Journal of Electric Power Science and Technology

Volume 38 | Issue 1 Article 26

4-10-2023

Multi-class electricity theft detection based on the CNN-LSTM hybrid model

Jinjin LI
Measurement Center of Guangxi Power Grid Co.,Ltd.,Nanning 530023,China

Jueyu CHEN
Measurement Center of Guangxi Power Grid Co.,Ltd.,Nanning 530023,China, jueyuchen@qq.com

Keying HUANG
Measurement Center of Guangxi Power Grid Co.,Ltd.,Nanning 530023,China

Recommended Citation
LI, Jinjin; CHEN, Jueyu; and HUANG, Keying (2023) "Multi-class electricity theft detection based on the CNN-LSTM hybrid model," *Journal of Electric Power Science and Technology*: Vol. 38: Iss. 1, Article 26.
DOI: 10.19781/j.issn.1673-9140.2023.01.026
Available at: https://jepst.researchcommons.org/journal/vol38/iss1/26

基于CNN-LSTM混合模型的多类别窃电行为检测

李金瑾,陈珏羽,黄柯颖

(广西电网有限责任公司计量中心,广西 南宁 530023)

摘　要:针对复杂电网环境下窃电行为难以准确检测的问题,提出一种基于CNN-LSTM混合模型的多类别窃电行为检测方法。首先基于卷积神经网络(CNN)良好的特征抽象能力提取一维用电数据的非周期性的局部特征,通过长短时记忆网络(LSTM)捕捉每日电能消耗数据间的相关性,提取周期性的用电特征建立特征融合层网络,再将CNN与LSTM提取的特征向量横向拼接获得新的融合向量,据此实现多类别窃电行为的准确检测。实验结果表明,本文提出方法能准确识别多类别窃电行为,相比现有检测方法检测结果更加全面准确。

关　键　词:窃电检测;多类别;卷积神经网络;长短时记忆网络;特征融合

DOI:10.19781/j.issn.1673-9140.2023.01.026　　**中图分类号**:TM73　　**文章编号**:1673-9140(2023)01-0226-09

Multi-class electricity theft detection based on the CNN-LSTM hybrid model

LI Jinjin, CHEN Jueyu, HUANG Keying

(Measurement Center of Guangxi Power Grid Co., Ltd., Nanning 530023, China)

Abstract: This paper addresses the difficulty of the accurately detecting electricity theft in complex grid environment and proposes a multi-category electricity theft detection method based on CNN-LSTM hybrid model. Firstly, the excellent feature abstraction ability of convolutional neural networks (CNN) is utilized to extract the non-periodic local features of one-dimensional electricity consumption data. Then, the long short-term memory (LSTM) is adopted to capture the correlation between daily power consumption data and extract periodic power consumption features to establish feature fusion layer network. After that, the feature vectors extracted by CNN and LSTM are horizontally splicing to obtain a new fusion vector. Based on this, the accurate detection of multiple types of electric theft behavior are realized. Experimental results show that the proposed method can accurately identify multiple types of electric theft behavior, and the detection results are more comprehensive and accurate than the existing detection methods.

Key words: electricity theft detection; multi-class; convolutional neural network; long short-term memory network; terminals

中国社会经济的快速发展与智能电网的高速建设密不可分,随着用电需求的增加,窃电手段层出不穷,给电力部门带来巨大的经济损失[1-2]。当前窃电稽查大都基于人工现场稽查,消耗大量的人力、物力与财力,而基于数据驱动研究为从海量数据中检测窃电用户提供新的前景[3],但现有窃电检

收稿日期:2021-11-08;**修回日期**:2021-12-27

基金项目:广西电网有限责任公司科技项目(GXKJXM20200020);国家自然科学基金(51777061)

通信作者:陈珏羽(1992—),女,硕士,主要从事电能计量研究;E-mail: jueyuchen@qq.com

测算法往往仅检测是否发生窃电行为，无法实现窃电行为的检测。而准确获取窃电行为的具体类别，对实际电网运行可针对性实现各类型窃电用户快速准确稽查[4]。因此，准确的多类别窃电行为检测技术的研究对于智能电网的安全稳定运行具有重大价值与意义[5]。

伴随智能电表与用电信息采集系统的普及，为基于数据驱动的窃电检测技术研究积累了大量计量数据。文献[6]提出基于支持向量机(support vector machine，SVM)的窃电检测技术，相比人工稽查，将窃电检测成功率由 3% 提高至 60%；文献[7]提出稀疏随机森林模型的用电异常检测方法，通过用户异常度累积的方式，有效降低误检率；文献[8]提出基于 XGBoost 窃电检测方法，并通过遗传算法进行特征筛选，有效提高窃电检测精度。然而上述窃电检测技术选取的窃电特征都是基于窃电行为的先验知识设计，不具有自适应性。

为实现窃电嫌疑的自适应检测，文献[9-10]提出基于卷积神经网络(convolutional neural networks，CNN)的窃电检测技术；文献[11]提出基于深度循环神经网络的异常用电检测方法；文献[12]提出密集型卷积神经网络(dense net，DN)和随机森林(RF)相融合的窃电检测方法。上述方法可有效实现用户用电异常检测，但并未对窃电行为进行进一步检测，无法对多类别窃电行为进行辨识。

为实现多类别窃电行为的准确检测，本文首先建立多类别窃电的行为模型，通过 CNN 挖掘用电数据的局部特征，同时基于长短时记忆(long short-term memory，LSTM)神经网络提取用电时序特征，建立时序特征与局部特征的融合层网络，构建多类别窃电行为特征数据集，以降低误检率、漏检率，提高多类别窃电行为检测准确率，据此提出基于 CNN-LSTM 混合模型的多类别窃电行为检测，最后通过实验分析验证本文提出方法的有效性与准确性。

1　多类别窃电行为模型

1.1　多类别窃电行为建模

窃电用户的用电并非随机用电，其目的是通过某些手段来减少电表所记录的电量，从而少缴电费。由于窃电用户自身的目的性，其用电行为同样具有一定用电规律，所以可根据此规律对窃电行为进行建模。综合当前广泛使用的多种虚假数据注入模型来模拟窃电行为[13]，本文使用 7 种模型来模拟多种窃电行为，如表 1 所示。

表 1　窃电样本构造公式

Table 1　Formula for constructing electric stealing samples

窃电类型	构造算式
1	$f_1(x_t^d)=ax_t^d, 0.1<a<0.8$
2	$f_2(x_t^d)=a_t^dx_t^d, 0.1<a_t^d<0.8$
3	$f_3(x_t^d)=\max\{x_t^d-\gamma,0\}, \gamma<\max(x_t^d)$
4	$f_4(x_t^d)=0$
5	$f_5(x_t^d)=\mathrm{mean}\{x_{t=1,2,\cdots,n}^d\}$
6	$f_6(x_t^d)=\beta\mathrm{mean}\{x_{t=1,2,\cdots,n}^d\}, 0.1<\beta<0.8$
7	$f_7(x_t^d)=x_{t=n,n-1,\cdots,1}^d$

表 1 中，类型 1 表示按照固定比例缩小计量电量；类型 2 表示连续按时变比例随机减少用电量；类型 3 表示根据阈值对实际用电负荷进行削减；类型 4 表示全时段上传零电量；类型 5 表示以用电量均值进行上传；类型 6 表示在用电量均值的基础上按固定比例缩小计量电量；类型 7 表示通过颠倒用电时序降低电费成本。不同类型窃电行为具有不同的用电特征，将某正常用户进行多种类型数据篡改模拟窃电的示例如图 1 所示。

由图 1 可知，由于窃电行为的多样性，各窃电用户的负荷曲线同样呈多样变化，因此实现多类别窃电行为检测的难度较大，需将用电数据中不同类型的特征进行融合以提高多类别检测的准确率。

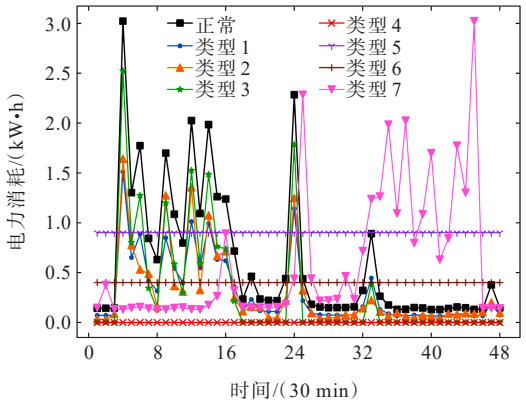


图 1　不同类别窃电行为示例

Figure 1　Example of the different types of electricity theft

1.2 评价指标

在多分类任务中，可用混淆矩阵来比较分类结果与实际测得值，从而直观地表示各类别的分类状态，多分类混淆矩阵如图2所示。

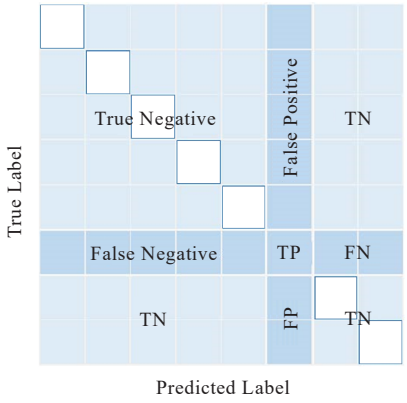


图2 多分类混淆矩阵

Figure 2 Multicategory confusion matrix

图2中，混淆矩阵列代表预测值为某一个类别，行代表真实标签为某一个类别。对于某类别而言，可将用户按真实标签与预测标签分为真阳性(true positive，TP)、假阳性(false positive，FP)、真阴性(true negative，TN)与假阴性(false negative，FN)。T_P表示将此类别用户正确地预测出来；F_P表示将非此类别的用户错误地预测为此类别；T_N表示将非此类别的用户正确地预测为非此类别；F_N表示将此类别用户错误地预测为非此类别。

以混淆矩阵为基础，可得分类器的多个评价指标，二分类中常用指标包括准确率(accuracy，ACC)、精确度(precision，PRE)、查全率(Recall)和F1分数(F1-Score，F1)。其中，A_{CC}为所预测样本中预测正确的比例；P_{RE}为所有预测为某类别的用户中实际为此类用户的占比；R_{ecall}为所有实际为某类别的用户中预测为此类用户的占比；F_1为与P_{RE}和R_{ecall}有关的综合指标[14]。各指标计算式分别如下：

$$A_{CC}=\frac{T_P+F_N}{T_P+T_N+F_P+F_N} \tag{1}$$

$$P_{RE}=\frac{T_P}{T_P+F_P} \tag{2}$$

$$R_{ecall}=\frac{T_P}{T_P+F_N} \tag{3}$$

$$F_1=\frac{2P_{RE}\cdot R_{ecall}}{P_{RE}+R_{ecall}} \tag{4}$$

式(1)～(4)中，T_P、F_P、T_N与F_N分别为分类结果的真阳性、假阳性、真阴性与假阴性的用户数量。

在计算多分类的评价指标时，需将n分类拆分为n个二分类进行计算，即可得到各个类别的评价指标结果。将各个类别的结果进行平均可得分类任务的整体指标，其中平均方式有2种，即宏平均(macro average)与微平均(micro average)。宏平均是将n个二分类的各指标结果直接求算数平均值，微平均则为n个二分类结果的T_P、F_P、T_N与F_N对应相加，再根据各指标算式进行计算。宏平均在计算的过程中不考虑各个类别的样本比例，因此不适合数据集不平衡的情况，所以本文采用微平均作为多分类结果的计算方式，使用A_{micro}、P_{micro}、R_{micro}、F_{1micro}分别表示多分类结果各类别准确率、精确度、召回率与F1分数的平均值，各指标计算式与二分类时A_{CC}、P_{RE}、R_{ecall}和F计算式形式相同，但变量含义不同，需将分类结果中每个类别的T_P、F_P、T_N与F_N样本数相加再进行计算。

由于实际电网中窃电用户所占比例较小，即窃电行为检测存在数据不平衡问题。因此，可用受试者工作特征曲线(receiver operating characteristic curve，ROC)评价分类器的性能，曲线每个点代表不同阈值下的分类效果，曲线越陡峭，即ROC曲线与坐标轴围成的面积越大，分类效果越好，曲线下面积被定义为Area Under Curve(AUC)。

2 CNN-LSTM模型窃电行为检测

2.1 卷积神经网络

用户用电数据时间序列特征的准确提取是实现多类别窃电行为识别的关键环节，CNN具有良好特征提取能力，其由输入层、卷积层(convolutional layer)、池化层(pooling layer)、全连接层(fully connected layer)及输出层组成[15]。卷积神经网络拥有表征学习能力，其对输入数据能够按其网络结构层层学习，且基于卷积神经网络提取特征效果明显以及对数据没有额外的特征工程要求，因此本文采用卷积神经网络对用户用电数据特征自适应提取。卷积神经网络模型的框架如图3所示。

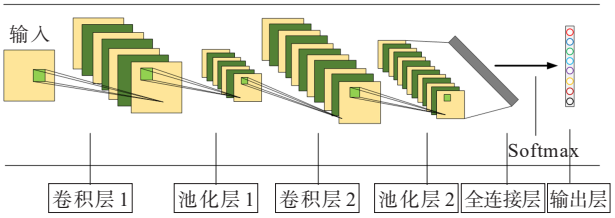


图 3　卷积神经网络模型框架

Figure 3　Architecture of the CNN model

图 3 中，卷积层为 CNN 的核心组成模块，由一组平行特征图（feature map）组成，通过卷积核（convolution kernel）对输入特征图进行卷积运算得到输出特征图，该特征图中所有元素均通过同一卷积核计算，即权值和偏置项共享。卷积运算如下：

$$\boldsymbol{x}_j^r = f\left(\sum \boldsymbol{x}_i^{r-1} \cdot \boldsymbol{k}_{i,j}^r + \boldsymbol{b}_j^r\right) \tag{5}$$

式中，$\boldsymbol{x}_j^r$ 为通过第 r 层卷积运算所得第 j 个输出特征图；$\boldsymbol{b}_j^r$ 为第 r 层网络第 j 个卷积核的偏置向量；$\boldsymbol{k}_{i,j}^r$ 为与第 i 个输入特征图运算的第 j 个卷积核；f 为非线性激活函数，以提高网络的拟合能力与稀疏性，在激活函数选择上，ReLU 函数相比 Sigmoid 函数有防止梯度弥散和计算速度快等优点，故模型中选择 ReLU 函数作为激活函数，其表达式为

$$R_{\mathrm{eLU}}(x) = \begin{cases} x, x > 0 \\ 0, x \geqslant 0 \end{cases} \tag{6}$$

式中，x 为卷积运算后得到的数据。

在卷积层之后接池化层，用于缩小模型大小，提高计算速度，同时提高所提取特征的鲁棒性，在减少冗余特征量的同时，保留用电行为主要特征，通过减少计算参量以达到降维效果，防止过拟合现象，可提高模型泛化能力。池化操作实际为一种下采样操作，其操作包括最大池化、均值池化、随机池化等。池化操作计算式为

$$p(i,j) = \frac{1}{w^2} \sum_{u=(i-1)w+1}^{iw} \sum_{v=(j-1)w+1}^{jw} a(u,v) \tag{7}$$

式中，$a(u,v)$ 为池化层输入矩阵中第 u 行 v 列的值；$p(i,j)$ 为池化层输出矩阵第 i 行 j 列的值；w 为参与集合区域的边值。

在最后一个池化层后接全连接层，将所有神经元进行全连接操作，可将所学分布式特征映射到样本标记空间，其模型可表示为

$$\boldsymbol{y} = \boldsymbol{w}\boldsymbol{x} + \boldsymbol{b} \tag{8}$$

式中，$\boldsymbol{x}$ 为全连接层的输入；$\boldsymbol{w}$ 为权值矩阵；$\boldsymbol{b}$ 为偏置向量。

2.2　长短时记忆神经网络

相比 CNN 良好局部特征的提取能力，循环神经网络（recurrent netural network，RNN）更能关注用户负荷数据中的时序特征[16]。将按日进行重构后的用电数据输入到 RNN 中，即可捕捉到每日电能消费数据间的相关性，进而提取到周期性的用电特征。LSTM 作为改进的循环神经网络，其通过特殊的“门”的内部机制来解决 RNN 输入长序列进行训练的过程中出现的梯度消失和梯度爆炸的问题，LSTM 单元的结构如图 4 所示。

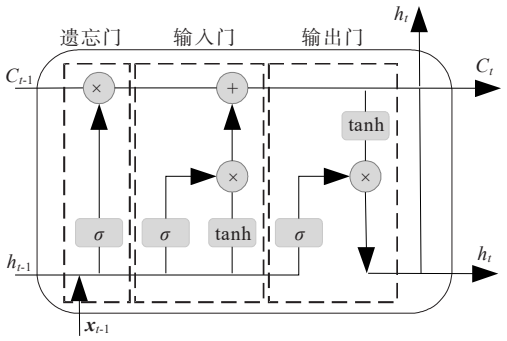


图 4　LSTM 单元结构

Figure 4　The architecture of LSTM cell

图 4 中，LSTM 的“门”包括 3 种类型，即遗忘门、输入门和输出门。遗忘门决定应丢弃或保留哪些信息，来自前一个隐藏状态的信息和当前输入信息同时传递至 sigmoid 函数中进行计算，输出在 0 至 1 之间的遗忘向量，表示信息被遗忘的程度，其计算式为

$$\boldsymbol{f}_t = \sigma(W_{\mathrm{f}}\boldsymbol{x}_t + U_{\mathrm{f}}h_{t-1} + b_{\mathrm{f}}) \tag{9}$$

式中，$\boldsymbol{x}_t$ 为 t 时间步时的输入特征向量；f_t、h_t 分别为在 t 时间步时经过遗忘门的激活值和隐藏层状态的值；W_{f}、U_{f} 分别为输入与隐藏层状态的权重；b_{f} 为各偏置值；$\sigma(x) = (1+\mathrm{e}^{-x})^{-1}$ 为 sigmoid 激活函数。

输入门用于更新神经元状态，首先将前一层隐藏状态的信息和当前输入的信息传递到 sigmoid 函数中进行计算可得：

$$i_t = \sigma(W_{\mathrm{i}}\boldsymbol{x}_t + U_{\mathrm{i}}h_{t-1} + b_{\mathrm{i}}) \tag{10}$$

式中，i_t 为在 t 时间步时经过输入门的激活值。

再将前一层隐藏状态的信息和当前输入的信息传递到 tanh 函数中进行计算，有

$$m_t = \tanh(W_{\mathrm{c}}\boldsymbol{x}_t + U_{\mathrm{c}}h_{t-1} + b_{\mathrm{c}}) \tag{11}$$

式中，$\tanh(x)=(e^{2x}-1)(e^{2x}+1)^{-1}$为双曲正切函数。

将式(10)、(11)得到的两向量点乘，再与式(9)得到的遗忘向量相加，即可得到更新后的神经元状态，即

$$C_t=f_t\cdot C_{t-1}+i_t\cdot nm_t \tag{12}$$

式中，C_t为隐藏状态的值；·为向量的点积。

输出门用来确定下一个隐藏状态的值。首先将当前的输入传递到sigmoid函数中进行计算得到输出为

$$o_t=\sigma(W_o x_t+U_o h_{t-1}+b_o) \tag{13}$$

将式(12)得到的神经元状态传递给tanh函数，将其输出与o_t点乘，即可确定隐藏状态所应携带的信息，其计算式为

$$h_t=o_t\cdot\tanh(C_t) \tag{14}$$

LSTM单元通过上述3个“门”的操作即可得到新的神经元状态和新的隐藏状态，再传递至下一时间步长。在传递完所有的时间步长之后，各LSTM单元隐藏状态值即为LSTM网络所提取到的特征。

2.3 CNN-LSTM模型构建与窃电行为检测

为实现多类别窃电行为准确检测，本文通过CNN提取用电数据非周期性特征，LSTM提取用电数据周期性特征，并构建特征融合层，将2个网络提取特征向量横向拼接得到新的融合向量，融合向量同时包含有非周期性局部特征与周期性时序特征，以此增强窃电行为特征敏感度，最后采用softmax激活函数得到不同窃电类型的概率输出，由概率大小得到模型预测的用户类别，实现多类别窃电行为检测，据此建立基于CNN-LSTM混合模型，如图5所示。

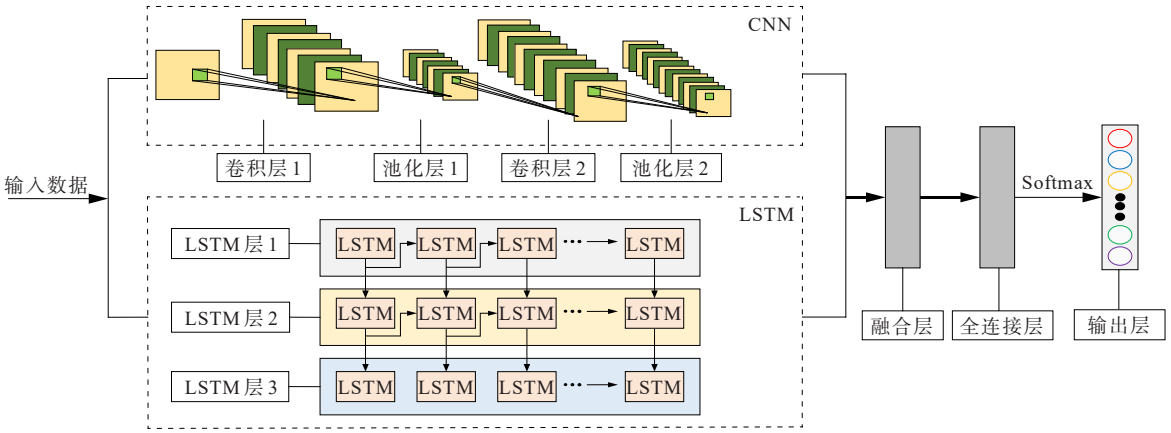


图5 CNN-LSTM窃电行为检测模型结构

Figure 5 CNN-LSTM structure of electric theft behavior detection

基于CNN-LSTM混合模型建立多类别窃电行为检测的算法流程图如图6所示，其主要步骤如下。

1) 将输入的电能用户用电数据分别转化为一维和二维时间序列，其中二维序列以天为周期，得到深度为15，宽度为48的二维用户用电数据。

2) 采用CNN网络提取一维用户用电数据的局部特征，得到n个特征向量，然后进行全局池化得到一个特征向量，再采用LSTM提取二维用户用电数据的时序特征，得到一个特征向量。

3) 将步骤2)中得到的2个不同类型的特征向量进行横向拼接，即可得到一个包含有非周期性局部特征与周期性时序特征的特征向量，即为融合特征向量。

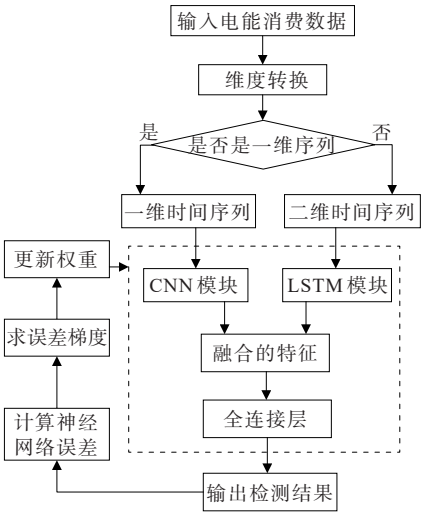


图6 CNN-LSTM算法流程

Figure 6 CNN-LSTM algorithm flow chart

4）将融合特征输入到全连接层中，并采用 softmax 激活函数得到每种用户类型的概率输出，概率最大的类别即为模型预测的用户类型，由此即可实现多类别窃电行为检测。

本文基于 Tensorflow 深度学习框架实现提出的 CNN-LSTM 混合模型，具体的网络参数设置如表 2 所示。

表 2　CNN-LSTM 模型参数设置

Table 2　CNN-LSTM model parameter setting

参数名称	参数值
卷积核大小	3*1
卷积步长	1
卷积输出层激活函数	Relu
池化层的步长	2
LSTM 层数	3
LSTM 神经元个数	32

3　实验与分析

3.1　实验数据与预处理

本文以 Irish Smart Energy Trail（ISET）发布的爱尔兰智能用电实验的实际用电数据为数据集，其包含爱尔兰 5 000 个居民和商业用户 535 d 的用电数据，以 30 min 为采样间隔，各用户每天采集 48 个数据点[17]，选择其中 3 000 个居民用户 15 d 负荷数据进行研究，由于所有用户同意将其用电数据研究使用，故假设该数据集用户为正常用户，随机选择正常用电数据按表 1 修改后作为窃电样本。

用电采集系统在采集数据过程中由于各种偶然因素会造成部分数据缺失，所以需对原始数据进行插值处理。针对缺失数据，若某用户连续缺失数据超过 7 d，将该用户缺失用电数据用 0 填补，否则采用三次多项式插值进行修正，算式为

$$L_3(t_k)=y_{k-2}l_{k-2}(t_k)+y_{k-1}l_{k-1}(t_k)+y_{k+1}l_{k+1}(t_k)+y_{k+2}l_{k+2}(t_k) \tag{15}$$

式中，t_k 为第 k 个数据点所在时刻，当其值出现缺失时，取待插值时刻相邻各 2 个数据点 $\{t_{k-2},t_{k-1},t_{k+1},t_{k+2}\}$ 对 t_k 时刻进行三次多项式插值；y 为对应时刻的函数值；l 为对应的插值基函数。

3.2　CNN-LSTM 窃电行为检测结果分析

本文将窃电样本与正常样本混合组成数据集，并将其按 3∶1∶1 将数据集分为训练集、测试集以及验证集，利用本文提出方法对训练集进行训练，各窃电类型检测的混淆矩阵如图 7 所示。

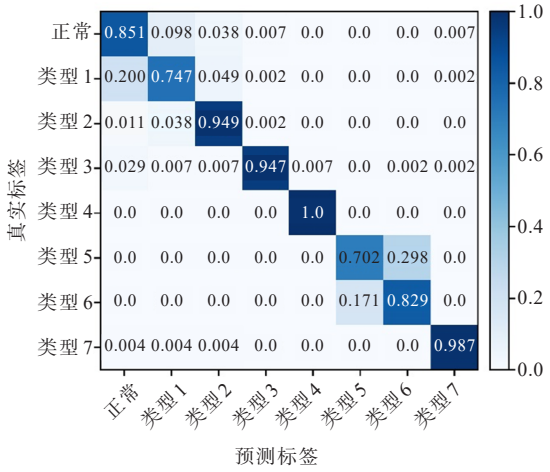


图 7　各窃电类型检测混淆矩阵

Figure 7　The confusion matrix of each type of electric theft detection

由图 7 可知，本文提出方法可准确识别 85.1% 的正常用户，对窃电类型 1、5、6 的识别准确率相对较低，对其他窃电类型的识别准确率均在 94% 以上。由图 7 左下角可知，20% 的正常用户被误识别为窃电类型 1，窃电类型 1 有 0.098 的概率被识别为正常用户。由窃电类型 1 的构造算式可知，类型 1 在正常用户用电数据的基础上乘倍数实现构造，与正常用户用电习惯较为相似，导致本文方法对两者易产生误判。由图 7 右下角可知，本文方法对窃电类型 5 和 6 有同样情况的误判，因其窃电类型 6 为在类型 5 的基础上乘以倍数构成，具有相似的用电趋势，区分困难。综上所述，无论针对用电规律差别较为显著用户样本还是难以分类样本，本文提出多类别窃电行为检测方法均具有很高的准确率。

各窃电类型检测的 ROC 曲线如图 8 所示。由图 8 可知，对于类型 2、3、4、7，A_{UC} 值均在 0.99 以上，本文提出方法具有较高的检测准确度，其他类别检测效果略低，但其 A_{UC} 仍能达到 0.97 以上。可见，对于所用类别的窃电行为，本文提出方法均能达到很高的检测准确度。

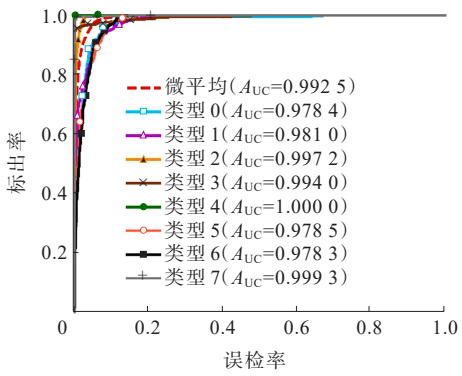


图8 各窃电类型检测ROC曲线

Figure 8 The ROC curve of each type of electric theft detection

3.3 与现有方法对比

为进一步验证本文算法的有效性,将本文所提方法与现有窃电检测方法进行对比分析,包括逻辑回归(logistic regression,LR)、梯度提升树(gradient boosting decision tree,GBDT)、CNN与LSTM。LR作为一种有监督学习的线性算法,常用于解决当前窃电检测问题;GBDT属于集成学习中Boosting算法,其通过采用加法模型,不断减小训练过程产生的残差实现样本分类;CNN通过卷积与池化操作提取负荷数据的局部特征进行用户分类,实现窃电行为检测;LSTM使用门控神经元学习时序数据的长期信息,通过提取负荷数据的全局特征来实现窃电行为检测。各算法的参数设置如表3所示,本文及现有算法的评价结果如表4所示。

表3 各算法参数

Table 3 Algorithm parameters

方法	参数
LR	C=50,采用L2正则项,t_{ol}=0.000 1
GBDT	l=0.1,$n_{estimators}$=100
CNN	与本文方法的CNN同结构
LSTM	与本文方法的LSTM同结构

表4 不同算法的评价结果

Table 4 Evaluation results of different algorithms

方法	A_{micro}/%	P_{micro}/%	F_{1micro}
LR	63.75	62.29	0.6252
GBDT	75.58	77.08	0.7533
CNN	85.72	86.20	0.8574
LSTM	72.83	75.61	0.7135
LSTM-CNN	87.56	87.91	0.8754

由表4可知,本文提出方法的A_{micro}、P_{micro}、F_{1micro}均为各检测算法中最高,对各类数据的P_{micro}为87.91%,具有较高的平均检测精确度。采用单一分类器LR的检测效果较差,其A_{micro}为63.75%,采用集成学习方法GBDT的A_{micro}为75.58%,效果较好,采用深度学习方法中CNN具有较高的A_{micro}等指标。本文方法将CNN和LSTM网络进行融合,提取负荷数据中的非周期性与周期性特征,进一步提高多类别窃电行为检测准确率。

为进一步研究和分析本文提出方法对多类别窃电行为的检测能力,与现有方法比较各类别窃电用户的检测准确率如表5所示。

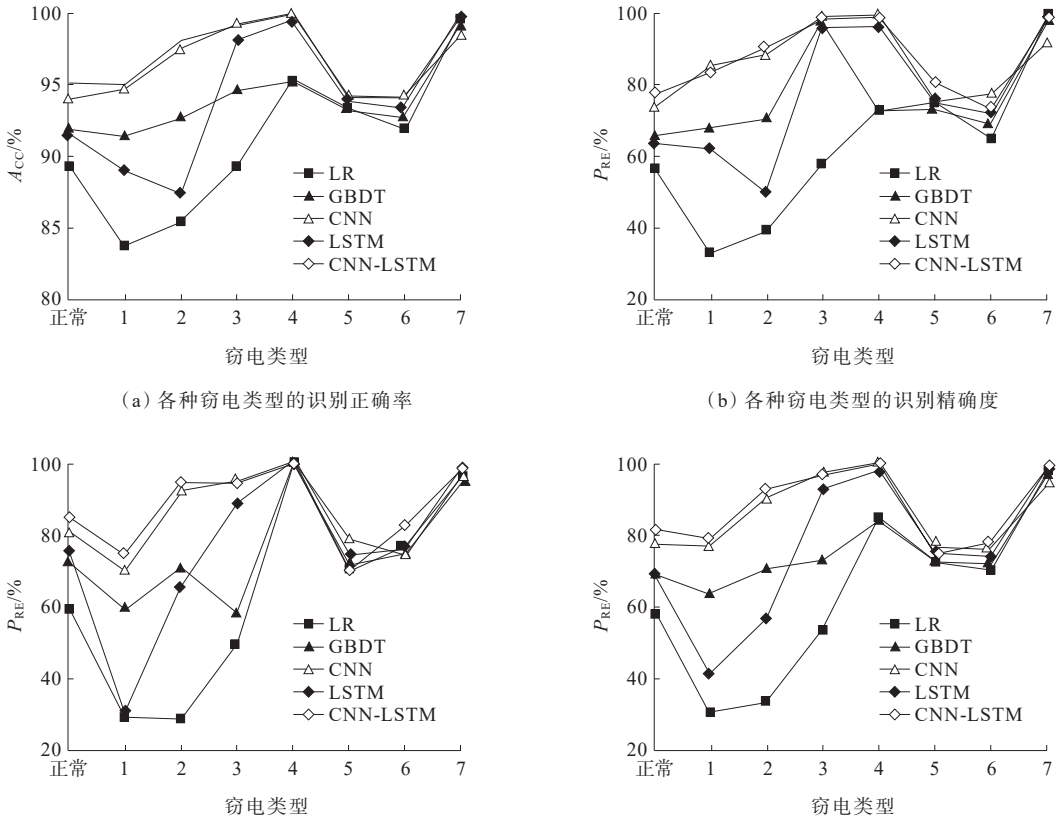
由表5可知,LR的分类效果相对较差,其余4种方法均对窃电类型1、5、6的检测效果相对较差,CNN-LSTM对窃电类型6相较于CNN和LSTM有显著提升。与LSTM对比可知,本文提出方法对窃电类型5的检测准确率略低于LSTM,对窃电类型7的检测能力与其相当,而对其他类型检测上均显著高于LSTM,与CNN的检测结果比较可知,本文提出方法在正常用户以及窃电类型1、2、6、7上的检测准确率均高于CNN,对窃电类型5上的检测效果略低,对其他窃电类型检测准确率与CNN相当。本文提出方法基于端对端的方式实现窃电行为多类别检测,融合LSTM提取的周期性特征及CNN非周期性特征,充分对数据特性分析,更好地捕捉负荷数据特征,从而提高多类别窃电行为检测方法的有效性与准确性。

表5 不同算法多类别用户的准确率比较

Table 5 Comparison of accuracy of different algorithms for multiple categories of users

窃电类型	不同算法多类别用户准确率				
	LR	GBDT	CNN	LSTM	CNN-LSTM
正常用户	0.598	0.736	0.811	0.756	0.851
1	0.289	0.596	0.700	0.302	0.747
2	0.291	0.713	0.927	0.653	0.949
3	0.496	0.582	0.951	0.891	0.947
4	1.000	1.000	1.000	1.000	1.000
5	0.700	0.718	0.787	0.744	0.702
6	0.767	0.751	0.744	0.762	0.829
7	0.960	0.951	0.964	0.987	0.987

为了进一步研究各种算法对各窃电类型的识别规律，本文将不同算法对7种窃电类型及正常用户的识别性能指标进行可视化，不同算法在多种窃电类型下的表现结果如图9所示。



（a）各种窃电类型的识别正确率　（b）各种窃电类型的识别精确度

（c）各种窃电类型的查全率　（d）各种窃电类型的F1分数值

图9　各算法多类别窃电类型检测结果比较

Figure 9　Comparison of detection results of different algorithms under different types of electric theft

由图9可知，本文提出方法在各种窃电类型的正确率、精确度、查全率和F1分数均表现最好，单一分类器LR的表现最差。从多窃电类型分析，5种分类算法对类型5～7的识别表现相近，但对类型1～4的识别表现差距较大，即本文提出方法和CNN对类型1～4的识别表现要远优于其他分类算法，表明本文提出的多类别窃电行为检测对窃电的特征提取能力更优。

4　结语

本文提出一种基于CNN-LSTM混合模型的多类别窃电行为检测方法，实验结果表明：本文建立的CNN-LSTM混合模型能有效提取一维用户负荷数据中非周期性特征与按日进行重构后的二维负荷数据的周期性特征；构建的特征融合层网络将获得特征向量横向拼接得到融合向量，同时包含了用电数据2种不同类型的特征，有效增强窃电行为特征敏感度，提升多类别窃电检测准确率。相比现有检测方法的单一判别结果，本文提出方法可输出各窃电行为的概率，识别窃电行为的多种类别，并有效提高了多类别窃电行为检测的整体准确度。

参考文献：

[1]　董立红，肖纯朗，叶鸥，等．一种基于CAEs-LSTM融合模型的窃电检测方法[J]．电力系统保护与控制，2022，50(21)：118-127.
DONG Lihong，XIAO Chunlang，YE Ou，et al．A detection method for stealing electric theft based on CAEs-LSTM fusion model[J]．Power System Protection and Control，2022，50(21)：118-127.

[2]　ZANETTI M，JAMHOUR E，PELLENZ M，et al．A tunable fraud detection system for advanced metering infrastructure using short-lived patterns[J]．IEEE Transactions on Smart

Grid,2019,10(1):830-840.

[3] 吕笃良,刘梦爽,桓露,等.基于重加权策略平衡损失与LSTM的窃电行为检测研究[J].智慧电力,2022,50(4):15-20+58.
LÜ Duliang, LIU Mengshuang, HUAN Lu, et al. Electricity stealing detection based on reweighted strategy balancing loss and LSTM[J]. Smart Power, 2022, 50(4):15-20+58.

[4] ZHENG K D, CHEN Q X, WANG Y, et al. A novel combined data-driven approach for electricity theft detection[J]. IEEE Transactions on Industrial Informatics, 2019,15(3):1809-1819.

[5] 王圆圆,白宏坤,王世谦,等.基于信息增益与Spearman相关系数的电力用户行为画像[J].电力工程技术2022,41,(4):220-228.
WANG Yuanyuan, BAI Hongkun, WANG Shiqian, et al. Power users' behavior portrait based on information gain and Spearman correlation coefficien[J]. Electric Power Engineering Technology,2022,41(4):220-228.

[6] NAGI J, YAP K S, TIONG S K, et al. Nontechnical loss detection for metered customers in power utility using support vector machines[J]. IEEE Transactions on Power Delivery,2010,25(2):1162-1171.

[7] 苏欣,田浩,秦昌龙,等.多变量数据聚类最优选择的用电关联分析算法[J].电网与清洁能源,2022,38(4):86-94+103.
SU Xin, TIAN Hao, QIN Changlong, et al. Electricity consumption association analysis algorithm for optimal selection of multivariate data clustering[J]. Power System and Clean Energy,2022,38(4):86-94+103.

[8] YAN Z Z, WEN H. Electricity theft detection base on extreme gradient boosting in AMI[J]. IEEE Transactions on Instrumentation And Measurement,2021,70:1-9.

[9] YAO D, WEN M, LIANG Z, et al. Energy theft detection with energy privacy preservation in the smart grid[J]. IEEE Internet of Things Journal,2019,6(5):7659-7669.

[10] JAVAID N, GUL H, BAIG S, et al. Using GANCNN and ERNET for detection of non technical losses to secure smart grids[J]. IEEE Access,2021,9:98679-98700.

[11] 严勤,邓高峰,胡涛,等.基于深度循环神经网络的异常用电检测方法[J].中国测试,2021,47(7):99-104.
YAN Qin, DENG Gaofeng, HU Tao, et al. Abnormal electricity detection method based on deep recurrent neural network[J]. China Measurement & TEST, 2021, 47(7):99-104.

[12] 蔡嘉辉,王琨,董康,等.基于DenseNet和随机森林的电力用户窃电检测[J].计算机应用,2021,41(S1):75-80.
CAI Jiahui, WANG Kun, DONG Kang, et al. Power user stealing detection based on DenseNet and random forest [J]. Journal of Computer Applications,2021,41(S1):75-80.

[13] 王永明,陈宇星,殷自力,等.基于大数据分析的电力用户行为画像构建方法研究[J7.高压电器,2022,58(10):173-179+187.
WANG Yongming, CHEN Yuxing, YIN Zili, et al. Research on construction method of power user behavior portrait based on big data analysis[J]. High Voltage Apparatus, 2022,58(10):173-179+187.

[14] FABIAN A N, GERARDO F, CHU C C. NTL detection in electric distribution systems using the maximal overlap discrete wavelet-packet transform and random undersampling boosting[J]. IEEE Transactions on Power Systems,2018,33(6):7171-7180.

[15] PEREIRA J, SARAIVA F. Convolutional neural network applied to detect electricity theft: a comparative study on unbalanced data handling techniques[J]. International Journal of Electrical Power & Energy Systems, 2021, 131(9):107085.

[16] BIAN J, WANG L, SCHERER R, et al. Abnormal detection of electricity consumption of user based on particle swarm optimization and long short term memory with the attention mechanism[J]. IEEE Access, 2021, 9: 47252-47265.

[17] 马宗彪,许素安,朱少斌,等.基于特征加权模糊聚类的电力负荷分类[J].中国电力,2022,55(6):25-32.
MA Zongbiao, XU Suan, ZHU Shaobin, et al. Power load classification based on feature weighted fuzzy clustering [J]. Electric Power,2022,55(6):25-32.