



电测与仪表

Electrical Measurement & Instrumentation

ISSN 1001-1390,CN 23-1202/TH

《电测与仪表》网络首发论文

题目：基于聚类特征及 seq2seq 深度 CNN 的家电负荷识别方法研究
作者：汪繁荣，向堃，吴铁洲
收稿日期：2020-08-25
网络首发日期：2020-10-16
引用格式：汪繁荣，向堃，吴铁洲. 基于聚类特征及 seq2seq 深度 CNN 的家电负荷识别方法研究. 电测与仪表.
<https://kns.cnki.net/kcms/detail/23.1202.TH.20201016.1636.014.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于聚类特征及 seq2seq 深度 CNN 的家电负荷识别方法研究

汪繁荣^{1,2}, 向堃¹, 吴铁洲¹

(1.湖北工业大学 太阳能高效利用及储能运行控制湖北省重点实验室, 武汉 430068; 2.无锡风繁伟业科技有限公司, 江苏 无锡 214171)

摘要: 非侵入式负荷分解技术能够挖掘用户内部信息获取各用电设备负荷信息, 使智能电网更加贴近日常生活, 为泛在电力物联网感知层建立提供有效数据。为解决传统非侵入式负荷分解方法输入数据复杂, 考虑因素较多, 采样硬件要求高以及辨识准确率较低等问题, 文章首先利用改进迭代 K 均值聚类提取用电设备运行状态建立负荷特征集, 之后将特征集输入构造的序列到序列的一维深层卷积神经网络模型以及序列到序列的单、双向长短时记忆网络等模型中进行负荷分解挖掘各设备运行状态。最后通过 REFITPowerData 数据集进行验证, 一维深层卷积神经网络模型虽然耗时较大但负荷识别准确率达到 93% 以上, 表明基于特征数据集及序列到序列的一维深层卷积神经网络非侵入式负荷分解方法与其他深度学习模型方法、人工神经网络方法相比表现出更显著的信息提取能力以及辨识能力。

关键词: 非侵入式负荷分解; 深度学习; 卷积神经网络; 序列到序列; 特征提取; 数据挖掘

中图分类号: TM714

文献标识码: A

文章编号: 1001-1390(2020)00-0000-00

Research on household appliance load identification method based on clustering features and seq2seq depth CNN

Wang Fanrong^{1,2}, Xiang Kun¹, Wu Tiezhou¹

(1. Hubei Key Laboratory for High-efficiency Utilization of Solar Energy and Operation Control of Energy Storage System, Hubei University of Technology, Wuhan 430068, China. 2. Wuxi Fengfan Weiye Technology Co., Ltd., Wuxi 214171, Jiangsu, China)

Abstract: Internal information of users can be mined by non-intrusive load decomposition (NILD) to obtain load information of various electrical equipments, which enables the smart grid to have a closer connection with daily life and provides effective data for the establishment of perception layer of ubiquitous power Internet of things (UPIoT). However, there exist some problems regarding traditional NILD, for example, the input data is complicated and a lot of factors need to be considered. Besides, sampling hardware is highly demanding and identification accuracy is relatively low. In order to solve these problems, the operating state of electrical equipment is extracted by using improved iterative K-means clustering to establish characteristic data set firstly. And then, constructed by inputting data set, sequence-to-sequence one-dimensional deep convolutional neural network (1-D-DCNN) and sequence-to-sequence LSTM and Bi-LSTM network model are decomposed to mine user information. Finally, verified by the REFITPowerData, the identification accuracy of 1-D-DCNN is over 93% though it is quite time-consuming. Compared with other deep learning model and artificial neural network methods, NILD based on characteristic data set and sequence-to-sequence 1-D-DCNN show more significant information extraction and identification capabilities.

Keywords: non-intrusive load decomposition, deep learning, convolutional neural network, sequence to sequence,

0 引言

非侵入式负荷监测 (Non-Intrusive Load Monitoring, NILM) 由 Hart 提出^[1], 也称为非侵入式负荷分解 (Non-Intrusive Load Decomposition, NILD)。它通过在用户用电系统入口处安装监测设备收集其用电负荷数据对用户内部用电负荷进行细分及能耗合理分配^[2]。由于智能电表的普及和大数据技术的发展, NILD 受到了更多关注。

特征选择是 NILD 中的重要环节, 一般来说有暂态特征和稳态特征两种^[3]。暂态特征主要是用电设备投切或状态类别变换时的参数变化特征。文献[4]利用小波分解提取暂态特征, 并用深度置信网络对其波形进行类别识别, 有较好的辨识精度, 但是其数据来自软件模拟数据, 若用于实际中还需进一步检验。文献[5]利用多尺度排列熵算法确定设备状态变化区间提取设备暂态特征然后利用 Yamamoto 算法进行区间检测, 辨识用电设备启停时刻实现负荷分解, 该方法有较好的降噪和辨识效果。文献[6]利用主成分分析法进行降维来构造低维度的特征空间, 之后通过 KNN 邻近算法完成负荷分解。这类基于暂态特征的 NILD 方法对于收集数据的软硬件设备要求较高, 在实际运用中有一定的局限性。稳态特征包括稳态电压、电流、有无功率等数据特征, 随着智能电表的普及与发展, 围绕稳态特征的 NILD 方法层出不穷。由于深度学习在各个领域表现突出^[7-8], 所以近年来其在非侵入式负荷分解领域的应用也逐渐增多。文献[9]利用深度学习的框架, 提出了新的非侵入式负荷分解模型, 其利用词嵌入处理输入数据然后利用单层双向长短时记忆网络和单层单向长短时记忆网络分别编码解码, 其中加入注意力机制, 最终得到负荷分解序列, 其分解效果明显。文献[10]将稳态特征数据图像化后进行辨识, 利用局部平均分解法对输入数据进行预处理然后基于卷积神经网络对家用电器的负荷特征进行提取, 最后经过 softmax 层概率输出分解结果。文献[11]利用循环卷积神经网络模型对总功率进行负荷分解, 在利用深度学习模型的时候没有建立负荷特征集, 导致精度、鲁棒性下降。以上文章都建立了比较优秀的深度学习应用模型, 但大多数深度学习的非侵入式负荷分解方法忽略了对多个用电器同时运行情况下进行分解分析比较, 其一是因为多个用电设备同时运行时会造成数据混叠难以分辨, 其二是输入数据复杂, 难以精确辨识。

为解决辨识精度低, 需要考虑多个用电设备同时工作等问题, 本文利用深度学习的框架, 提出一种基于改进 k-means 迭代聚类算法数据集结合序列到序列 (sequence-to-sequence, seq2seq) 的一维深层卷积神经网络的 (One-Dimensional Deep Convolutional Neural Network, 1-D-DCNN) 非侵入式

负荷分解方法(km++-1-D-DCNN), 通过数据的训练和测试, 与再次建立的 seq2seq 单、多层, 单、双向长短时记忆(Long Short-Term Memory, LSTM、Bilateral Long Short-Term Memory, Bi-LSTM)网络模型以及常用的广义回归神经网络 (General Regression Neural Network, GRNN)模型、反向传播神经网络(Back Propagation Neural Network, BPNN)模型进行对比, 根据所有设备共同的状态识别准确率以及单个设备各状态识别准确率等评价指标验证本文方法的有效性。

1 用电设备负荷特征集建立

将深度学习模型应用于非侵入式负荷分解问题时, 建立用于网络训练所需的用电设备负荷特征数据集是关键环节^[12]。由于长时间采集用电设备负荷具有较大的实现难度, 且种种原因导致用电负荷会存在真空区间即没有用电设备运行状态, 因此, 通过获取典型用电设备短时间运行数据或将长时间数据进行截断以提取负荷特征, 然后生成大量设备训练数据是目前一种较为理想可行的方法。

1.1 设备状态特征提取

用电设备的不同运行状态特征可以通过聚类方法对负荷数据进行挖掘得到。在聚类过程中, 簇中心数目会影响聚类效果即影响用电设备运行状态的提取, 簇中心数目过少会造成聚类不彻底, 部分近似质点无法分开, 导致用电设备提取不全; 簇中心数目过多则会使用电设备运行状态类别增加, 甚至聚类过细产生用电设备实际所不存在的运行状态, 进而增加分解算法的复杂度。本文在 k-means 聚类算法的基础上利用 k-means++算法进行簇中心初始化并引入轮廓值指标以筛选更加有效的聚类数目, 然后迭代执行避免结果陷入局部最优, 最后使用新的初始簇中心位置寻求更为合理的聚类方案。轮廓值是该点与其自身聚类中的点的相似程度的度量。第 i 个点的轮廓值 A_i 定义为:

$$A_i = (b_i - a_i) / \max(a_i, b_i) \quad (1)$$

式中 a_i 是从第 i 个点到与 i 相同的簇集中其他点的平均距离; b_i 是从第 i 个点到不同簇集中的点的最小平均距离。轮廓值的范围为-1 到 1。较高的轮廓值表示 i 与其自身的簇集完全匹配, 若大多数点具有较高的轮廓值, 则聚类结果是合适的。反之, 则聚类结果不合适。k-means++算法使用启发式方法寻找 k-means 聚类的质心, 改进了原 k-means 聚类算法中 Lloyd^[13]算法的运行时间和最终解质量。

k-means++算法按如下方式选择质心, 假设簇数为 K ^[14]:

(1) 从数据集 X 中随机均匀选择一个观测值

该值为第一个质心 c_1 ;

(2) 计算从每个观测值到 c_1 的距离, 将 c_j 和观测值 m 之间的距离表示为 $d(x_m, c_j)$;

(3) 从 X 中随机选择下一个质心 c_2 , 概率为:

$$\frac{d^2(x_m, c_1)}{\sum_{j=1}^n d^2(x_j, c_1)} \quad (2)$$

(4) 选择中心 c_j 有以下操作:

1) 计算从每个观测值到每个质心的距离, 并将每个观测值分配给其最近的质心。

2) 对于 $m = 1, \dots, n$ 和 $p = 1, \dots, j-1$, 从 X 中随机选择质心 c_j , 概率为:

$$\frac{d^2(x_m, c_p)}{\sum_{\{h; x_h \in C_p\}} d^2(x_h, c_p)} \quad (3)$$

式中 C_p 是最接近质心 c_p 的所有观测值的集合, 而 x_m 属于 C_p 。也就是说, 选择每个后续中心

时, 其选择概率与它到已选最近中心的距离成比例。

(5) 重复步骤 (4), 直到选择了 K 个质心。

本文提出的改进迭代 k-means 聚类方法步骤如下:

1) 给定数据集 X , 选择 K 个初始簇中心 (质心), 使用 k-means++ 算法进行簇中心初始化;

2) 计算所有观测值到每个质心的点到簇质心的距离;

3) 两阶段迭代更新: 一阶段批量更新, 将每个观测值分配给离质心最近的簇。二阶段在线更新, 只要将观测值重新分配给另一质心可减少簇内点到质心距离平方和的总和, 就对该观测值执行此分配;

4) 计算每个簇中观测值的平均值, 以获得 K 个新质心位置;

5) 重复步骤 2) ~ 步骤 4), 直到簇分配不变, 或达到最大迭代次数;

6) 再次重复 1) ~ 步骤 5), 迭代 n 次, 寻找距离总合最小的解。

1.2 数据预处理及设备负荷特征的挖掘

本文使用英国思克莱斯大学发布的 REFITPowerData^[15]数据集, 该数据集包含21个家庭2013年10月到2015年5月的10种不同用电设备的单个负荷有功功率数据和总负荷有功功率数据, 采样时间为8秒。因为有些家庭的数据不全, 本文选取了2号家庭的冰箱、洗衣机、电视、水壶4个主要用电设备的有功功率数据。电视为小功率用电设备, 水壶为大功率用电设备, 冰箱为长时间工作设备, 洗衣机为多状态工作设备, 这些设备能够较好的验证后文分解方法的性能。为了建立用电设备负荷特征集, 首先对原始数据进行预处理, 本文取各用电设备有功功率数据中前100万个采样点数据并将各个用电设备运行状态的有功功率消耗数据提出且剔除相同时间的功率消耗数据, 设立提取参数以保证数据的有效性, 表1展示了各用电设备参数。数据筛选后, 为保证工作状态和非工作状态的区别, 需将采样点数量大于8万点的用电设备数据序列随机抽取8万点, 再用0补长序列长度到10万点, 反之若数据小于8万点则直接将原序列用0补长至10万点长度。

表1 主要用电设备提取参数及序列长度

Tab.1 Extraction parameters and sequence length of main electrical equipment

用电设备名称	最小功率 /W	最大功率 /W	提取后序列长度/点	输入序列长度/点
冰箱	50	110	392 582	1 000
洗衣机	20	3 000	41 155	1 000
电视	30	60	112 411	1 000
水壶	200	3 000	8 920	1 000

数据筛选后, 将采样点数量大于8万点的用电设备数据序列随机抽取8万点, 再用0补长序列至10万点长度, 反之直接将原序列用0补长至10万点长度。接下来利用本文所提出的聚类方法对各用电设备运行状态进行提取建立负荷特征集, 输入数据为1维数据, 距离度量选用绝对差之和, 重复聚类次数为100。各用电设备运行状态聚类结果如表2所示,

表2 各设备运行状态聚类结果

Tab.2 Cluster results of running status of each equipment

用电设备	轮廓值					最优聚类数目	簇中心值
	K=2	K=3	K=4	K=5	K=6		
冰箱	0.996 2	0.752 1	0.615 0	/	/	2	0.85

洗衣机	0.996 8	0.775 5	0.87	0.825 6	0.821 3	4	0,103,220,2063
电视	0.992 2	0.928 9	0.496 1	/	/	2	0,46
水壶	0.999 9	0.960 7	0.942 6	/	/	2	0,272 8

表中 K 为聚类数目,即用电设备运行状态数量(包含不工作状态,即簇中心值为 0),一般来说家用电器运行状态不会超过 6 个,所以 K 取值最大为 6。冰箱、电视和水壶均为 2 种运行状态,其轮廓值在 $K=2$ 时也是最大的,洗衣机虽然在 $K=2$ 时轮廓值最大但就实际运行情况来看其运行状态不只 2 个,所以取次大值 $K=4$ 为其最终结果。图 1 展示了冰箱和洗衣机的聚类效果。冰箱的聚类效果较好,每个簇中轮廓值均较大且无负值点,洗衣机虽有负值点但数量较少,且其他簇分布良好。

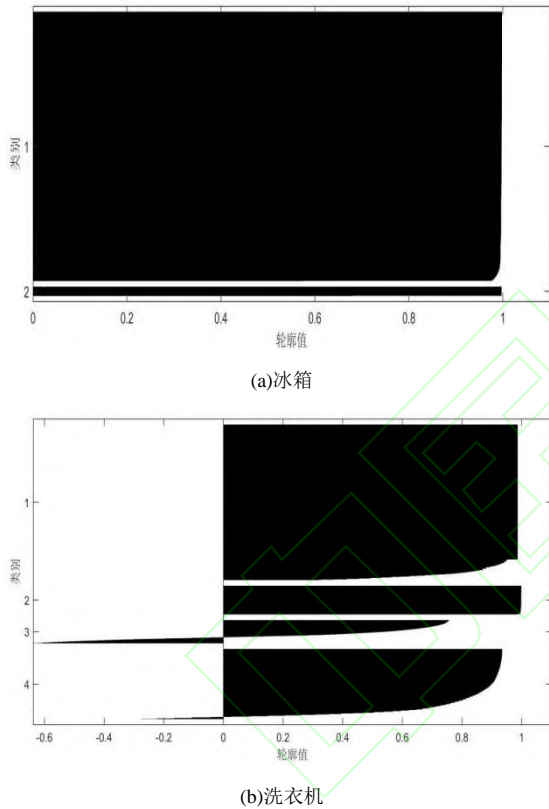


图 1 冰箱和洗衣机聚类效果

Fig.1 Clustering effect of refrigerator and washing machine

2 深度学习与非侵入式负荷分解

深度学习已经广泛应用于多个领域,作为机器学习的一个重要分支,简单来说其领域的多种算法就是深层的神经网络算法。卷积神经网络在图像处理领域一直以来有出色的表现,处理二维数据时具有很大优势^[14]。就上一节建立的设备状态特征数据集,本文将输入的一维数据(总有功功率)进行增维扩增至 4 维矩阵作为模型的输入,完成数据转化,实质上数值没有改变,提出一种序列到序列的一维深

层卷积神经网络模型进行负荷辨识,完成非侵入式负荷分解,分解模型流程如图 2 所示。另外,在负荷分解的过程中,每个用电设备都有其时间特性或其它未知特性,所以本文还构建了 LSTM 和 Bi-LSTM 网络模型进行负荷辨识。



图 2 非侵入式负荷分解模型流程图

Fig.2 Flow chart of non-intrusive load decomposition model

2.1 卷积神经网络

卷积神经网络中包含卷积层、池化层、全连接层等基本功能层。各个层具有独立的神经元,不同层神经元间相互连接,同层神经元间无连接^[16]。卷积层计算公式为:

$$X_j^l = f\left(\sum_{i \in M_j} X_j^{l-1} \otimes K_{ij}^l + b_j^l\right) \quad (4)$$

式中 X_j^{l-1} 表示 $l-1$ 层的第 j 个滤波器特征:

K_{ij}^l 表示第 l 层的第 j 个滤波器特征和第 $l-1$ 个滤波器特征的卷积核函数; $f(\cdot)$ 是激活函数; b_j^l 是 l 层的第 j 个偏置参数; \otimes 代表卷积运算方式。一般来说卷积层还有 2-D 卷积层、3-D 卷积层、转置卷积层和分组卷积层等。池化层计算公式为:

$$L_j^l = f\left(a_j^l \text{down}(L_j^{l-1}) + b_j^l\right) \quad (5)$$

式中 a_j^l 和 b_j^l 分别表示网络层之间的权重和偏置参数; $\text{down}(\cdot)$ 为池化函数; $f(\cdot)$ 是激活函数。池化层用来降低分辨率,常用的池化层有平均池化

层、最大池化层以及全局平均池化层。另外激活函数有Relu函数、Elu函数以及Tanh函数等，本文选用Relu函数，Relu函数对输入的每个元素执行阈值运算，其中任何小于零的值都设置为零。经过卷积层和池化层的计算后原始数据特征便映射到了隐含特征空间，通过全连接层标记这些特征空间便能提取原数据特征，然后利用softmax层概率分类。softmax层是将标量映射概率输出，若有 k 个标量 $x_1 - x_k$ 则运算公式为：

$$z_k = \text{soft max}(x_k) = \frac{\exp(x_k)}{\sum_{i=1}^k \exp(x_i)} \quad (6)$$

式中 $z_k \in [0,1]$ ， $\forall k$ ，并满足 $\sum_{i=1}^k z_k = 1$ 。最后

利用分类层来输出分类序列，分类层计算具有互斥类的多类分类问题的交叉熵损失，其定义如下：

$$\text{loss} = \frac{1}{N} \sum_{t=1}^N \sum_{j=0}^{v-s} h_{t,j} \quad (7)$$

$$h_{t,j} = -y_{t,j} \log \tilde{y}_{t,j} - (1 - y_{t,j}) \log(1 - \tilde{y}_{t,j}) \quad (8)$$

式中 $y_{t,j}$ 为 t 时刻真实值为 j 的概率，在 j 与 t 时刻真实值相等时该概率为1，其他情况下为0； $\tilde{y}_{t,j}$ 为 t 时刻模型计算出的分类值为 j 的概率， $y_{t,j}$ 与 $\tilde{y}_{t,j}$ 越近，则 $h_{t,j}$ 越接近0，反之则增大。

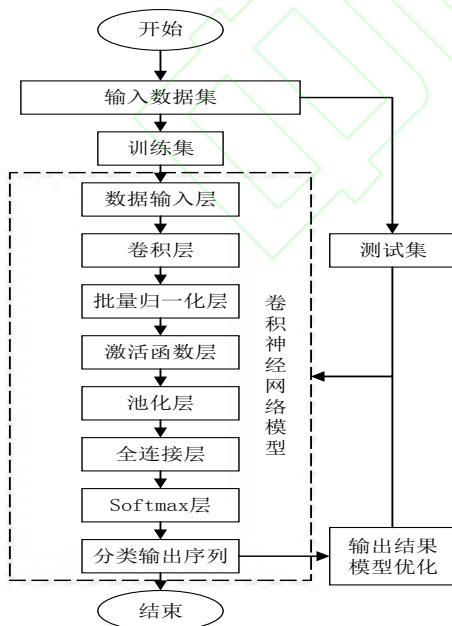


图3 卷积神经网络运算流程图

Fig.3 Flow chart of convolution neural network operation

后文使用到的网络层还有批量归一化层、dropout层（脱落层）等，批量归一化层对网络中的激活值和梯度传播进行归一化，使网络训练成为更简单的优化问题，dropout层则可以防止过度拟合。简单的卷积神经网络模型运算流程图如图3所示，另外模型得到结果后会通过反向传播方式调整模型网络参数。

2.2 一维深层卷积神经网络

图4为深层卷积神经网络的层构图，本文在常用的卷积神经网络利用二维数据图分类的基础上，改进输入层数据形式(本质还是通过2维卷积来实现分类)，达到1维时间序列输入实现分类的目的，并加深网络层数使其提取特征鲁棒性更强，分类结果更加精确。一维深层神经网络的输入是总有功率序列转化的4-D矩阵。总有功率序列为 $1 \times n$ 行向量，将其重构成 $1 \times 1 \times 1 \times n$ 的矩阵，其中前两维表示 1×1 像素，第3维1表示灰度颜色通道，第4维 n 表示序列长度。深层卷积网络由输入层、A类层、B类层、全连接层、softmax层和分类层组成，共47层。A类层一共10组40层，每组由2-D卷积层、批量归一化层、Relu函数层以及最大池化层构成，每层2-D卷积层过滤器数量分别为8/16/32/64/128/256/512/1024/2048/4096，大小均为 1×1 。B类层由2-D卷积层(过滤器数量为8 192，大小为 1×1)、批量归一化层和Relu函数层组成。为了使该模型辨识效果更好，在训练模型时加入了更多的网络参数，为防止模型过度拟合加入了L2正则化因子^[17]，其值为0.000 5，使用具有动量的梯度随机下降(SGDM)优化器训练网络，动量大小为0.97，梯度阈值为0.5，最大迭代数为300，在每一代训练时增加数据混淆，初始学习率为0.01，为了效果更佳设定学习率梯度下降，每过50代学习率在原有基础上乘0.2。

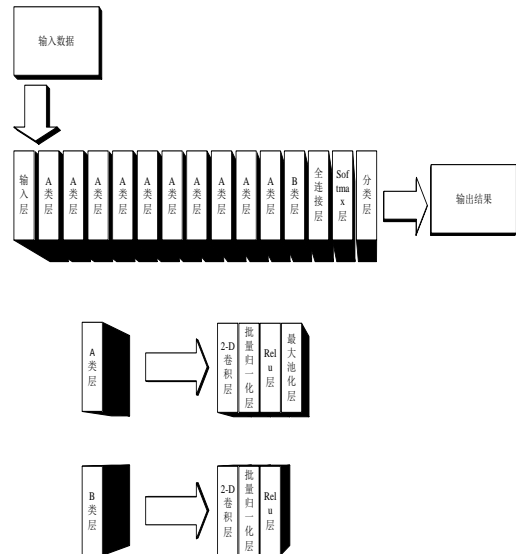


图4 深层卷积神经网络结构图

Fig.4 Structure diagram of deep convolutional neural network

2.3 LSTM 和 Bi-LSTM 神经网络

LSTM是一种特殊的循环神经网络(RNN),它解决了长依赖问题,在处理时间序列时有良好的表现^[18]。随着研究深入,根据双向循环神经网络(Bi-RNN)特性,一种Bi-LSTM神经网络也被广泛运用,其实实现了信息双向传递,在处理已知序列时,性能优于单向LSTM神经网络^[19]。

为了验证文中所提方法的有效性,本文还建立了seq2seq的LSTM神经网络和seq2seq的Bi-LSTM神经网络模型。LSTM神经网络由序列输入层、LSTM层、全连接层、dropout层(值为0.1)、softmax层以及分类层组成。网络训练参数设置如下,使用具有动量的梯度随机下降(SGDM)优化器训练网络,动量大小为0.97, L2正则化因子为0.000 5, 梯度阈值为2, 最大迭代数400, 小批量训练大小为500, 初始学习率0.01, 每过50代学习率在原有基础上乘0.2。Bi-LSTM神经网络由序列输入层、Bi-LSTM层、全连接层、dropout层(值为0.2)、softmax层以及分类层组成。网络训练参数设置如下,使用具有动量的梯度随机下降(SGDM)优化器训练网络,动量大小为0.97, L2正则化因子为0.000 5, 梯度阈值为1, 最大迭代数300, 小批量训练大小为500, 初始学习率0.01, 每过50代学习率在原有基础上乘0.2。由于实验测试得知增加LSTM层和Bi-LSTM层层数会导致辨识效果很差, 所以不再建立更深的LSTM和Bi-LSTM神经网络。

3 实验分析

本文将运用第1节已经处理过的用电设备特征数据集结合第2节介绍的深度学习分类模型来完成非侵入式负荷分解。为验证本文提出模型的性能,选取4种典型用电设备,根据家庭2的用电情况整理了12种设备用电状态,每种用电设备与其对应的运行状态编码方式如表3所示,其中冰箱、电视、水壶有2种状态(2表示运行,1表示非运行),洗衣机有4种状态(2、3、4为3个不同运行状态,1表示非运行)。网络的训练集一共12组12 000个采样点,每组1 000个数据,测试集与训练集不重复,为10 000个采样点数据序列,该序列具有时序性。网络输入为用电设备总有功功率序列,即 $1 \times 1 \times 1 \times 12000$ 4-D矩阵,输出为12个。

表3 用电设备运行编码数据表

Tab.3 Code data table of running state of electrical equipment

状态编号	用电设备运行状态			
	冰箱	洗衣机	电视	水壶
1	2	1	1	1
2	1	1	2	1

3	1	1	1	2
4	1	1	1	1
5	1	2	1	1
6	1	3	1	1
7	1	4	1	1
8	2	2	1	1
9	2	3	1	1
10	2	4	1	1
11	2	1	2	1
12	2	1	1	2

3.1 评价指标及仿真结果

非侵入式负荷分解评价指标有很多,本文选取了6种评价指标:准确度、精确率、召回率、F分数、均方根误差以及平均功率失真率。其计算方法如下:

$$P_1 = \frac{QP}{N} \quad (9)$$

$$P_2 = \frac{TP}{TP + FP} \quad (10)$$

$$P_3 = \frac{TP}{TP + EP} \quad (11)$$

$$F = \frac{2P_2 \times P_3}{P_2 + P_3} \quad (12)$$

$$RMSE = \left\{ \left[\frac{(x_1 - x)^2 + (x_2 - x)^2 + \dots}{N} \right]^{0.5} \right. \quad (13)$$

$$R = \frac{|P_5 - P_6|}{P_6} \quad (14)$$

式中 QP 为模型分解结果与实际结果相同的采样点数量; N 为测试采样点总量; TP 表示用电设备实际处于工作状态且分解结果也处于工作状态的点数量; FP 表示用电设备实际处于工作状态但分解结果处于非工作状态的点数量; EP 表示用电设备实际处于非工作状态但模型分解结果处于工作状态的点数量; $x_1 \dots x_n$ 表示模型分解后拟合功率; x 表示实际功率。 P_5 表示单个用电设备分解后拟合总功率, P_6 表示单个用电设备实际总功率。

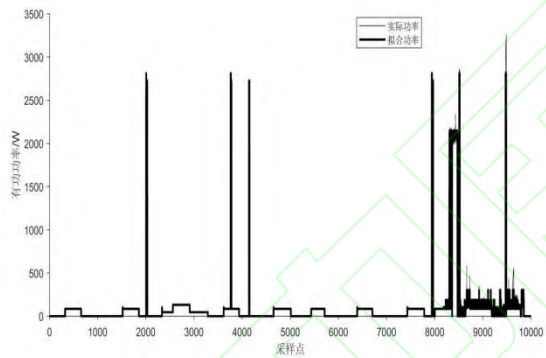
P_1 、 P_2 、 P_3 、 F 均为非侵入式负荷分解的基本

评价指标,已被大多研究者研究到比较高的水平, $RMSE$ 能够反映模型分解后拟合功率曲线的效果,其值受数值影响较大,但在比较相似数据时能够展示拟合水平高低,其值越小表明拟合效果越好。各个模型的分解结果如表4所示(表中的准确度是指4种设备共同识别正确时的准确度),整体来说深度学习

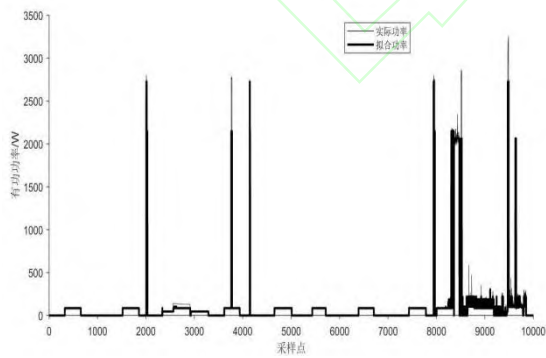
模型较常用神经网络模型有更高的准确度和更低的均方根误差,但耗时较高,深度学习模型中深层 CNN 模型表现最为出色,但是会消耗更多的时间, Bi-LSTM 模型表现略好于 LSTM 模型,为了确保高准确率我们愿意花时间以训练模型,由于常用神经网络模型的准确度和 RMSE 值不理想,所以后文不再对其进行讨论。将各个模型分解之后的结果结合表 2 中的簇中心值对总有功功率进行重构拟合,拟合效果如图5所示,与前文所得结果一致,1-D-DCNN 模型的拟合效果最佳, Bi-LSTM 模型次之, LSTM 效果一般。

表 4 各模型分解结果及耗时

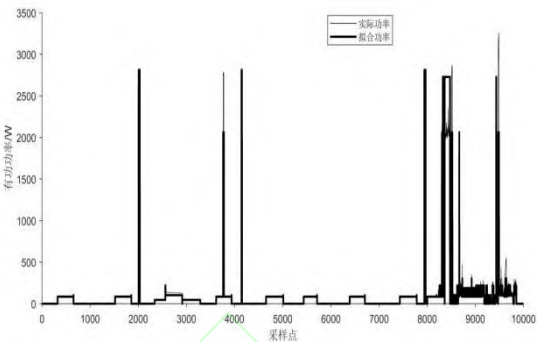
Tab.4 Decomposition results and time consumption of each model					
评价	深度学习模型			常用神经网络模型	
指标	LSTM	Bi-LSTM	1-D-DCNN	GRNN	BP
P1	88.43%	88.83%	93.79%	79.41%	70.04%
RMSE	105.795 9	100.527 8	24.904 7	746.367	904.703 8
时间	28分7秒	46分31秒	144分24秒	耗时非 常少	4分2秒



(a)1-D-DCNN 模型分解拟合效果图



(b)Bi-LSTM 模型分解拟合效果图



(c)LSTM 模型分解拟合效果图

图 5 各深度学习模型分解效果对比图

Fig.5 Comparison chart of decomposition effects of various deep learning models

单从拟合效果以及准确度还无法说明模型性能优劣,表 5 展示了每种模型将总有功功率分解后得到每个用电设备状态的准确度。由于部分用电设备各状态功率相似且设备运行存在多个用电设备同时运行等情况,所以有些设备状态类型没有被很好地辨识出来,就各个设备状态辨识效果来看深层 CNN 模型辨识效果更好,没有未辨识出来的状态, Bi-LSTM 效果次之,有 2 个状态未辨识出来, LSTM 效果最差有 4 个状态未辨识出来。由于 1-D-DCNN 模型在多个方面都表现优秀,对其进行更为细致的分析。1-D-DCNN 模型分解结果各项评价指标如表 6 所示,将该模型的分解结果进行重构拟合。图 6 展示了 1-D-DCNN 模型分解各用电设备运行状态效果。洗衣机、水壶各方面指标表现良好,冰箱、电视召回率稍低即出现了一些不该出现的设备状态,电视的平均功率失真率较高,表明功率较小的设备不易识别。

表 5 各深度学习模型设备状态辨识准确度

Tab.5 Accuracy of equipment status identification of each deep learning model

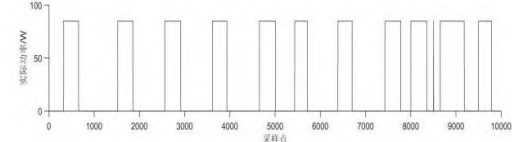
模型种类	各设备运行状态辨识准确度(%)											
	1	2	3	4	5	6	7	8	9	10	11	12
LSTM	91.90	99	0	98.38	79.92	84.34	10.94	62.42	65.79	0	0	0

Bi-LSTM	91.35	96.33	10.34	98.7	81.50	85.54	86.72	45.71	5.26	100	0	0
1-D-DCNN	93.55	99.67	93.1	98.87	73.15	24.82	62.50	69.89	94.74	93.62	99.72	23.33

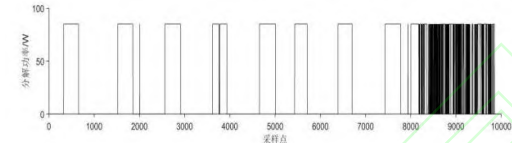
表 6 1-D-DCNN 模型分解各评价指标数据

Tab.6 Data of each evaluation index after decomposition of the 1-D-DCNN model

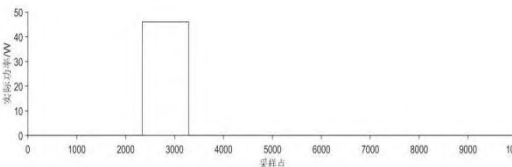
各项评价	用电设备			
	冰箱	洗衣机	电视	水壶
准确度 P_1	95.31%	96.28%	98.3%	99.99%
精确率 P_2	97.62%	97.67%	99.97%	99.99%
召回率 P_3	97.57%	98.55%	98.33%	100%
F 分数	97.60%	98.1%	99.14%	99.99%
RMSE	18.6981	27.6717	6.0248	32.8317
平均功率				
失真率	0.65%	3.09%	15.5%	3.77%



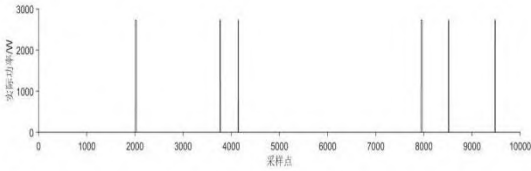
(a)冰箱



(b)洗衣机



(c)电视



(d)水壶

图 6 深层卷积神经网络分解结果与实际结果对比图

Fig.6 Comparison graph between the decomposition results of deep convolutional neural network and actual results

从单个用电设备的分解效果来看，该模型下较好的非侵入式负荷辨识效果，虽然冰箱和电视的分解结果中出现了一些辨识错误的点位，这些点位可能处于多设备同时运行状态，由于当多设备运行时会造成总功率相似，但是其他时间段分解效果很好。总体来说本文所提出的模型分解性能高于 LSTM 和 Bi-LSTM 以及一般神经网络模型，具有实际运用价值。

4 结束语

充分挖掘用电设备运行特征可以为用户和供电公司带来便利，本文转换了传统 CNN 模型输入数据维度以及改变其深度建立了新的深度学习模型，与 seq2seq LSTM 以及 seq2seq Bi-LSTM 模型相比，seq2sqe 1-D-DCNN 模型具有更高的辨识准确率。另外，在建立特征集时本文在传统 k-means 算法的基础上运用 k-means++算法初始化簇中心，利用重复迭代聚类操作避免局部最小，并提出轮廓值指标以确定聚类数目，经验证本文所提 KM++-1-D-DCNN 方法具有较好的 NILD 效果。由于输入数据仅为总有功率，所以对硬件设备要求不会过高，具有一定的实用性。未来，还能进一步提升模型泛化性以及数据集扩容至云数据，计算更大的数据集提升辨识精度，同时随着智能电表的发展，作为输入因素的数据也可多维化。

参考文献

- [1] HART G W. Nonintrusive appliance load monitoring[J]. Proceedings of the IEEE, 1992, 280(12): 1870-1891.
- [2] 田正其, 徐晴, 李如意, 等. 一种融合时间特征的非侵入式负荷辨识决策方法[J/OL]. 电测与仪表: 1-9[2020-08-26]. <http://kns.cnki.net/kcms/detail/23.1202.TH.20200803.1424.016.html>.
- Tian Zhengqi, Xu Qing, Li Ruyi, et al. A non-intrusive load identification and decision-making method integrating time characteristics[J/OL]. Electrical measurement and instrumentation: 1-9 [2020-08-26]. <http://kns.cnki.net/kcms/detail/23.1202.TH.20200803.1424.016.html>.
- [3] 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.
- [4] 马雷. 非侵入式居民电力负荷监测方法研究[D]. 东南大学, 2018.
- Ma Lei. Research on Non-intrusive Resident Electric Load Monitoring Method[D]. Southeast University, 2018.
- [5] 高浩瀚, 张利, 梁军, 等. 基于改进排列熵算法和 Yamamoto 算法的非侵入式用电设备状态变化检测[J]. 电力自动化设备, 2020, 40(1): 192-197.
- Gao Haohan, Zhang Li, Liang Jun, et al. Non-intrusive electrical equipment state change detection based on improved permutation entropy algorithm and Yamamoto algorithm[J]. Electric Power Automation Equipment, 2020, 40(1): 192-197.
- [6] 张啸宇. 非侵入式负荷分解方法研究[D]. 山东大学, 2019.
- Zhang Xiaoyu. Research on Non-intrusive Load Decomposition Method [D]. Shandong University, 2019.
- [7] Hinton G, Deng Li, Yu Dong, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups[J]. IEEE Signal Processing Magazine, 2012, 29(6): 82-97.
- [8] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016.
- [9] 王轲, 钟海旺, 余南鹏, 等. 基于 seq2seq 和 Attention 机制的居民用户非侵入式负荷分解[J]. 中国电机工程学报, 2019, 39(1): 75-83, 322.
- Wang Ke, Zhong Haiwang, Yu Nanpeng, et al. Non-intrusive load decomposition of residential users based on seq2seq and Attention mechanism[J]. Chinese Journal of Electrical Engineering, 2019, 39(1): 75-83, 322.
- [10] 张玉天, 邓春宇, 刘沅昆, 等. 基于卷积神经网络的非侵入负荷辨识算法[J/OL]. 电网技术: 1-7[2020-04-23]. <https://doi.org/10.13335/j.1000-3673.pst.2019.0653>.
- Zhang Yutian, Deng Chunyu, Liu Yuankun, et al. Non-intrusive load identification algorithm based on convolutional neural network[J/OL]. Power grid technology: 1-7[2020-04-23]. <https://doi.org/10.13335/j.1000-3673.pst.2019.0653>.
- [11] 余登武, 刘敏. 基于深度循环卷积模型的非侵入式负荷分解方法[J/OL]. 电测与仪表: 1-9[2020-08-26]. <http://kns.cnki.net/kcms/detail/23.1202.TH.20200604.1036.004.html>.
- Yu Dengwu, Liu Min. Non-intrusive load decomposition method based on deep cyclic convolution model[J/OL]. Electrical measurement and instrumentation: 1-9 [2020-08-26]. <http://kns.cnki.net/kcms/detail/23.1202.TH.20200604.1036.004.html>.
- [12] 汪繁荣, 向堃, 刘辉. 基于改进 AP 聚类与优化 GRNN 的非侵入式负荷分解研究[J]. 工程科学与技术, 2020, 52(4): 56-65.
- Wang Fanrong, Xiang Kun, Liu Hui. Research on Non-intrusive Load Decomposition Based on Improved AP Clustering and Optimized GRNN[J]. Advanced Engineering Sciences, 2020, 52(4): 56-65.
- [13] Lloyd S P. Least Squares Quantization in PCM's[J]. IEEE Transactions on Information Theory, 1982, 28(2): 129-136.
- [14] David Arthur, Sergei Vassilvitskii. K-Means++: The Advantages of Careful Seeding[C]//Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007, New Orleans, Louisiana, USA, January, 7-9, 2007. ACM, 2007.
- [15] Murray D, Stankovic L, Stankovic V. An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study[J]. Scientific Data, 2017, 4: 160122.
- [16] 杨国亮, 曾建尤, 王志元. 基于深层卷积神经网络的图像美感度分类[J]. 中北大学学报: 自然科学版, 2018, 39(4): 101-107.
- Yang Guoliang, Zeng Jianyou, Wang Zhiyuan. Image aesthetic classification based on deep convolutional neural network[J]. Journal of North University of China: Natural Science Edition, 2018, 39(4): 101-107.
- [17] Ng A Y. Feature selection, L1 vs. L2 regularization, and rotational invariance[C]//Proceedings of the Twenty-first International Conference on Machine Learning. Banff, Alberta, Canada: ACM, 2004.
- [18] 王晖南, 薛建立, 刘佳易. 基于智能电表数据的非侵入式负载监测[J/OL]. 电测与仪表: 1-6[2020-08-26]. <http://kns.cnki.net/kcms/detail/23.1202.TH.20200407.1439.008.html>.
- Wang Huinan, Xue Jianli, Liu Jiayi. Non-intrusive load monitoring based on smart electric energy meter data [J/OL]. Electrical measurement and instrumentation: 1-6 [2020-08-26]. <http://kns.cnki.net/kcms/detail/23.1202.TH.20200407.1439.008.html>.
- [19] 何恒靖, 王灏, 肖勇, 等. 基于 Bi-LSTM 算法的非侵入式负荷监测模

型[J]. 南方电网技术, 2019, 13(2): 20-26.

He Hengjing, Wang Hao, Xiao Yong, et al. Non-intrusive load monitoring model based on Bi-LSTM algorithm[J]. Southern Power System Technology, 2019, 13(2): 20-26

作者简介:



汪繁荣（1989—），男，博士，硕士生导师，讲师，从事电网控制与智能制造研究。Email: wfr@whu.edu.cn

向堃（1996—），男，通信作者，硕士研究生，从事电网控制与智能制造、非侵入式负荷监测与分解。Email: 358674729@qq.com

吴铁洲（1966—），男，博士，教授，从事储能控制及能耗管理、智慧用电系统。

收稿日期: 2020-08-25; 修回日期: 2020-10-13

(杜景飞 编发)

