

基于多头注意力机制的用户窃电行为检测

肖丁 张琦璠 纪厚业

北京邮电大学计算机学院 北京 100876

(dxiao@bupt.edu.cn)

摘要 窃电对社会和经济发展造成了重大损害。如何基于电力大数据来检测用户恶意窃电行为,已受到学术界和工业界的广泛关注。针对传统方法依赖于手工特征、行为序列表征不足和检测精度差等问题,提出了一种基于多头注意力机制的窃电检测模型(Electricity Theft Detection Based on Multi-Head Attention,ETD-MHA)。该模型基于双向门控循环神经网络(Bidirectional Gated Recurrent Unit,BiGRU)充分捕获用户用电行为序列的时序特征,引入多头注意力机制来进一步增强关键特征的区分度,并通过加深网络来提高学习效果。在爱尔兰和中国国家电网智能电表数据集上进行了大量的实验,结果表明,相比传统的逻辑回归(Linear Regression,LR)、支持向量机(Support Vector Machine,SVM)、随机森林(Random Forest,RF)等多种算法,所提模型展现出了明显的优势。例如,在爱尔兰智能电表数据集上,其 AUC 值相比 LR 算法最高提升了 34.6%。

关键词: 智能电网;窃电检测;深度学习;门控循环神经网络;多头注意力机制

中图分类号 TP183

Electricity Theft Detection Based on Multi-head Attention Mechanism

XIAO Ding,ZHANG Yu-fan and JI Hou-ye

School of Computer Science,Beijing University of Posts and Telecommunications,Beijing 100876,China

Abstract Electricity theft causes significant damage to social and economic development. How to detect malicious electricity theft based on power big data has been widely concerned by academia and industry. Aiming at the problems of traditional methods relying on manual features,insufficient behavior sequence representation,poor detection accuracy,etc.,this paper proposes an electricity theft detection model based on multi-head attention mechanism (ETD-MHA).The bidirectional gated recurrent unit is used to fully capture the time features of the electricity consumption behavior sequence,and the distinction of key features is gradually enhanced in the multi-head attention mechanism,and finally,the learning effect is improved by deepening the networks.Extended experiments are conducted on the smart meter datasets of Ireland and China State Grid.The results show that the proposed method achieves better performance compared with the linear regression (LR),support vector machine (SVM),random forest (RF),and other traditional algorithms.For example,the AUC value of the proposed model is improved by up to 34.6% compared to the LR algorithm.

Keywords Smart grids,Electricity theft detection,Deep learning,Gated recurrent neural network,Multi-head attention mechanism

1 引言

随着智能电网的发展以及智能终端的普及,我国电力系统的规模在不断扩大,对提升电力公司经济效益、降低经济损失提出了更高的要求。电力公司的损失主要来自两个方面^[1],即技术损失(Technical Losses,TL)和非技术损失(Non-Technical Losses,NTL)。TL是由内部电阻或能源损耗造成的^[2],往往不可避免;而NTL通常是由用户恶意窃电行为引起的,这种不当行为导致的损失在电力公司的总损失中占据了较大比重。根据官方的统计,美国、英国和加拿大每年的窃电损失分别约为60亿美元、1.73亿美元和1亿美元。因此,

窃电检测(Electricity Theft Detection,ETD)对于避免经济损失至关重要。尽管许多窃电检测方法在不同场景下取得了较好的效果,但仍存在许多弊端。

一方面,现有大多数方法通常使用传统的机器学习算法来分析用户的用电行为。这些算法包括线性回归(Linear Regression,LR)^[3]、人工神经网络(Artificial Neural Network,ANN)^[4]、支持向量机(Support Vector Machine,SVM)^[5-8]、XGBoost^[9]等。但是,这些浅层机器学习方法往往基于手工特征,无法提取深层特征,检测性能较差。另一方面,重要的时序信息缺乏有效的建模。用电记录作为特殊的时间序列数据,具有一定的周期性、季节性等时序特性。另外,序列通常

到稿日期:2021-01-22 返修日期:2021-07-23

基金项目:国家自然科学基金(U20B2045,61772082)

This work was supported by the National Natural Science Foundation of China(U20B2045,61772082).

通信作者:张琦璠(zhangyufan@bupt.edu.cn)

是大规模的,且不同的序列点往往具有不同的作用,即在不同时间出现的相同电力消耗可能会产生一系列不同的影响。例如,由于制冷或供暖设备的电耗较高,夏季或冬季的低电耗比春秋季节更令人怀疑。可见,学习行为的序列表征对于窃电行为的检测至关重要。

最近,深度学习在诸如图像分类、序列建模等领域获得了巨大的成功。在深度学习模型中,循环神经网络(Recurrent Neural Network,RNN)正是为处理带有时间序列标签的数据而设计的,能够捕捉用电数据中的时间相关性,在时序数据分析上获得了很好的性能。门控循环单元(Gated Recurrent Unit,GRU)通过引入自循环机制,在一定程度上缓解了RNN中存在的梯度消失问题,更易于学习长期依赖信息。为了进一步提升模型的性能,研究者在不同方面对其进行了改进,包括加深网络层数、融合其他模型、引入注意力机制等。注意力机制^[10]的主要思想是能够让模型过滤无关信息以及关注关键信息,从而更适应大规模序列数据的学习,其已成为序列建模等领域研究的热点。

为了更好地表征用电行为序列的深层特性,提高模型处理大规模序列数据的能力,本文提出了一种基于多头注意力机制的窃电检测模型(ETD-MHA)。首先,基于用电记录的时序特征,并考虑序列的前后两个方向,采用双向门控循环单元(BiGRU)网络提取用电序列的局部顺序信息。接着,结合多头注意力提取序列的全局关键信息,加强细粒度的特征表达,并通过堆叠方式来加深网络层数,进一步提高模型特征的抽象性和鲁棒性。最后,在爱尔兰智能电表和中国国家电网数据集上,将ETD-MHA算法与LR,SVM,RF,MLP,CNN等算法进行了对比实验,实验结果表明所提模型获得了较好的效果,验证了其有效性。

2 相关工作

窃电检测可看作是一类特殊的异常检测或入侵检测问题。现有的方法主要分为两种:基于硬件的方法和基于大数据的方法。随着智能终端的广泛铺设,用户用电数据也呈现海量增长,因此大数据相关的方法引起了更多的关注。专家们通过排查电力系统中的故障信息和走访勘察的形式来检测用户是否存在窃电行为,因此可以采用有监督的学习模型对用电数据进行分析。在早期的研究中,这一类方法主要使用基于特征工程的机器学习方法^[4-8]来建立模型。文献[4]基于日负荷曲线,采用极限学习机(Extreme Learning Machine,ELM)算法学习历史用电数据的行为模式,检测的最高准确率为54.61%。文献[5]除了考虑日负荷曲线,还将用户评级作为输入特征,将SVM作为分类器,最终准确率超过了60%。然而,这些浅层方法不能处理高维数据和抽取深层次的特征,低检测率仍然不能满足实际应用。

近年来,随着深度学习概念的兴起,各种类型的深度神经网络模型已应用到用电数据分析中^[11-17]。文献[11]提出了一种新的非线性深度神经网络NeuCast,实现了用电负荷预测和异常检测。文献[13]提出了一种基于堆叠去相关自编码器(Uncorrelating AutoEncoder,UAE)的特征提取方法,并结合

支持向量机检测是否窃电。文献[15]提出了一种基于宽度和深度卷积神经网络(Convolutional Neural Network,CNN)结构的窃电检测方法,可以同时捕获用电行为的全局特征和周期性,在国网提供的数据集上达到了90%以上的准确率。文献[16]提出了一种基于RNN的深度架构,该架构能够学习序列的时间相关性,在爱尔兰数据集上达到了94%的召回率。文献[17]基于RNN和MLP,综合考虑了日用电曲线的时序特征和用户类别等外部影响因素,提出的深度融合模型性能有明显的改进。注意力机制最早在图像领域被提出,后来逐渐被应用到序列建模中,但在窃电检测中鲜有应用。Bahdanau等^[18]首先提出将注意力与递归神经网络相结合,以实现机器翻译中的对齐。Vaswani等^[19]则进一步扩展了注意力,提出的多头注意力机制在机器翻译中取得了最好的效果。受前人工作的启发,本文方法通过双向门控循环神经网络和多头注意力机制学习序列的表征,最后通过加深网络来进一步提高窃电检测的准确性。

3 窃电检测框架

基于堆叠循环神经网络和注意力机制网络的窃电检测模型的总体框架如图1所示,它由3个关键组件组成:输入层、表示层和输出层。该模型的总体公式如下:

$$\hat{y} = f(x|\theta) \quad (1)$$

其中, θ 表示模型的参数, \hat{y} 是模型输出的预测值。

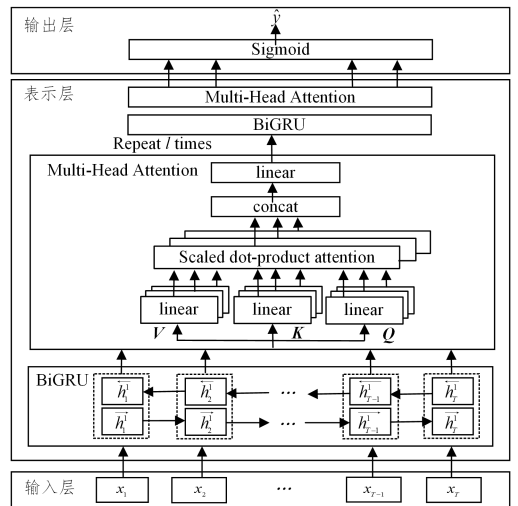


图1 窃电检测框架

Fig.1 Electricity theft detection framework

窃电检测问题可看作是离散的二分类任务,模型将用电记录 $x = [x_1, x_2, \dots, x_T]$ 划分为以下预定义类别之一(异常或正常):

$$y = \begin{cases} 1, & \text{if record is abnormal} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

3.1 输入层

由于在实际的电网数据中存在着大量噪声和异常点,以及鉴于数据规格化操作的需要,输入层需要对数据进行预处理。基于用电数据的时序特征,预处理的基本步骤包括缺失值处理、异常值处理和归一化处理。

3.1.1 缺失值处理

对于记录中缺失的数据,可以通过线性插值来恢复,具体表达式如下:

$$f(x_t) = \begin{cases} \frac{x_{t-1} + x_{t+1}}{2}, & \text{if } x_t \in \text{NaN}, x_{t-1} \text{ and } x_{t+1} \notin \text{NaN} \\ 0, & \text{if } x_t \in \text{NaN}, x_{t-1} \text{ or } x_{t+1} \in \text{NaN} \\ x_t, & \text{if } x_t \notin \text{NaN} \end{cases} \quad (3)$$

其中, $x_t \in \text{NaN}$ 表示 x_t 是空值或是非数值。

3.1.2 异常值处理

为了去除用电序列中的异常点,使用 3σ 定律来恢复不准确的值。

$$f(x_t) = \begin{cases} \text{avg}(\mathbf{x}) + 2 * \text{std}(\mathbf{x}) \\ x_t \end{cases} \quad (4)$$

其中, $\text{avg}()$ 和 $\text{std}()$ 分别用于计算序列 \mathbf{x} 的均值和标准差。

3.1.3 归一化处理

为减小数据中不同属性大小对计算的影响,对异常值处理后的用电序列进行归一化处理,具体表达式如下:

$$f(x_t) = \frac{x_t - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})} \quad (5)$$

其中, $\min()$ 和 $\max()$ 分别用于计算序列 \mathbf{x} 的最小值和最大值。

3.2 表示层

表示层实际上是一个由多层堆叠组成的深度神经网络,每一子层主要包括两个部分:双向 GRU 和多头注意力网络。双向 GRU 子层可以有效获取序列的上下文顺序信息;多头注意力子层则能够从不同的维度和表示子空间中学习到相关的信息,充分捕捉长距离的序列特征。这两个子层相互补充,能有效提高模型的行为分析能力。通过堆叠层,网络可以学习到更多抽象的潜在特征并增强学习效果。

3.2.1 双向 GRU 子层

将用电行为视作用电量序列,基于门控循环单元(GRU)的网络模型能有效提取用户的用电行为特征。GRU 使用循环神经网络中的门控机制,可在不使用单独的存储单元的情况下跟踪序列的状态。其中,重置门 r_t 通过控制上一时间步的隐藏状态的输入来丢弃和预测无关的历史信息,更新门 z_t 用于遗忘和选择记忆,它们一起控制将信息更新到新状态 \mathbf{h}_t 或候选状态 $\tilde{\mathbf{h}}_t$ 的过程。

在时刻 t , GRU 计算新状态的表达式为:

$$\mathbf{h}_t = (1 - z_t) \odot \mathbf{h}_{t-1} + z_t \odot \tilde{\mathbf{h}}_t \quad (6)$$

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_h \mathbf{x}_t + \mathbf{r}_t \odot (\mathbf{U}_h \mathbf{h}_{t-1}) + \mathbf{b}_h) \quad (7)$$

$$z_t = \sigma(\mathbf{W}_z \mathbf{x}_t + \mathbf{U}_z \mathbf{h}_{t-1} + \mathbf{b}_z) \quad (8)$$

$$r_t = \sigma(\mathbf{W}_r \mathbf{x}_t + \mathbf{U}_r \mathbf{h}_{t-1} + \mathbf{b}_r) \quad (9)$$

其中, \mathbf{x}_t 表示时刻 t 的用电记录向量, σ 是 sigmoid 激活函数。

BiGRU 包含前向层 $\overrightarrow{\text{GRU}}$ 和后向层 $\overleftarrow{\text{GRU}}$ 。前向隐藏状态 $\overrightarrow{\mathbf{h}}_t$ 和向后隐藏状态 $\overleftarrow{\mathbf{h}}_t$ 的计算式如下:

$$\overrightarrow{\mathbf{h}}_t = \overrightarrow{\text{GRU}}(\mathbf{x}_t), \overleftarrow{\mathbf{h}}_t = \overleftarrow{\text{GRU}}(\mathbf{x}_t) \quad (10)$$

然后,将它们拼接起来,获得 t 时刻的候选状态 $\mathbf{h}_t = [\overrightarrow{\mathbf{h}}_t, \overleftarrow{\mathbf{h}}_t]$, 该状态记录了 \mathbf{x}_t 的用电序列信息。

最后,双向 GRU 层的输出向量如下:

$$\text{BiGRU}(\mathbf{x}_t) = [\mathbf{h}_1, \dots, \mathbf{h}_T] \quad (11)$$

3.2.2 多头注意力子层

注意力可以选择性地关注序列中的关键信息。为了进一步提高模型处理大规模序列数据的能力,本文通过引入注意力机制来过滤无关信息,以捕获更为细粒度的特征。从 transformer 模型^[18]中得到启发,本文采用多头自注意力机制提取不同语义空间中的依赖关系。多头注意力(Multi-Head Attention, MHA)机制本质上是多个自注意力(Self-Attention)结构的结合。

在自注意力层中,采用缩放点积注意力(Scaled Dot-Product Attention, SDPA)来计算输入向量序列查询 \mathbf{Q} 、键 \mathbf{K} 、值 \mathbf{V} 的注意力输出。

$$\text{SDPA}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V} \quad (12)$$

其中,输入向量来自 BiGRU 层的输出向量,即 $\mathbf{Q} = \mathbf{K} = \mathbf{V} = \text{BiGRU}(\mathbf{x}_t)$, d 是网络隐藏层的大小,也是 \mathbf{Q} 和 \mathbf{K} 向量的维数。通过执行缩放操作,使得 softmax 函数的参数在使用较大尺寸的键时不会变得过大。

在多头注意力机制中,为了从不同的维度和表示子空间中学习到用电特征的相关信息,通过多次使用不同的参数对向量 \mathbf{Q} 、矩阵 \mathbf{K} 和矩阵 \mathbf{V} 做一个线性变换,并将结果输入到 SDPA 中,以获取不同的注意力输出。那么每个头的注意力输出 head_i 可以表示为:

$$\text{head}_i = \text{SDPA}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V) \quad (13)$$

其中, \mathbf{W}_i^Q , \mathbf{W}_i^K 和 \mathbf{W}_i^V 是线性变换参数。对 \mathbf{Q} , \mathbf{K} 和 \mathbf{V} 的不同参数的线性变换是“多头”的实质。

在 n 次计算缩放点积注意力之后,使用输出权重矩阵 \mathbf{W}^o 连接 n 个并行头 head_1 至 head_n ,得到的值即为 MHA 的结果。

$$\text{MHA}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{concat}[\text{head}_1, \dots, \text{head}_n]\mathbf{W}^o \quad (14)$$

总的来说,多头注意力机制层是将 BiGRU 层输出的由向量 \mathbf{h}_1 至 \mathbf{h}_t 合并形成的矩阵,经多次不同参数的缩放点积注意力计算,形成多个头而后合并进行线性变换。

3.2.3 堆叠策略

深度拓扑对实现良好的性能至关重要。在含有 l 层的堆叠式深度神经网络中,我们使用上一子层的输出作为当前子层网络的输入,具体的每个子层的输出向量 \mathbf{E}^i 通过式(15)来计算:

$$\mathbf{E}^i = \begin{cases} \text{SubLayer}(\mathbf{E}^{i-1}), & 1 < i \leq l \\ \mathbf{x}, & i = 1 \end{cases} \quad (15)$$

最终,我们将用电记录 \mathbf{x} 经过 l 层神经网络的输出得到的表示特征 \mathbf{E}^l 馈送到输出层。

3.3 输出层

在输出层,使用带有非线性激活层的多层感知器(MLP)来计算异常概率 \hat{y} 。

$$\hat{y} = \sigma(\mathbf{W}_o \cdot \mathbf{E}^l + \mathbf{b}_o) \quad (16)$$

其中, \mathbf{W}_o 和 \mathbf{b}_o 是输出层的权重矩阵, σ 是 sigmoid 激活函数。交叉熵损失 L 的定义如下:

$$L = -\sum_i \hat{y}_i \log \hat{y}_i + (1 - \hat{y}_i) \log (1 - \hat{y}_i) \quad (17)$$

最后,通过最小化预测值 \hat{y}_i 和真值 y_i 之间的损失 L 来训练模型。

在实际情况下,为了检测 NTL,我们将地区变压器记录电量 $E_{TM}(t)$ 与分布式电表记录电量总和 $E_{SM}(t)$ 进行比较,若满足式(18),则说明检测到窃电。

$$E_{TM}(t) > E_{SM}(t) + E_{TL}(t) + \epsilon \quad (18)$$

其中, ϵ 是 TL 的计算误差。

4 实验

4.1 实验数据

本文实验使用的数据来自爱尔兰可持续能源管理局和中国国家电网提供的用户用电数据,两个数据集具有不同的时间尺度和粒度。

数据集 1:爱尔兰可持续能源管理局^[20]发布的用电数据包含了居民和商业等类别,用户从 2009 年 7 月到 2010 年 12 月长达 535 天的日用电记录,每条记录包含了一天中的 48 次采样点。参考文献[9]中的正常样本过滤方法和窃电样本生成方法,将过滤后的居民类别用户日用电记录作为正常样本,并随机选择 20% 生成窃电样本,混合形成实验所用的合成数据集。

数据集 2:中国国家电网公司(State Grid Corporation of China, SGCC)的数据集包含从 2014 年 1 月到 2016 年 10 月 1035 天内 42372 位用户的用电记录。有 3615 位(8.5%)用户已被专业人员确认为窃电。依照文献[14]中的预处理方法,我们将数据进行 2D 折叠。表 1 列出了训练模型所使用的预处理后的数据集的详细信息。

表 1 数据集的详细信息

Table 1 Datasets details

| | SEAI 数据集 | SGCC 数据集 |
|--------|----------|----------|
| 样本数量 | 121 440 | 42372 |
| 正常样本数量 | 101 200 | 38757 |
| 异常样本数量 | 20240 | 3615 |
| 时间序列长度 | 48 | 147 |

4.2 基准和评价指标

为了评估本文模型的有效性,实现了如下基准方法:机器学习方法 LR, SVM, RF, 以及深度学习方法 MLP, CNN。参照文献[15]设置的最优参数,将测试结果调到最优。其中 LR, SVM, RF 是近几年常用的浅层模型, MLP 和 CNN 的层数与本文模型的设置相同。本文实验中的评价指标如下。

(1) F_1 值是精确率和召回率的调和平均值,定义为:

$$F_1 = \frac{2PR}{(P+R)} \quad (19)$$

其中, P 和 R 分别是精确率和召回率。 F_1 值越大,表示模型分类效果越好。

(2) 曲线下面积(AUC)指随机选择的正例的排名高于随机选择的负例的排名的概率。如果根据对正例样本的预测对所有样本按升序排序进行评分,则 AUC 的计算式如下:

$$AUC = \frac{\sum_{i \in \text{positiveClass}} \text{rank}_i - \frac{M(1+M)}{2}}{M \times N} \quad (20)$$

其中, rank_i 表示样本 i 的排序值, M 和 N 分别是正样本和负样本的数量。

(3) 平均精度(MAP)通常用于判断信息检索的质量。为了计算 MAP,根据预测得分对测试集的标记进行排序,然后按以下方式在前 k 个标记上定义精度。

$$P@k = \frac{Y_k}{k} \quad (21)$$

其中, Y_k 表示在 k 个标记中正确预测的窃电记录的数量。

4.3 实验设置

在基于 NVIDIA GeForce GTX 1080 * 2 的服务器上进行实验,利用 Keras 实现模型。以 SEAI 数据集为例,参数设置如下:隐藏层大小设置为 128,注意力机制头数设置为 8,模型深度设置为 4。在训练中, batchSize 设置为 128,学习率 learning rate 设置为 0.001。为了降低模型过拟合的风险,设置纵向堆叠层间的 dropout 概率为 0.5,使用 Adam 来优化模型。

4.4 异常检测的结果

本文将训练比例分别设置为 50%, 70% 和 80%, 进行 3 组实验。本文测试了 ETD-MHA 模型和对比算法(LR, SVM, RF, MLP 和 CNN)在两个真实数据集上的性能,结果如表 2 和表 3 所列。对比分析实验结果可以得到如下结论:

(1) 在大多数情况下,基于深度学习的方法(MLP, CNN 和 ETD-MHA)优于传统机器学习方法(LR, SVM 和 RF)。深层神经网络可以学习到异常样本和正常样本的隐藏特征。而相比 MLP, CNN 能更好地捕获局部相关性,取得较好的效果。

(2) 在所有数据集上,本文提出的 ETD-MHA 模型在 3 组实验上的 AUC, F_1 , $P@100$ 和 $P@200$ 都取得了最好的结果。例如,在训练数据占比为 80% 的 SGCC 数据集上, F_1 值相比基准模型最高提升了 31.9%, 而 $P@100$ 达到了 99.2%, 相比基准模型最高提升了 51.2%。性能实验的结果说明,用电行为的序列信息在窃电行为检测过程中发挥了重要作用。结合双向 GRU 子层和注意力机制子层的 ETD-MHA 模型可以有效整合历史序列中的信息,提升窃电行为检测算法的性能。

表 2 ETD-MHA 模型在 SEAI 数据集上的表现

Table 2 Performance of the ETD-MHA model on SEAI dataset

| Methods | Training ratio=50% | | | | Training ratio=70% | | | | Training ratio=80% | | | |
|---------|--------------------|------|-------|-------|--------------------|------|-------|-------|--------------------|------|-------|-------|
| | AUC | F1 | P@100 | P@200 | AUC | F1 | P@100 | P@200 | AUC | F1 | P@100 | P@200 |
| LR | 68.3 | 45.2 | 66.4 | 60.7 | 70.9 | 47.2 | 68.9 | 61.8 | 71.6 | 49.9 | 69.2 | 60.5 |
| SVM | 74.4 | 46.7 | 70.9 | 66.2 | 77.0 | 48.1 | 75.4 | 64.9 | 79.1 | 50.2 | 76.0 | 66.4 |
| RF | 79.9 | 62.9 | 90.6 | 88.4 | 77.8 | 63.3 | 95.5 | 91.5 | 78.0 | 66.1 | 92.3 | 88.5 |
| MLP | 85.0 | 75.8 | 89.9 | 88.0 | 85.9 | 77.7 | 95.0 | 93.3 | 86.4 | 78.8 | 95.0 | 92.5 |
| CNN | 89.4 | 84.7 | 94.6 | 93.4 | 91.9 | 85.2 | 96.2 | 95.6 | 93.5 | 87.1 | 96.9 | 95.9 |
| ETD-MHA | 95.3 | 88.5 | 98.4 | 97.8 | 96.2 | 91.2 | 98.9 | 98.7 | 96.4 | 92.7 | 99.4 | 99.2 |

¹⁾ Sustainable Energy Authority of Ireland, SEAI

表 3 ETD-MHA 模型在 SGCC 数据集上的表现

Table 3 Performance of the ETD-MHA model on SGCC dataset

| Methods | Training ratio=50% | | | | Training ratio=70% | | | | Training ratio=80% | | | |
|---------|--------------------|------|-------|-------|--------------------|------|-------|-------|--------------------|------|-------|-------|
| | AUC | F1 | P@100 | P@200 | AUC | F1 | P@100 | P@200 | AUC | F1 | P@100 | P@200 |
| LR | 67.7 | 40.8 | 64.6 | 56.7 | 69.2 | 43.4 | 66.7 | 57.8 | 70.6 | 46.7 | 65.6 | 57.8 |
| SVM | 71.8 | 45.6 | 68.6 | 59.2 | 72.8 | 46.9 | 72.4 | 60.7 | 74.1 | 48.4 | 73.5 | 62.0 |
| RF | 73.2 | 44.5 | 90.8 | 86.7 | 73.7 | 47.4 | 92.6 | 88.6 | 73.9 | 49.2 | 90.5 | 85.4 |
| MLP | 77.6 | 47.7 | 94.0 | 89.6 | 78.6 | 50.8 | 96.9 | 93.3 | 78.2 | 53.0 | 95.0 | 90.0 |
| CNN | 88.1 | 50.3 | 95.6 | 94.8 | 90.2 | 54.3 | 96.9 | 95.6 | 91.1 | 55.1 | 96.9 | 95.2 |
| ETD-MHA | 90.3 | 56.3 | 98.1 | 97.6 | 93.6 | 59.3 | 98.8 | 97.1 | 94.5 | 61.6 | 99.2 | 97.2 |

4.5 变体模型对比实验

为了评估所提模型关键模块的有效性,设计了以下 4 种变体模型,模型各层的参数均一致。

(1)ETD-MHA-NeGRU:基于所提模型,去掉 BiGRU 子层,仅保留 MHA 子层。

(2)ETD-BiGRU:基于所提模型,去除多头注意力层,仅利用 BiGRU 子层提取特征。

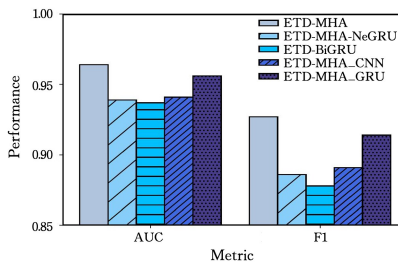
(3)ETD-MHA_CNN:基于所提模型,将各层 BiGRU 均替换为 CNN。

(4)ETD-MHA_GRU:基于所提模型,将原模型的各层 BiGRU 均替换为 GRU。

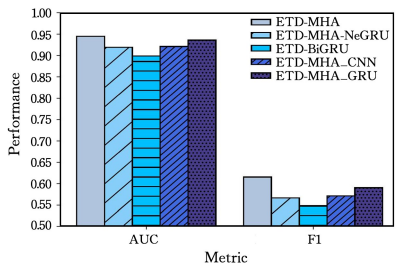
观察图 2 所示的所提模型和变体模型在两个数据集上的实验结果,可以得出如下结论:

(1)ETD-MHA-NeGRU 缺少 BiGRU 子层对底层局部时序特征的提取,且没有考虑时序先后,而时序信息在序列建模中很重要,因此检测性能较差。在 SGCC 数据集上,AUC 值相比本文提出的模型降低了 2.8%。因此,BiGRU 子层在窃电检测框架中必不可少。

(2)与缺少注意力子层的 ETD-BiGRU 相比,ETD-MHA 模型在两个数据集上的 AUC 值和 F1 值都取得了更好的效果。例如,在 SGCC 数据集上,AUC 值和 F1 值分别有 5.1% 和 12.2% 的提升。这说明注意力机制考虑了不同序列点的作用,可以捕获序列点之间有价值的信息,过滤无关信息,可以有效地从历史序列中提取信息,得到更佳的效果。



(a) 模型在 SEAI 数据集上的表现



(b) 模型在 SGCC 数据集上的表现

图 2 变体模型的对比实验

Fig. 2 Variant model comparison experiments

(3)ETD-MHA_CNN 和 ETD-MHA_GRU 的效果仅次于最优模型,可见本文设计的窃电模型的框架设计合理,且所提模型 ETD-MHA 中使用的 BiGRU 组件学习了序列中前后两个方向的内在联系,提取局部时序特征的能力强于 GRU 和 CNN。例如,在 SGCC 数据集上,本文模型的 AUC 值相比 ETD-MHA_CNN 和 ETD-MHA_GRU 分别提升了 2.6% 和 1.0%。

总的来说,BiGRU 子层对输入的用电量序列提供了更高级的特征表示,多头注意力机制是性能提升的关键,不仅与 BiGRU 子层互相补充,还能有效地捕捉较长距离的序列特征,模型分类能力也显著增强。

4.6 参数分析

基于 SEAI 数据集,我们检查了两个重要的超参数的敏感性:堆叠层 l 和多头注意力机制的 head 数量 n 。如图 3(a) 所示,随着堆叠神经网络层数的增加,模型的效果具有一定的提升。综合性能和效率考虑,设置栈式堆叠循环神经网络的层数 l 为 4。在多头注意力层,head 的数量会影响对不同特征的关注度,head 过多或过少都可能造成有效特征干扰或缺失,从而影响模型分类效果。如图 3(b) 所示,本文设置 head 数目分别为 1,2,4,8 和 16,可知最佳的 head 数目为 8。

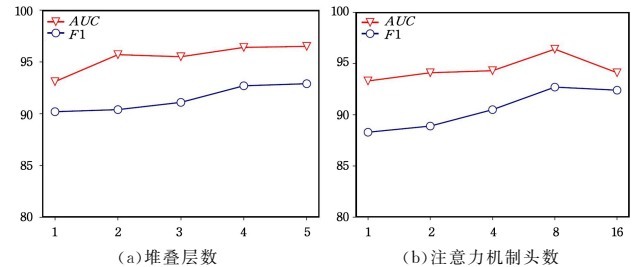


图 3 参数分析

Fig. 3 Parameters analysis

结束语 窃电会造成大量经济损失,因此一直受到电力公司和研究者的关注,利用海量数据开展窃电检测已成为国内外的研究热点。针对现有窃电检测方法依赖于手工特征、行为序列表征不足、检测精度差等问题,本文提出了基于多头注意力机制的窃电检测模型 ETD-MHA。该模型的优点在于其利用 BiGRU 网络来获取序列中前后两个方向的内在联系,得到局部顺序信息,并利用多头注意力机制加强特征表达,获取全局关键信息,最后通过加深网络层数来提升学习效果。实际数据的算例分析和对比测试展示了该模型相比其他现有模型的优越性。

参考文献

[1] DEPURU S S S R, WANG L, DEVABHAKTUNI V. Electricity

- theft; Overview, issues, prevention and a smart meter based approach to control theft[J]. *Energy Policy*, 2011, 39(2): 1007-1015.
- [2] NAVANI J P, SHARMA N K, SAPRA S. Technical and Non-Technical Losses in Power System and Its Economic Consequence in Indian Economy[J]. *International Journal of Electronics & Computer Science Engineering*, 2012, 1(2): 757-761.
- [3] AKOUEMO H N, POVINELLI R J. Probabilistic anomaly detection in natural gas time series data[J]. *International Journal of Forecasting*, 2016, 32(3): 948-956.
- [4] NIZAR A H, DONG Z Y, WANG Y. Power Utility Nontechnical Loss Analysis With Extreme Learning Machine Method[J]. *IEEE Transactions on Power Systems*, 2008, 23(3): 946-955.
- [5] 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.
- [6] DEPURU S S S R, WANG L, DEVABHAKTUNI V, et al. A hybrid neural network model and encoding technique for enhanced classification of energy consumption data[C] // 2011 IEEE Power and Energy Society General Meeting. IEEE, 2011: 1-8.
- [7] DEPURU S S S R, WANG L, DEVABHAKTUNI V. Support vector machine based data classification for detection of electricity theft[C] // 2011 IEEE/PES Power Systems Conference and Exposition. IEEE, 2011: 1-8.
- [8] NAGI J, YAP K S, TIONG S K, et al. Improving SVM-Based Nontechnical Loss Detection in Power Utility Using the Fuzzy Inference System[J]. *IEEE Transactions on Power Delivery*, 2011, 26(2): 1284-1285.
- [9] JOKAR P, ARIANPOO N, LEUNG V C. Electricity Theft Detection in AMI Using Customers' Consumption Patterns[J]. *IEEE Transactions on Smart Grid*, 2017, 7(1): 216-226.
- [10] LUONG M T, PHAM H, MANNING C D. Effective approaches to attention-based neural machine translation[J]. arXiv: 1508.04025, 2015.
- [11] CHEN P, LIU S, SHI C, et al. NeuCast: Seasonal Neural Forecast of Power Grid Time Series[C] // Twenty-Seventh International Joint Conference on Artificial Intelligence IJCAI-18. International Joint Conferences on Artificial Intelligence Organization. 2018: 3315-3321.
- [12] ZHUANG S J, YU Z Y, GUO W Z, et al. Short Term Load Forecasting via Zoneout-based Multi-time Scale Recurrent Neural Network[J]. *Computer Science*, 2020, 47(9): 105-109.
- [13] HU T Y, GUO Q L, SUN H B. Nontechnical loss detection based on stacked uncorrelating autoencoder and support vector machine[J]. *Automation of Electric Power Systems*, 2019, 43(1): 119-127.
- [14] MENG Z, XU X. A Hybrid Short-Term Load Forecasting Framework with an Attention-Based Encoder-Decoder Network Based on Seasonal and Trend Adjustment[J]. *Energies*, 2019, 12(24): 4612.
- [15] ZHENG Z, YANG Y, NIU X, et al. Wide and Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids[J]. *IEEE Transactions on Industrial Informatics*, 2017, 14(4): 1606-1615.
- [16] NABIL M, MAHMOUD M, ISMAIL M, et al. Deep recurrent electricity theft detection in AMI networks with evolutionary hyper-parameter tuning[C] // 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). IEEE, 2019: 1002-1008.
- [17] CHEN Z, MENG D, ZHANG Y, et al. Electricity theft detection using deep bidirectional recurrent neural network[C] // 2020 22nd International Conference on Advanced Communication Technology (ICACT). IEEE, 2020: 401-406.
- [18] BAHDANAU D, CHO K, BENGIO Y. Neural Machine Translation by Jointly Learning to Align and Translate[J]. arXiv: 1409.0473, 2014.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. arXiv: 1706.03762, 2017.
- [20] ISSDA. Data from the commission for energy regulation[EB/OL]. [2019-07-01]. <http://www. ucd. ie/issda/data/commissionforenergyregulationcer/>.



XIAO Ding, born in 1966, lecture. His main research interests include software engineering and artificial intelligence.



ZHANG Yu-fan, born in 1996, postgraduate. Her main research interests include power systems and artificial intelligence.

(责任编辑:柯颖)