

基于监督学习的非侵入式负荷监测算法比较

涂京, 周明, 宋旭帆, 周光东, 李庚银

(华北电力大学 新能源电力系统国家重点实验室, 北京 102206)

摘要:非侵入式负荷监测(NILM)能够在不干扰用户正常用电的情况下,低成本地实现用户用电设备类型的识别和用电负荷的分解,因此非常适用于家庭用户用电监测。大量智能电表在家庭用户中的安装为居民NILM提供了数据支撑,也使得居民NILM研究成为热点。基于家庭负荷稳态电流样本,采用负荷电流谐波系数作为负荷分类特征,建立了基于多层感知器(MLP)神经网络、 k -近邻算法、逻辑回归、支持向量机的4种NILM分类模型,利用BLUED数据库对4种分类器进行训练和测试,对比分析其在识别精度、训练时间、识别速度和抗噪能力方面的表现,并对其在家庭负荷识别中的应用效果进行对比研究。结果表明,4种分类器中MLP神经网络具有总体最优的分类效果和计算性能,更适用于家庭用户负荷监测。

关键词:非侵入式负荷监测;负荷识别;多层感知器神经网络;支持向量机;监督学习

中图分类号:TM 721;TM 764

文献标识码:A

DOI:10.16081/j.issn.1006-6047.2018.12.019

0 引言

智能电网发展带来的优势之一是电力用户与电网的友好互动^[1],即用户在价格或激励机制下主动调整自身用电行为,实现削峰填谷、节能降耗,同时降低用户用电开支。用户负荷监测是了解用户用电特性的主要方法,为分析用户用电行为、调整用电方式提供依据和指导,是用户参与电网双向互动的支撑技术。因此,研究有效的负荷监测方法对了解用户用电特性、开展智能用电、制定价格或激励政策具有重要的指导意义。

Hart教授于1982年提出了非侵入式负荷监测NILM(Non-Intrusive Load Monitoring)方法^[2],通过监测用户端口处的电气信息,采用数学算法根据用户总体用电功率分解得到各个用电设备的工作状态,进而分析各个设备的用电情况。相比于侵入式监测,NILM的成本低、安装维护简单,同时具有较高的准确度,实用性强,适用于家庭负荷监测等低成本监测场景^[3],具有广泛的应用前景。

家庭智能电表的普及使得NILM成为国内外智能用电研究的热点之一。总体而言,NILM算法可以归纳为监督学习算法和无监督学习算法。监督学习算法的识别准确度较高,应用较广,但用于模型训练的数据集需要人工预先标记。无监督学习算法可以使用无标记数据,但识别准确度较低,目前在事件检测和特征提取阶段有较多应用。在监督学习算法的研究方面,文献[4]应用了动态模糊C均值聚类 and k -近邻 k -NN(k -Nearest Neighbor)识别算法对家庭负荷进行识别,能够对低功率设备实现90%的识别准确

率,效果较好;文献[5]采用主成分分析法和Fisher有监督判别算法实现负荷识别,对运行机理不同的负荷分类效果较好,对运行机理相同的负荷进行区分则需要进一步探讨;文献[6]采用粒子群优化算法优化人工神经网络的训练参数,将构建的神经网络应用于家庭负荷识别,测试结果表明该方法显著提高了识别准确性和计算效率;文献[7]利用弹性反向传播算法训练多层神经网络,通过训练所有可能的设备组合实现负荷分解,但随着设备种类增加,输出神经元和训练样本将呈指数增长;文献[8]采用支持向量机(SVM)对1 Hz的采样信号进行负荷识别,识别了95%的电热负荷,但对小功率电器的识别效果不佳;文献[9]增加3次谐波作为负荷特征,采用遗传算法求解多特征目标函数,分解精度较高,但是遗传算法的求解速度较慢,难以适用于在线分析监测;文献[10]则提出一种基于改进鸡群算法的负荷监测方法,同样存在计算速度较慢的不足。以上文献基于监督学习算法实现了NILM,识别准确度较高,但需要预先获取负荷样本以训练模型。在无监督学习算法的研究方面,文献[11]提出了基于滑动窗双边累积和的无监督负荷投切事件检测算法,采用IEEE 30节点系统进行了仿真测试,表明所提算法能有效检测负荷投切事件,但并未对家庭负荷的投切进行单独测试;文献[12]利用均值漂移聚类算法和权重聚类算法实现了负荷特征提取,可用于建立NILM负荷特征数据库或负荷状态监测。总体而言,无监督学习算法因使用无标记数据存在识别精度不高的问题,因此通常应用于负荷投切事件检测、负荷特征提取等环节。本文着重对基于监督学习的NILM算法进行对比研究。

虽然目前各文献提出了多种家庭NILM的监督学习算法,但所提出的算法通常基于不同的实验环境,且部分文献仅仅分析了算法的准确性,对算法的

收稿日期:2018-05-28;**修回日期:**2018-09-12

基金项目:国家重点研发计划项目(2016YFB0901100)

Project supported by the National Key Research and Development Program of China(2016YFB0901100)

快速性缺乏研究,彼此之间难以进行横向对比。针对此不足,本文建立了 4 种具有代表性的监督学习分类算法,并对它们进行对比研究。首先,在负荷特征选取上,选用容易获取的智能电表提供的稳态电流作为特征提取源,采用离散傅里叶变换 DFT (Discrete Fourier Transform) 对家庭负荷稳态电流进行谐波分解,提取电流的奇次谐波作为负荷特征;其次,分别建立了多层感知器 MLP (Multi-Layer Perceptions) 神经网络、 k -NN、逻辑回归 LR (Logistic Regression)、SVM 这 4 种分类模型;然后,利用卡内基梅隆大学提供的 BLUED 数据库^[13]对 4 种分类器进行了训练和测试,对比分析了它们在识别精度、训练时间、识别速度、抗噪性能等方面的表现,并对其在 NILM 中的应用效果进行了总结。

1 基于 DFT 的负荷电流奇次谐波特征提取

负荷特征作为负荷识别的主要依据,其对不同类别的设备应具有足够的区分能力。早期麻省理工学院利用设备投切前/后的有功变化量 ΔP 、无功变化量 ΔQ 作为负荷特征^[14],具有获取简单、设备投资小的优点,但对功率相近的设备则会发生误判。为了提高识别准确率,在增大采样频率的基础上,电流谐波可作为负荷分类的有效特征;通常采用傅里叶变换提取谐波系数。由于谐波特征能够反映设备的运行特性,而且维数较高,因此具有较好的负荷区分能力。电流谐波特征还可利用其暂态过程谐波分析提取特征。另外,基于电压-电流轨迹图或电压噪声的特征提取方法亦有少量应用,但电压噪声特征由于需要极高的采样频率而难以推广。

本文采用易获取的负荷电流作为特征提取源,利用傅里叶变换进行负荷电流谐波分解,取谐波幅值较大的谐波次数作为负荷特征,以便获取较好的区分能力。对离散采样的负荷电流样本应用 DFT 获取负荷特征的方法如式(1)所示。

$$X(f) = \sum_{n=0}^{N_f-1} i(n) e^{-j \frac{2\pi f n}{N_f}} \quad (1)$$

其中, $i(n)$ 为家用电器周期电流采样点,每周周期采样 N_f 次,从 0 标记到 N_f-1 ; f 为谐波次数, $f=0, 1, \dots, N_f-1$; $X(f)$ 为复数形式的 DFT 的谐波系数。

将 $X(f)$ 前乘以归一化系数 $1/N_f$ 以符合傅里叶级数分析的习惯,利用欧拉公式将其整理为实数计算表达式:

$$X'(f) = \frac{1}{N_f} X(f) = \frac{1}{N_f} \sum_{n=0}^{N_f-1} i(n) \left(\cos \frac{2\pi f n}{N_f} - j \sin \frac{2\pi f n}{N_f} \right) \quad (2)$$

$$a_f = \text{Re}[X'(f)] = \frac{1}{N_f} \sum_{n=0}^{N_f-1} i(n) \cos \frac{2\pi f n}{N_f} \quad (3)$$

$$b_f = \text{Im}[X'(f)] = -\frac{1}{N_f} \sum_{n=0}^{N_f-1} i(n) \sin \frac{2\pi f n}{N_f} \quad (4)$$

其中, a_f 为 $X'(f)$ 的实部, b_f 为 $X'(f)$ 的虚部,两者共同作为 f 次谐波的参数。

对于 DFT 而言,电流谐波的幅值随着谐波次数的增大而减小。因此,为了减少噪声对分类的影响,应选择幅值较大的低次谐波作为主要分类特征;另外,可以证明,如果电流波形满足式(5),即电压正半周下电流波形与电压负半周下电流波形关于电流轴 $y=0$ 对称,则偶次谐波(包含直流分量)为 0。

$$i\left(\frac{N_f}{2}+n\right) = -i(n) \quad n=0, 1, \dots, \frac{N_f}{2}-1 \quad (5)$$

通常家用电器的电流波形均近似满足上述要求。以电冰箱为例,图 1 和图 2 分别为电冰箱在某工作状态下的周期电流波形及其直流分量、基波和 2~13 次谐波的幅值。

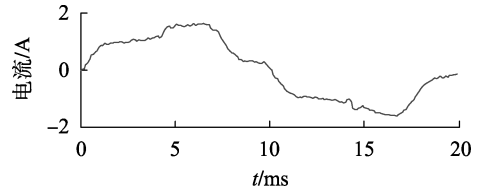


图 1 电冰箱在某工作状态下的周期电流波形

Fig.1 Periodic current waveform of a refrigerator under some working condition

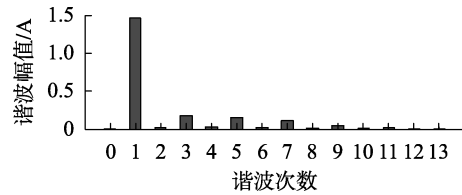


图 2 电流谐波幅值

Fig.2 Amplitude of current harmonic

从图 2 可以看出,偶次谐波幅值明显小于奇次谐波幅值,且 13 次谐波的幅值已经很小,更高次谐波的幅值更小且易受噪声干扰。因此,本文选择基波和 3、5、7、9、11、13 次谐波作为负荷分类的特征。

2 4 种监督学习分类算法模型

2.1 MLP 神经网络分类模型

MLP 神经网络是将单一感知器模型进行多层叠加形成的通用神经网络模型^[15],由一个输入层、若干隐含层和一个输出层组成。MLP 神经网络被证明具有任意逼近连续函数的性能^[16],作为分类算法可以划分复杂的非线性决策边界。

利用 MLP 神经网络进行负荷识别,其输入节点数目与样本负荷特征维数一致,其输出节点数目与所需分类类别数目一致,所需分类类别由标记类别

的训练样本预先确定。网络隐含层层数及其所含神经元的数目则由网络功能的复杂程度确定,用于负荷识别的神经网络通常采用单隐层结构,包含 4~60 个神经元。网络的激励函数采用 Sigmoid 函数,代价函数采用正则化的交叉熵代价函数,训练算法采用反向传播法计算网络参数梯度,具体实现过程见附录 A1。

2.2 k -NN 分类模型

k -NN 算法是一种用于分类的非参数统计方法,通过计算分类对象与其他训练样本的距离,选出最近的 k 个训练样本,其中样本数量最多的类别则作为该分类对象的类别,即分类对象的类别由其最近的 k 个邻居中大多数邻居的类别所决定。

k -NN 算法的逻辑简单、清晰,能够形成复杂的分类边界,但对局部训练样本的构成非常敏感,且参数 k 需要预先选定, k 的大小在一定程度上影响分类算法的效果。另外, k -NN 算法的“多数表决”机制会在训练样本类别分布偏斜时出现缺陷,样本数量较多的类别将大幅影响分类结果,通常利用距离 $1/d$ (d 为待识别对象到邻居的距离)进行加权,赋予近处的邻居较大的权重以处理样本分布偏斜。本文采用 $1/d$ 加权的 k -NN 算法作为家用电器负荷识别的分类模型,并利用遍历法进行样本识别,具体的计算实现过程见附录 A2。

2.3 LR 分类模型

LR 作为一种经典分类算法被广泛应用于各类分类问题,其采用 Sigmoid 函数嵌套线性回归的形式处理二分类问题。LR 在特征空间中应用线性超平面将样本划分为正类和负类,并使得样本分类错误尽可能小。

LR 为二分类的分类器,利用“一对多”^[17]策略将其扩展为多类别分类器。“一对多”策略对每个负荷类别单独建立一个分类器,该分类器以某一负荷类别为正类、其他负荷类别为负类进行训练,对于 K 个负荷类别则训练 K 个分类器。对于某个待分类的负荷样本,将其同时输入所有分类器进行计算,选择输出最大的分类器的负荷类别作为该样本的类别。另外,采用正则化的对数似然函数作为分类器的代价函数,训练算法采用共轭梯度法。正则化的对数似然函数如式(6)所示。

$$C = -\frac{1}{N} \sum_{i=1}^N [y_i \ln y'_i + (1 - y_i) \ln(1 - y'_i)] + \frac{\lambda}{2N} \sum_{j=1}^M \theta_j^2 \quad (6)$$

其中, N 为训练样本数量; y_i 、 y'_i 分别为第 i 个训练样本的期望输出、实际输出; M 为参数数量; θ_j 为第 j 个参数; λ 为正则化项的权重系数。LR 的具体实现过程见附录 A3。

2.4 SVM 分类模型

SVM 是一种基于监督学习的大间距分类器,其核心理念为在特征空间中寻找尽可能大的分类间隔以有效分隔正类和负类。分类间隔通常由正、负类边界若干向量支撑形成,因而称之为 SVM。

SVM 作为二分类的分类器,其对于多分类问题同样采用“一对多”策略,对每个负荷类别单独训练分类器。SVM 应用核函数可以适用于非线性分类问题,本文采用可生成非线性分类边界的高斯核函数 SVM 分类器,其代价函数为软间隔最大化代价函数,具体实现过程参见附录 A4。

2.5 NILM 监督学习算法的理论特点分析

4 种监督学习算法中,MLP 神经网络、 k -NN 算法和基于核函数的 SVM 是非线性负荷分类器,能够根据负荷特征进行复杂的分类边界划分,理论上具有较高的识别准确率;而 LR 为线性分类器,仅能在高维负荷特征空间中使用线性超平面进行负荷分类,对非线性的特征边界则易出现错误。考虑到不同类别负荷亦可能具有相近的负荷特征,因此,理论上具有非线性分类能力的 MLP 神经网络、 k -NN 算法和 SVM 应当具有更好的负荷识别准确率。

从模型建立复杂度的角度来看,MLP 神经网络的模型结构最为复杂,参数数量最多,因此训练模型所需计算时间最长,收敛速度较慢,其训练时间视规模不同需要数分钟至数小时不等。相较而言,LR 和 SVM 由于结构简单、参数数量较少,训练速度较快,通常在秒级时间即可完成训练。 k -NN 算法作为一种基于实例的监督学习算法,无需事先建立模型,没有训练过程。

从模型应用的角度来看, k -NN 算法需要将待识别的负荷样本与所有已知的负荷样本进行对比,然后才能完成负荷类型识别,因而 k -NN 算法虽然没有模型训练过程,但是在 NILM 的负荷识别阶段却需要大量的计算,且计算时间随负荷样本的增加而增加。相对而言,MLP 神经网络、LR 和 SVM 都是将待识别的负荷样本输入训练好的模型,进行固定的数学计算继而完成负荷类型识别,因此负荷识别速度较快。理论上,MLP 神经网络、LR 和 SVM 在 NILM 的应用中应当具有较好的实时性,而利用 k -NN 算法进行负荷识别时需要适当地确定负荷样本的规模以确保负荷识别的实时性。

3 NILM 监督学习分类算法性能对比分析

本文从识别精度、训练时间、识别速度和抗噪能力方面对 MLP 神经网络、 k -NN 算法、LR 和 SVM 这 4 种 NILM 分类算法进行对比分析。首先对 4 种分类器进行训练和测试,对比其不同电器设备上的识别准确率;然后对 4 种分类器的收敛状况、训练时

间、识别速度等特性进行对比,分析各分类器的训练特点;最后对训练样本和测试样本添加电流噪声,对比分析 4 种分类器在不同噪声水平下的识别效果。

3.1 仿真测试环境

本文采用卡内基梅隆大学公开的 BLUED 数据库进行仿真测试,该数据库使用 12 kHz 的采样频率,在 1 个美国家庭的电力端口处连续采集了一周的电压、电流数据,并标记了各个电器设备的投切时刻。以每次投切事件前、后端口处的稳态电流之差作为投切负荷电流样本,每次投切事件均提取 100 个负荷电流样本,并根据电压校正电流波形相位。训练数据采用 580 次负荷投切事件,共计 58 000 个负荷电流样本作为训练集,包含了 8 个设备类别(7 种常用设备和 1 种其他设备);测试数据采用数据库中具有相同设备类别的另外 248 次负荷投切事件,共计 24 800 个负荷电流样本作为测试集。

仿真测试软件平台使用 MATLAB R2010a,硬件平台使用 Intel i5-2410M 2.3 GHz,4 GB RAM 电脑,4 种分类算法均在该平台上编程实现,并使用相同的数据集进行测试,利用相同的计算环境从而保证算法性能分析的有效性。

3.2 识别精度对比分析

对 4 种分类器使用相同的训练集和测试集,根据分类器在测试集中的识别效果对比分析不同分类器的分类精度。表 1 为 MLP 神经网络、 k -NN 算法、LR 和 SVM 在测试集上的识别准确率。表中,总体准确率为测试集中准确识别的样本占测试集总样本的比例。从表 1 中可以看出,MLP 神经网络、 k -NN 算法、SVM 几乎能够对各类电器设备进行准确分类,而 LR 在冰箱、浴室顶灯和未知设备的识别上出现了较多错误。这是因为 MLP 神经网络、 k -NN 算法、基于高斯核的 SVM 均是非线性分类器,能够形成不规则的负荷分类边界,而 LR 仅能进行负荷的线性分类,对于非线性分类边界负荷会出现明显的分类错误。因此,从识别精度来看,MLP 神经网络、 k -NN 算法、SVM 明显优于 LR,更适用于 NILM。

表 1 4 种分类器的识别准确率

Table 1 Recognition accuracy of four classifiers

设备	识别准确率/%			
	MLP 神经网络	k -NN 算法	LR	SVM
厨房切碎机	99.9	99.9	99.9	99.9
冰箱	99.8	99.8	83.7	99.6
空气压缩机	99.9	99.8	99.6	99.8
吹风机	99.9	99.9	99.9	99.9
后院灯	99.8	99.9	99.8	99.9
浴室顶灯	99.6	99.8	95.7	99.4
卧室灯	99.9	99.6	99.8	99.8
其他设备	99.9	99.9	57.9	99.9
总体准确率/%	99.8	99.8	85.2	99.7

3.3 训练及识别过程对比分析

采用 3.1 节所述训练集对 MLP 神经网络和 LR

进行训练,然后进行负荷识别准确率测试,它们在训练和测试中的识别准确率与迭代训练次数的关系如图 3 所示。

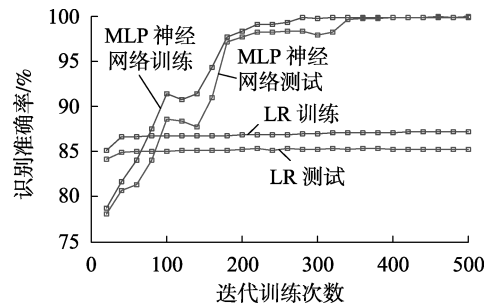


图 3 MLP 神经网络和 LR 的训练和测试过程

Fig.3 Training and testing process of MLP neural network and LR

由图 3 可见,相比于 LR,MLP 神经网络的收敛速度较为缓慢,但最终能获得较好的识别准确率;LR 能较快收敛于全局最优解,但通常其识别准确率较低。对于 k -NN 和 SVM 分类器而言, k -NN 分类器是基于实例的学习算法,其对于每个测试样本可直接计算识别样本所属类别,不存在训练过程,且识别准确率较高;SVM 分类器的参数优化问题可转化为一个凸二次规划^[18],收敛速度较快,其识别准确率与 k -NN 分类器相当。

表 2 为 4 种分类器训练至收敛后的测试情况,可以看出 k -NN 算法由于其基于实例学习的特殊性,无训练时间,而 SVM 和 LR 的训练速度相对较快,时间大约在 10 s 级,MLP 神经网络的训练时间则达到了分钟级,训练速度较慢。因而,MLP 神经网络虽然具有较好的识别准确率,但是其收敛速度慢,训练时间长,在 NILM 应用中其只能采用离线训练方式,而其他算法可以在 NILM 装置中进行模型训练,灵活性较好。

表 2 4 种分类器训练、测试结果

Table 2 Training and testing results of four classifiers

分类器	训练时间/s	单样本识别时间/ μ s	识别准确率/%
MLP 神经网络	223.2	1.81	99.8
k -NN 算法	0	1.89×10^3	99.8
LR	18.5	0.46	85.2
SVM	7.92	243.6	99.7

从分类器的识别速度来看,各分类器表现了与训练时间近乎相反的特性。LR 和 MLP 神经网络在微秒级时间即能够实现负荷识别,SVM 则需 243.6 μ s,而 k -NN 算法则需要 1.89 ms 才能识别 1 个负荷电流样本,识别时间约为 SVM 的 7.8 倍。考虑到工程中 NILM 装置的 CPU 较慢且存在人机交互、远程通信等其他工作模块,具有较快识别速度的 LR 和 MLP 神经网络将更适用于 NILM 装置的在线识别。

另外,分类器由于其训练方式不同,在不同规模

的训练集下其训练时间差异较大,同时也会影响分类器对新样本的识别速度。图4为4种分类器在5%~100%的训练集样本(即每次投切事件分别取5~100个电流样本)下,各个分类器训练时间与训练样本规模的关系。

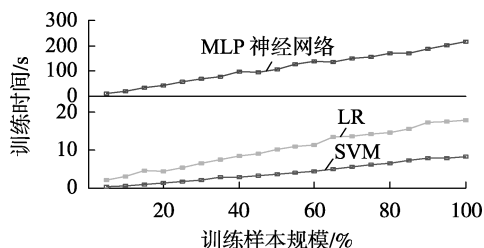


图4 分类器训练时间与训练样本规模的关系

Fig.4 Relationship between training time of classifiers and training sample size

图4中, k -NN算法由于其特殊性,不存在参数训练过程,因而无训练曲线;对于MLP神经网络和LR而言,通常采用梯度下降的方法进行训练,其训练时间和训练样本数量成正比;对于SVM而言,当样本维数确定时,其训练时间的复杂度约为 $O(N_{sv}^3 + NN_{sv})$ ^[19],其中 N_{sv} 为支持向量的个数, N 为训练样本个数,在本节的训练方式下,不同样本规模下的 N_{sv} 变化较小,因此SVM训练时间的复杂度近似退化为 $O(N)$,即SVM的训练时间与训练样本规模呈线性关系。总体来看,MLP神经网络的训练时间还是远大于LR和SVM方法,大体达到分钟级;各类算法的训练时间与训练样本数量正相关。基于以上特性,MLP神经网络仅能用于NILM离线训练方式,LR和SVM在训练负荷样本数量较小时能在NILM装置中训练,训练负荷样本较多时 also 需采用离线训练方式,然后将训练好的模型载入NILM装置应用。

图5为在5%~100%的训练集样本下,各分类器单个负荷样本识别时间与训练样本规模的关系。由图5可以看出, k -NN算法单个负荷样本的识别时间和训练样本数量呈线性关系,MLP神经网络和LR单个负荷样本的识别时间稳定不变,与训练样本数

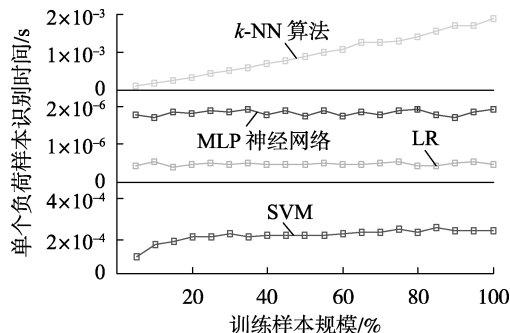


图5 分类器单个负荷样本识别时间与训练样本规模的关系

Fig.5 Relationship between recognition time of single load sample and training sample size

量无关,SVM单个负荷样本的识别时间则表现为训练样本数量的对数函数。一般而言,负荷训练样本的数量越大,模型越精确,负荷识别效果越好;但随着训练样本增加, k -NN算法单个负荷样本的识别时间线性增大,影响了NILM识别的实时性;而MLP神经网络、LR和SVM单个负荷样本的识别时间较为稳定,且识别速度快,适用于NILM在线负荷识别。

3.4 抗噪能力对比分析

本节先在3.1节所述的训练集和测试集电流样本信号加入高斯白噪声,再对4种分类器进行训练和测试,以对比分析不同方法的抗噪能力。噪声强度采用信噪比SNR(Signal-to-Noise Ratio)进行衡量,SNR越小,则噪声相对越强。各个分类器的识别准确率与负荷电流SNR的关系如图6所示。

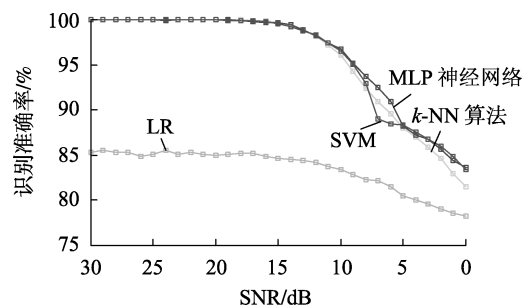


图6 4种分类器的识别准确率与SNR的关系

Fig.6 Relationship between recognition accuracy of four classifiers and SNR

由图6可以看出,随着SNR逐渐下降、噪声强度逐渐增大,各分类器的识别准确率均出现了不同程度的下降。MLP神经网络、 k -NN算法、SVM的识别准确率从SNR为19 dB时开始下降,下降过程中MLP神经网络的识别准确率略高于 k -NN算法和SVM;随着噪声强度的增大,LR的识别准确率也不断下降,降幅相对较小。在有噪声的情况下,MLP神经网络、 k -NN算法、SVM这3种分类器的识别准确率仍然显著优于LR。

图7为SVM的训练时间和单个负荷样本识别时间与SNR的关系。由图7可以看出,随着噪声强度增强,SVM的训练时间和单个负荷样本识别时间出现了急剧增加。可见随着SNR减小、噪声强度增大,SVM的模型训练和负荷识别时间快速增大,并

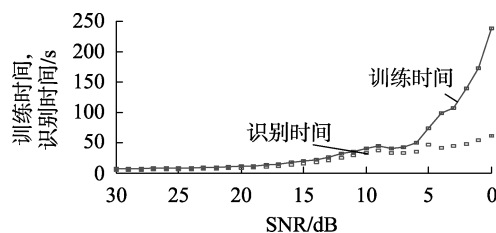


图7 SVM训练时间、识别时间与SNR的关系

Fig.7 Relationship between training time and recognition time of SVM and SNR

在 SNR 为 4 dB 时超过了 k -NN 算法的识别时间。SVM 出现训练、识别时间急剧增加的原因是:当负荷样本的噪声强度增大时,支持向量个数 N_{sv} 快速增大,SVM 的训练和识别时间与 N_{sv}^3 成正比,因而急剧增加。相比而言,其他 3 种分类器在不同噪声强度下的训练时间和识别时间均稳定在表 2 所示数据附近,性能较为稳定。

从抗噪能力角度分析,SVM 在 NILM 中仅能用于负荷电流噪声较低场合,在高噪声环境下 SVM 的实时性将大幅度降低;而 MLP 神经网络、LR 和 k -NN 算法在不同电流噪声水平下的识别时间比较稳定,且 MLP 神经网络的识别准确率最优。

3.5 综合性能对比分析

表 3 为 4 种分类器的性能综合对比。由表 3 可知,LR 的计算速度和稳定性最优,但其识别精度低,提升有限且非常困难; k -NN 算法简单、可靠,能够划分高度复杂的分类边界,但是其识别需要大量计算时间和内存容量,应用于 NILM 在线识别受限;SVM 能够最小化结构风险,负荷识别准确率较好,但其训练、识别时间对噪声较为敏感,适用于低噪声环境下的 NILM 场合;相对而言,MLP 神经网络虽然训练较慢,但其判断准确、识别快速、稳定性好,对于通常采用的离线训练、在线负荷识别的家庭 NILM 最为适用。

表 3 4 种分类器的性能对比

Table 3 Performance comparison among four classifiers

性能	MLP 神经网络	k -NN	LR	SVM
识别准确率	高	高	中	高
训练时间	分钟级	—	10 s 级	10 s 级
识别时间	微秒级	毫秒级	微秒级	100 μ s 级
训练时间复杂度与训练样本数量的关系	$O(N)$	—	$O(N)$	$O(N)$
识别时间复杂度与训练样本数量的关系	常数	$O(N)$	常数	$O(\lg N)$
噪声下的识别准确率	良好	良好	中	良好
噪声下的计算稳定性	好	好	好	差

4 结论

NILM 能够根据家庭电力入口处的整体电气信息分析用户内部各个设备的工作状态,为家庭负荷监测和智能用电提供决策指导。本文选取家庭用电设备稳态电流的奇次谐波作为负荷特征,建立了 4 种具有代表性的监督学习分类算法——MLP 神经网络、 k -NN 算法、LR 和 SVM,对比了它们在 NILM 中的应用效果,并比较分析了 4 种算法在识别精度、训练时间、识别速度和抗噪性能方面的表现,指出了各自的特点。分析结果表明:在 4 种分类器中,MLP 神经网络最适用于家庭 NILM。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] 王守相,葛磊蛟,王凯. 智能配电系统的内涵及其关键技术[J]. 电力自动化设备,2016,36(6):1-6.
WANG Shouxiang, GE Leijiao, WANG Kai. Main contents and key technologies of smart distribution system[J]. Electric Power Automation Equipment, 2016, 36(6): 1-6.
- [2] 程祥,李林芝,吴浩,等. 非侵入式负荷监测与分解研究综述[J]. 电网技术,2016,40(10):3108-3117.
CHENG Xiang, LI Linzhi, WU Hao, et al. A survey of the research on non-intrusive load monitoring and disaggregation[J]. Power System Technology, 2016, 40(10): 3108-3117.
- [3] 刘博. 非侵入式电力负荷监测与分解技术[D]. 天津:天津大学,2014.
LIU Bo. Non-intrusive power load monitoring and disaggregation technique[D]. Tianjin: Tianjin University, 2014.
- [4] AZAZA M, WALLIN F. Supervised household's loads pattern recognition[C] // Electrical Power and Energy Conference (EPEC). [S.l.]: IEEE, 2016: 1-5.
- [5] 祁兵,程媛,武昕. 基于 Fisher 有监督判别非侵入式居民负荷辨识方法[J]. 电网技术,2016,40(8):2484-2491.
QI Bing, CHENG Yuan, WU Xin. Non-intrusive household appliance load identification method based on Fisher supervised discriminant[J]. Power System Technology, 2016, 40(8): 2484-2491.
- [6] CHANG H H, LIN L S, CHEN N, et al. Particle swarm optimization based nonintrusive demand monitoring and load identification in smart meters[J]. IEEE Transactions on Industry Applications, 2013, 49(5): 2229-2236.
- [7] 李如意,王晓换,胡美璇,等. RPROP 神经网络在非侵入式负荷分解中的应用[J]. 电力系统保护与控制,2016,44(7):55-61.
LI Ruyi, WANG Xiaohuan, HU Meixuan, et al. Application of RPROP neural network in nonintrusive load decomposition[J]. Power System Protection and Control, 2016, 44(7): 55-61.
- [8] DUFOUR L, GENOUD D, JARA A, et al. A non-intrusive model to predict the exible energy in a residential building[C] // Wireless Communications and Networking Conference Workshops (WCNCW). [S.l.]: IEEE, 2015: 69-74.
- [9] 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.
SUN Yi, CUI Can, LU Jun, et al. A non-intrusive household load monitoring method based on genetic optimization[J]. Power System Technology, 2016, 40(12): 3912-3917.
- [10] 刘兴杰,曹美晗,许月娟. 基于改进鸡群算法的非侵入式负荷监测[J]. 电力自动化设备,2018,38(5):235-240.
LIU Xingjie, CAO Meihan, XU Yuejuan. Non-intrusive load monitoring based on improved chicken swarm optimization algorithm[J]. Electric Power Automation Equipment, 2018, 38(5): 235-240.
- [11] 牛卢璐,贾宏杰. 一种适用于非侵入式负荷监测的暂态事件检测算法[J]. 电力系统自动化,2011,35(9):30-35.
NIU Lulu, JIA Hongjie. A transient event detection algorithm for nonintrusive load monitoring[J]. Automation of Electric Power Systems, 2011, 35(9): 30-35.
- [12] DONG M, MEIRA P C M, XU W, et al. Non-intrusive signature extraction for major residential loads[J]. IEEE Transactions on Smart Grid, 2013, 4(3): 1421-1430.
- [13] ANDERSON K, OCNEANU A, BENITEZ D, et al. BLUED: a fully labeled public dataset for event-based non-intrusive load monitoring research[C] // Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD). Beijing, China:

- IEEE, 2012; 1-5.
- [14] HART G W. Nonintrusive appliance load monitoring[J]. Proceedings of the IEEE, 1992, 80(12): 1870-1891.
- [15] 阎平凡, 张长水. 人工神经网络与模拟进化计算[M]. 北京: 清华大学出版社, 2005: 10-37.
- [16] KREINOVICH V Y. Arbitrary nonlinearity is sufficient to represent all functions by neural networks; a theorem[J]. Neural Networks, 1991, 4(3): 381-383.
- [17] BISHOP C M. Pattern recognition and machine learning[M]. New York, USA: Springer-Verlag, 2006: 181-192.
- [18] 王彬, 郭文鑫, 向德军, 等. 基于改进支持向量机和两步式聚类分析的电网关键断面辨识和精细规则生成方法[J]. 电力自动化设备, 2017, 37(9): 166-170, 177.
WANG Bin, GUO Wenxin, XIANG Dejun, et al. Grid critical section detection and fine operational rule generation based on improved support vector machine and two-step clustering analysis [J]. Electric Power Automation Equipment, 2017, 37(9): 166-170, 177.
- [19] MOUNTRAKIS G, IM J, OGOLE C. Support vector machines in remote sensing: a review[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2011, 66(3): 247-259.

作者简介:



涂京

涂京(1992—),男,湖北十堰人,博士研究生,主要研究方向为智能用电及大数据分析(**E-mail**: tujing_128@163.com);

周明(1967—),女,湖北黄冈人,教授,博士,主要研究方向为电力系统优化运行、电力系统规划与可靠性、电力市场等(**E-mail**: zhouming_ncepu.edu.cn);

宋旭帆(1993—),男,河北石家庄人,硕士研究生,主要研究方向为负荷监测及负荷特性分析(**E-mail**: xfsong2015@163.com)。

Comparison of supervised learning-based non-intrusive load monitoring algorithms

TU Jing, ZHOU Ming, SONG Xufan, ZHOU Guangdong, LI Gengyin

(State Key Laboratory for Alternate Electrical Power System with Renewable Energy Sources,
North China Electric Power University, Beijing 102206, China)

Abstract: NILM (Non-Intrusive Load Monitoring) can realize the type identification of the user's electrical equipment and the load decomposition with low costs and without disturbing the users' normal electricity consumption, which is especially suitable for residential load monitoring. The installations of massive smart meters for domestic consumers provide data support for residential NILM, which makes the residential NILM become a hot research topic. Based on the samples of residential load steady-state current, the harmonic coefficients of the load steady-state current are used as the load classification characteristics, and four NILM classification models based on MLP (Multi-Layer Perceptions) neural network, k -nearest neighbor algorithm, logistic regression and SVM (Support Vector Machine) are established. The four types of classifiers are trained and tested with the BLUED database, their performances of identification accuracy, training time, identification speed and anti-noise property are compared, and their application effects in residential load identification are contrasted. Results show that, among the four classifiers, the MLP neural network has the optimal classification effect and computational performance, which is more suitable for monitoring the residential load.

Key words: non-intrusive load monitoring; load identification; multi-layer perceptions neural networks; support vector machines; supervised learning

附 录

A1 MLP 神经网络的具体实现过程

MLP 神经网络是将单一感知器模型进行多层叠加形成的通用神经网络模型，由 1 个输入层、若干隐含层和 1 个输出层组成。其具体结构如图 A1 所示。

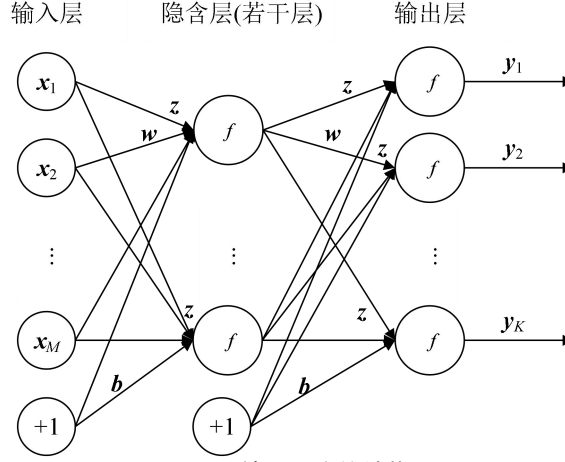


图 A1 MLP 神经网络的结构图
Fig.A1 Structure of MLP neural network

对于 NILM 而言，若模型的训练集 Φ 中有 N 个负荷样本，每个样本的负荷特征为 M 维，总计有 K 个负荷类别，则第 i 个负荷样本 $\mathbf{x}_i = [x_{i,1}, x_{i,2}, \dots, x_{i,M}]$ ，其负荷类别 $\mathbf{y}_i = [y_{i,1}, y_{i,2}, \dots, y_{i,K}]$ ，当 $\mathbf{x}_i$ 属于第 j 类负荷时， $y_{i,j}=1$ ，否则 $y_{i,j}=0$ 。由此构建模型的训练集 $\Phi = \{(\mathbf{x}_i, \mathbf{y}_i)\}$ 。

对于隐含层和输出层而言，其激励函数 f 在 NILM 中选取为 Sigmoid 函数如下：

$$f(z) = \frac{1}{1 + e^{-z}} \quad (\text{A1})$$

$$\mathbf{z} = \mathbf{a}^T \mathbf{w} + \mathbf{b} \quad (\text{A2})$$

其中， $\mathbf{z}$ 为上一层神经元的输出向量 $\mathbf{a}$ 在该神经元处的加权和； $\mathbf{w}$ 为加权的权重； $\mathbf{b}$ 为 +1 神经元的权重。

在 MLP 神经网络的训练过程中，首先进行前向传播计算，即从输入层的负荷样本 $\mathbf{x}_i$ 开始，按照加权求和、计算激励函数的顺序，逐层计算隐含层的激励函数输出，并以激励函数输出作为下一层的输入值，直至获得输出层激励函数值 $\mathbf{y}'_i$ 。

然后计算正则化的神经网络代价函数，即：

$$C = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K [y_{i,j} \ln y'_{i,j} + (1 - y_{i,j}) \ln (1 - y'_{i,j})] + \frac{\lambda}{2N} \sum \|\mathbf{w}\|^2 \quad (\text{A3})$$

对于代价函数而言，首先计算无正则化项时各个网络权重的梯度。根据代价函数和激励函数，运用求导的链式法则，可以求得输出层向量 $\mathbf{z}$ 关于代价函数的梯度 $\nabla \mathbf{z}$ ；然后再根据链式法则，可求得上一次输出向量 $\mathbf{a}$ 和两层之间网络权重向量 $\mathbf{w}$ 的梯度如下：

$$\nabla \mathbf{w} = \nabla \mathbf{z} \times \mathbf{a} \quad (\text{A4})$$

$$\nabla \mathbf{a} = \mathbf{w}^T \times \nabla \mathbf{z} \quad (\text{A5})$$

再次通过链式法则反复计算上一层的 $\nabla \mathbf{z}$ 、 $\nabla \mathbf{w}$ 、 $\nabla \mathbf{a}$ ，直至求解至输入层为止。然后在原先求得的网络权重梯度上增补正则化项以获得关于网络权重完整梯度 $\nabla \mathbf{w}'$ ，可应用梯度下降法或其他基于梯度的迭代方法进行求解。

基于梯度下降的迭代公式如下：

$$\nabla \mathbf{w}' = \nabla \mathbf{w} + \frac{\lambda}{N} \mathbf{w} \quad (\text{A6})$$

$$\mathbf{w} = \mathbf{w} - \alpha \nabla \mathbf{w}' \quad (\text{A7})$$

其中， α 为迭代学习率，需要预先确定。

当经过多次迭代，代价函数基本不再下降后，则认为模型收敛，MLP 神经网络模型的训练过程结束。

负荷识别过程中，将待识别负荷样本 $\hat{\mathbf{x}}$ 代入 MLP 神经网络模型进行前向计算，从而求得网络输出 $\hat{\mathbf{y}}=[\hat{y}_1, \hat{y}_2, \dots, \hat{y}_K]$ ，选取其中数值最大的一项所对应的负荷类型即为待识别负荷的负荷类型。

此外，亦可使用 MATLAB 自带的神经网络工具箱实现上述过程，具体可参考文献[A1]。

A2 k -NN 分类模型的具体实现过程

k -NN 算法是一种用于分类的非参数统计方法，其原理简洁明了，能够形成复杂的分类边界，但对局部训练样本的构成非常敏感。因此，本文采用了 $1/d$ 加权的 k -NN 算法作为 NILM 的负荷分类模型，应用遍历法进行样本距离计算，其具体实现步骤如下：

- a. 计算待识别负荷样本 $\hat{\mathbf{x}}$ 与训练集 $\Phi=\{(\mathbf{x}_i, \mathbf{y}_i)\}$ 中每个负荷样本的距离 $d_i=\|\hat{\mathbf{x}}-\mathbf{x}_i\|^2$ ；
- b. 从计算得到的 d_i 中选取前 k 个距离最小的 d_i 及相应的负荷样本 $(\mathbf{x}_i, \mathbf{y}_i)$ ，此 k 个负荷样本构成最近邻集合 $\mathcal{N}$ ，其与待识别负荷样本 $\hat{\mathbf{x}}$ 的距离形成距离列向量 $\mathbf{D}=[d_i]$ ；
- c. 构建负荷类别与最近邻集合 $\mathcal{N}$ 的关系矩阵 $\mathbf{H}=[h_{i,j}]$ ，其中 $h_{i,j}$ 表示 $\mathcal{N}$ 中的第 i 个负荷样本是否属于第 j 个负荷类别，若属于则 $h_{i,j}=1$ ，否则 $h_{i,j}=0$ ；
- d. 用距离列向量 $\mathbf{D}$ 乘以关系矩阵 $\mathbf{H}$ ，计算所得的行向量中数值最大的项对应的负荷类别即为待识别负荷样本 $\hat{\mathbf{x}}$ 的负荷类别。

此外，MATLAB 同样自带 k -NN 算法工具包，亦可用该工具包实现上述过程。

A3 LR 的具体实现过程

LR 作为一种经典分类算法广泛应用于各类分类问题，其采用 Sigmoid 函数嵌套线性回归的形式处理二分类问题。本文采用“一对多”策略将 LR 扩展为多分类器，具体实现方法如下。

“一对多”策略下的 LR 训练过程中，需要对每个类别训练一个 LR 分类器。对于第 k 个负荷类别，先将训练集 $\Phi=\{(\mathbf{x}_i, \mathbf{y}_i)\}$ 进行适当修改，若 $\mathbf{x}_i$ 属于该负荷类别，则定义其负荷类别为 1，否则为 0，由此构建新训练集 $\Phi_k=\{(\mathbf{x}_i, \mathbf{y}_i^k)\}$ 。

然后计算正则化 LR 的代价函数如下：

$$\mathbf{z} = \mathbf{w}^T \mathbf{x}_i + b \quad (\text{A8})$$

$$\hat{\mathbf{y}}_i^k = \frac{1}{1 + e^{-z}} \quad (\text{A9})$$

$$C = -\frac{1}{N} \sum_{i=1}^N [y_i^k \ln \hat{y}_i^k + (1 - y_i^k) \ln (1 - \hat{y}_i^k)] + \frac{\lambda}{2N} \sum_{j=1}^M w_j^2 \quad (\text{A10})$$

其中， $\mathbf{w}=[w_1, w_2, \dots, w_M]$ 为 LR 模型的权重向量， b 为截距变量，两者为 LR 模型的参数。

应用链式求导法则获取 $\mathbf{w}$ 和 b 的梯度：

$$\nabla \mathbf{w} = \mathbf{w} \times \frac{1}{N} \sum_{i=1}^N (\hat{\mathbf{y}}_i^k - \mathbf{y}_i^k) \quad (\text{A11})$$

$$\frac{\partial C}{\partial b} = \frac{1}{N} \sum_{i=1}^N (\hat{\mathbf{y}}_i^k - \mathbf{y}_i^k) \quad (\text{A12})$$

然后可按照附录 A1 中的梯度下降法进行求解，此处不再赘述。

“一对多”策略需要对每个负荷类别均按照上述过程训练一个 LR 分类器。在负荷识别过程中，将待识别负荷样本 $\hat{\mathbf{x}}$ 代入各类别的 LR 分类器中，按式 (A8)、(A9) 计算得到 K 个类别的 $\hat{\mathbf{y}}^k$ ，其中数值最大项所对应的负荷类别即为待识别负荷样本 $\hat{\mathbf{x}}$ 的负荷类别。

A4 SVM 的具体实现过程

SVM 是一种基于监督学习的大间距分类器，其核心理念为在特征空间中寻找尽可能大的分类间隔以有效分隔正类和负类。SVM 是一种二分类器，因此同样应用“一对多”策略将 SVM 扩展为多分类器。

“一对多”策略下的 SVM 训练过程中，需要对每个类别训练一个 SVM 分类器。对于第 k 个负荷类别，先将训练集 $\Phi=\{(\mathbf{x}_i, \mathbf{y}_i)\}$ 进行适当修改，若 $\mathbf{x}_i$ 属于该负荷类别，则定义其负荷类别为 1，否则为 -1，由此构建出新训练集 $\Phi_k=\{(\mathbf{x}_i, \mathbf{y}_i^k)\}$ 。

SVM 分类器的标准形式如下：

$$\begin{cases} \min_{\{\mathbf{w}, \mathbf{b}, \xi_i\}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad \mathbf{y}_i^k (\mathbf{w} \cdot \mathbf{x}_i + \mathbf{b}) \geq 1 - \xi_i \quad i=1, 2, \dots, N \\ \xi_i \geq 0 \quad i=1, 2, \dots, N \end{cases} \quad (\text{A13})$$

其中， $\mathbf{w}$ 和 $\mathbf{b}$ 为 SVM 的参数； C 为惩罚系数，其值越大则对分类错误惩罚越大，需预先给定。

根据对偶原理，式 (A13) 的对偶形式为：

$$\begin{cases} \min_{\{\alpha\}} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i^k y_j^k (\mathbf{x}_i \cdot \mathbf{x}_j) - \sum_{i=1}^N \alpha_i \\ \text{s.t.} \quad \sum_{i=1}^N \alpha_i y_i^k = 0 \\ 0 \leq \alpha_i \leq C \quad i=1, 2, \dots, N \end{cases} \quad (\text{A14})$$

为了加强 SVM 的分类能力，需将负荷特性映射到高维空间，则需定义映射函数 $\phi(\mathbf{x})$ ，以 $\phi(\mathbf{x})$ 替换式 (A14) 中的 $\mathbf{x}$ 。同时，可以看出，在对偶问题中，主要涉及负荷样本 $\mathbf{x}_i$ 与 $\mathbf{x}_j$ 的内积，因此可以不显式定义映射函数 $\phi(\mathbf{x})$ ，仅需定义其内积即可，此内积函数称之为核函数，相应的方法称之为核技巧。

本文应用高斯函数 G 作为 SVM 的核函数，因此，式 (A14) 转化为：

$$\begin{cases} \min_{\{\alpha\}} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i^k y_j^k G(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^N \alpha_i \\ \text{s.t.} \quad \sum_{i=1}^N \alpha_i y_i^k = 0 \\ 0 \leq \alpha_i \leq C \quad i=1, 2, \dots, N \end{cases} \quad (\text{A15})$$

上述模型是一个凸二次规划模型，采用 Cplex 等求解器即可进行模型求解。

同时，基于高斯核函数 SVM 输出函数为：

$$z(\mathbf{x}) = \sum_{i=1}^N \alpha_i y_i^k G(\mathbf{x}_i, \mathbf{x}) + \mathbf{b} \quad (\text{A16})$$

“一对多”策略需要对每个负荷类别均按照上述过程训练一个 SVM 分类器。负荷识别过程中，将待识别负荷样本 $\hat{\mathbf{x}}$ 代入各类别的 SVM 分类器中，计算每个类别的输出，其中数值最大的项所对应的负荷类别即为待识别负荷样本 $\hat{\mathbf{x}}$ 的负荷类别。MATLAB 同样自带 SVM 算法工具包，亦可用该工具包实现上述过程。SVM 理论具体可参考文献[A2]。

参考文献：

- [A1] 丛爽. 面向 MATLAB 工具箱的神经网络理论与应用[M]. 北京：中国科学技术大学出版社，2009：20-100.
[A2] 李航. 统计学习方法[M]. 北京：清华大学出版社，2012：95-134.