE BAETS L, RUYSSINCK J, DEVELDER C, et al. Appliance classification using VI trajectories and convolutional neural networks[J]. Energy and Buildings, 2018, 158: 32-36.
- [9] 王守相, 刘天宇. 计及用电模式的居民负荷梯度提升树分类识别方法[J]. 电力系统及其自动化学报, 2017, 29(9): 27-33.
- [10] 宋旭帆, 周明, 涂京, 等. 基于 k-NN 结合核 Fisher 判别的非侵入式负荷监测方法[J]. 电力系统自动化, 2018, 42(6): 73-80.
- [11] 周明, 宋旭帆, 涂京, 等. 基于非侵入式负荷监测的居民用电行为分析[J]. 电网技术, 2018, 42(10): 3268-3276.
- [12] HASTIE T, ROSSET S, ZHU J, et al. Multi-class AdaBoost[J]. Statistics and Its Interface, 2009, 2(3): 349-360.
- [13] 翟夕阳, 王晓丹, 雷蕾, 等. 基于多类指数损失函数逐步添加模型的改进多分类 AdaBoost 算法[J]. 计算机应用, 2017, 37(6): 1692-1696.
- [14] 黄新波, 李文君子, 宋桐, 等. 采用遗传算法优化装袋分类回归树组合算法的变压器故障诊断[J]. 高电压技术, 2016, 42(5): 1617-1623.

(责任编辑: 方素华)