

郑征<sup>1</sup> 谭磊<sup>2</sup> 周楠<sup>1</sup> 韩军伟<sup>1</sup> 高晶<sup>2</sup> 翁理国<sup>3</sup>

# 基于多头注意力卷积网络的电力负荷预测

## 摘要

预测居民用电相当于预测一个多元时间序列.针对多个传感器信号的特定窗口能够利用预测模型提取不同的特征来预测用电量.然而,由于时间序列内部特征存在着不规则的模式,包括电力属性之间隐藏的相关性,使得负荷预测准确率不高.为了提取复杂的不规则电力模式,选择性地学习时空特征以减少电力属性间的平移方差,本文提出了一种基于多头注意力的卷积神经网络深度学习模型.相较于单纯的时间序列模型,该模型利用卷积和加权机制对电力属性和有功功率间的局部相关性进行建模.它利用 softmax 函数和点积运算的注意力分数来模拟电力需求的瞬态和脉冲特性,有效地对瞬时脉冲功耗进行预测.在美国加州大学欧文分校(University of California, Irvine, UCI)家庭用电数据集共 2 075 259 个时间序列上的实验表明,所提出的模型与现有方法相比,准确率得到了较大提升.

## 关键词

深度学习;注意力机制;时间序列预测;负荷预测

中图分类号 TP183

文献标志码 A

收稿日期 2021-11-29

资助项目 国家自然科学基金(61773219)

作者简介

郑征,教授级高级工程师,研究方向为电力大数据分析及应用.335537837@qq.com

1 国网河南省电力公司经济技术研究院,郑州,450052

2 河南九域腾龙信息工程有限公司,郑州,450052

3 南京信息工程大学 江苏省大数据分析技术重点实验室,210044

## 0 引言

居民用电预测可被定义为一个多元时间序列预测问题<sup>[1]</sup>.在能源管理系统(Energy Management System, EMS)领域,包括最近备受关注的智能电网服务,基于历史电力消耗和其他电力属性预测未来电力需求的过程是必不可少的<sup>[2]</sup>.在电力规划阶段主要的问题是对电力消耗的性质进行建模,而目前这种建模仅限于线性回归等单纯的时间序列模型<sup>[3]</sup>.

在负荷预测领域,机器学习的目标是利用所有客户的历史数据生成数据库,预测未来的电力需求.移动平均<sup>[4]</sup>和支持向量回归<sup>[5]</sup>是电力预测的代表性方法,它们成功地模拟了终端用户的长期行为,提高了预测精度.然而,在高时间分辨率条件下<sup>[6]</sup>针对暴露行为模式的短期行为建模的预测任务中,需要一种更加复杂和实用的时间序列建模方法.

深度学习由于其优秀的特征提取能力,使得其在负荷预测领域被重点关注.但深度学习的电力预测模型<sup>[7]</sup>仍然面临两个问题:一是有功功率消耗模式与其他电力属性之间存在多重共线性,二是电力消耗的瞬态和脉冲性.虽然目前已经提出利用卷积运算提取电力数据局部相关性的方法,但同时或单独使用对有功功率有关键影响的用电设备会导致深度学习模型性能的下降.具体来说,没有针对多元和脉冲特性进行建模的模型在能量峰值时性能会下降.在深度学习兴起之前,大多数时间序列建模方法都是基于符号-动力方法<sup>[8]</sup>.由于时间轴上的轻微移动会导致两个时间序列之间的距离变大,Lin 等<sup>[9]</sup>提取并建模旋转不变符号,构造了模式袋.为了建立时不变特征并进行非线性映射来预测电力消耗,Tso 等<sup>[10]</sup>提出一种神经网络,并与现有基于规则和符号的预测方法进行了性能比较.在基于机器学习的电力需求预测方法中,自回归综合移动平均、决策树以及神经网络的预测效果最好,其非线性映射能力备受关注<sup>[11]</sup>.

深度学习模型包括学习时间门函数的长短时记忆网络(Long Short-Term Memory, LSTM)<sup>[12-13]</sup>和提取功率谱间局部相关性的卷积神经网络(Convolutional Neural Network, CNN),在电力负荷预测领域取得了显著的成绩.此外,利用 CNN 和 LSTM 层作为构建块,自编码器(AE)<sup>[14]</sup>和对抗学习模型(如生成对抗网络(GAN)<sup>[15]</sup>)作为概率设计方法,间接地验证了深度学习在电力预测中的有效性.

然而,考虑到时间序列的特征(非恒定均值和方差、多周期性、高波动性、跳跃性等),电力负荷预测是一项困难的任务.Mocanu等<sup>[16]</sup>将随机预训练阶段引入到神经网络的电力负荷预测中,并在不同的时间条件下对其进行评估.受限玻尔兹曼机(RBM)是一种代表性的无监督学习神经网络模型,其目标是最小化层之间的KL散度<sup>[17]</sup>.堆叠RBM适用于从电力负荷数据中学习先验分布,对提高预测性能有显著贡献.Kong等<sup>[18]</sup>尝试利用循环神经网络(RNN)和LSTM来预测电力消耗,并验证了可行性,显著提高了预测性能.

针对预测序列的不同特点,研究人员提出了预测模型的优化方法.对深度学习框架或损失函数进行优化使得模型的预测性能有了显著的提高. Shi等<sup>[19]</sup>在时间序列建模之前引入自编码器提取多元特征,采用池化操作的RNN有选择地更新梯度.此外,一种利用遗传算法优化影响预测性能的关键因素时滞参数的方法也被提出<sup>[20]</sup>.与遗传算法相比,基于粒子群优化算法(PSO)的负荷预测模型具有更快的收敛速度和更大范围的空间探索,因其性能优于现有的深度学习模型而受到关注<sup>[21]</sup>.还有一些研究对损失函数进行优化,如分位数损失函数在神经网络中的适应性<sup>[22-23]</sup>.

从基于深度学习方法的电力消耗预测的相关研究中可以看出,深度学习是进行电力预测的有效手段.本文提出了一种基于多头注意力的深度学习模型,利用时间序列建模中卷积和加权机制来对电力属性和有功功率间的局部相关性进行建模.其核心思想是通过卷积运算从多变量电力属性中提取特征,通过门控运算对功率谱进行时间序列建模,采用基于softmax函数运算的概率数据定位方法,有效地预测瞬时功耗和脉冲功耗.综上所述,本文方法中注意力的权重函数可以模拟包括能量峰值在内的时间序列电力数据的短期模式.

本文的其余部分组织如下:第1部分说明了如何通过具有多头注意力的深度学习模型有选择性地提取电力属性;第2部分通过各种实验来评估模型的性能,包括多头注意力向量的可视化以及与最新模型的比较;第3部分对全文进行了总结,并对未来的发展方向进行了讨论.

## 1 基于多头注意力的卷积循环神经网络

本节描述了CNN-LSTM网络的结构,该网络利

用多头注意力提取时空特征,利用通道注意力提取特征通道间的联系,并从功率谱中建模电力消耗.多头注意力利用softmax和点积运算实现多目标关注,以模拟暂态和脉冲电力需求值.

### 1.1 网络结构概述

该方法的主要目的是利用CNN-LSTM网络进行电力需求预测.本文利用多头注意力机制并将其表示为函数 $\phi$ 来提取时空特征并预测未来电力需求.由于存在多层叠加表示的复杂非线性映射,本文采用直接预测策略<sup>[24]</sup>,避免了递归策略中偏差的累积.直接预测策略的公式表示如下:

$$\hat{y}_t = f_h(\mathbf{X}_{t-h}^\omega; \Theta_h) + e_{t,h}, \quad (1)$$

其中: $f_h$ 表示直接预测模型; $\mathbf{X}_{t-h}^\omega = [\mathbf{R}_{t-h}, \dots, \mathbf{R}_{t-h-\omega}]$ 带有时滞 $\omega$ ;  $\Theta_h = [\psi_h, \theta_h]$ 表示 $h$ 层的模型参数,其中 $\psi_h$ 是一组超参数, $\theta_h$ 是一组参数; $e_{t,h}$ 是模型的预测误差.电力需求模型提前 $h$ 步预测.输入电力属性向量序列 $\mathbf{R}_t$ 由总体有功功率、无功功率、电压、电流和三个子指标组成.

众所周知,卷积循环神经网络在信号处理领域具有数据驱动的滤波学习优势,主要是用来提取时空特征.图1展示了带有多头注意力的CNN-LSTM的总体架构.它主要分为两个阶段:

1) 数据预处理阶段定义训练CNN-LSTM模型所需的超参数.在定义超参数之前,对每个电力属性和滑动窗口进行最小-最大标准化.滑动窗口由时间滞后的输入信号 $\omega$ 的长度定义,其步幅参数 $\tau$ 决定了时间步长重叠的数量.对每个时段的电力消耗数据进行采样,时间分辨率为1 min、15 min、30 min、45 min、60 min、1 d和1周(7 d).

2) 建模阶段根据定义的超参数调整CNN-LSTM模型的权重.为了从功率谱中提取时空特征,本文构建了一个卷积池操作对电力属性之间的隐藏相关性进行建模,并利用循环记忆单元的门控操作来对时间序列数据之间的时间相关性进行建模.同时,本文还设计了一种多头注意力机制和一种通道注意力机制,前者用于将激活值解释为概率并构建相关矩阵,置于每个卷积层和循环层之间,后者用于提取特征通道间的联系,滤除冗余通道信息.此外,将类激活映射(Class Activation Map, CAM)放入最后一层卷积层进行网络输出分析,通过对CNN最顶层特征图的权值进行求和,对感受野进行局部定位.

### 1.2 卷积循环神经网络

利用神经回归模型对电力消耗进行建模的主要

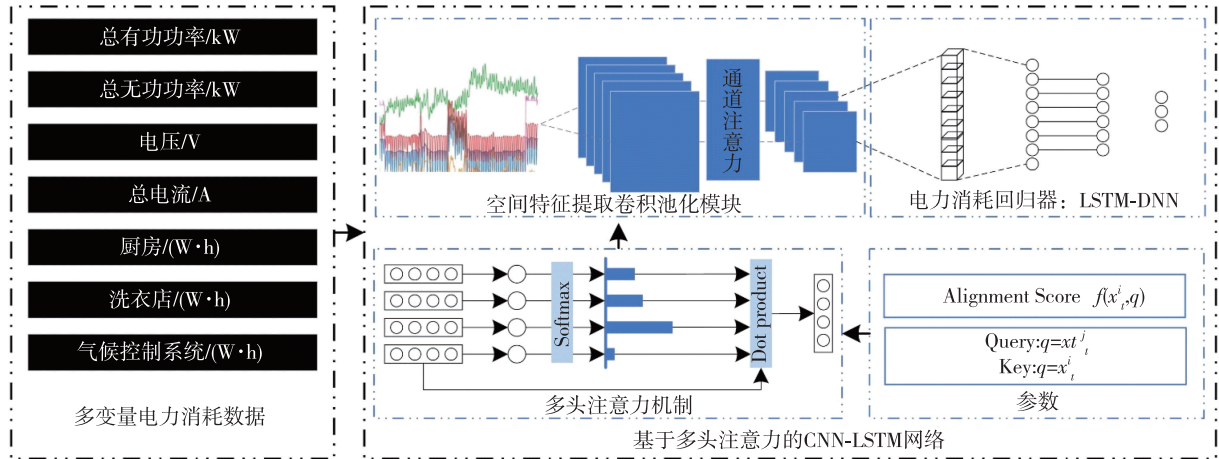


图1 基于多头注意力的CNN-LSTM网络整体结构

Fig. 1 Overall structure of CNN-LSTM based on multi-headed attention

问题在于从有限的电力消耗数据样本中提取时空特征<sup>[25]</sup>.首先构造CNN和LSTM,从时间序列的电力消耗数据中学习特征.两种深度学习模型在保持时空特征互补关系的前提下,以前后顺序的方式进行组合.考虑到CNN中的卷积 $f_c$ 和池化操作成功应用于信号处理领域,因此也适用于建模电力消耗序列,利用滑动窗口信号之间的局部相关性提取特征.卷积运算通过学习滤波器提取隐藏的相关性,以减少特征之间的平移方差,并保持电力属性之间的空间关系.假设给定 $t$ 个时间步长,使用 $m \times 1$ 大小的滤波器 $W$ ,第 $l$ 层第 $a$ 个节点和电力属性序列 $R_\tau$ 中第 $\tau$ 个元素进行卷积运算:

$$f_c^l(X_t) = \left[ \sum_{a=0}^{m-1} W_a^{l-1} R_\tau \right]_{\tau=1}^{t-\omega} \quad (2)$$

由于卷积运算 $f_c$ 中被扭曲和复制的输出向量的维数随着卷积滤波器的数量而增加,因此通过最大池化运算提取临近节点激活的特征.池化是指CNN为了减少网络参数而进行的降维过程,以便于更快的计算<sup>[26]</sup>.最大池化操作对 $k \times 1$ 区域下的特征选择和降维有影响.本文所提出的一维卷积池化操作旨在从每个属性的功率谱中提取空间特征,并将编码向量序列传递给LSTM.通过滑动窗预处理,卷积池化函数 $f_c$ 提取的空间特征包含了窗口尺寸 $\omega$ 的时间序列信息.LSTM的核心思想是适应由输入门、遗忘门和输出门组成的门控运算,并在时间步长 $t$ 时用细胞状态 $C_t$ 和隐藏值 $h_t$ 生成编码向量 $f_t$ :

$$f_t = h_t = o_t \cdot \tanh(C_t), \quad (3)$$

( $\cdot$ )表示元素相乘.CNN-LSTM提取时空特征后,利用经典的多层感知器(MLP)、激活函数 $\sigma$ 和权重矩

阵 $W^l$ 完成回归函数 $\phi$ :

$$f^l(X_t) = \sigma(W^l f_t^{l-1}(f_c^{l-2}(X_t)) + b^l), \quad (4)$$

其中,MLP的最后一层使用线性激活函数,输出标量值 $\hat{y}$ 表示为电力预测.CNN-LSTM回归器通过最小化均方误差(MSE)表示的损失函数,利用梯度下降优化的反向传播算法进行更新.MSE损失函数如下,其中 $n$ 为观测次数:

$$L_{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (5)$$

### 1.3 多头注意力机制

注意力用于计算两个来源的元素之间的相关分数.直观地说,注意力机制是一种计算查询与键之间的相似性,并提取与查询相关的值进行加权和的操作.

假设给定窗口 $X_t^\omega = [R_t, \dots, R_{t-\omega}]$ ,其时间步长 $t$ ,查询 $q$ 为时空特征向量表示,注意力通过相容性函数 $f(R_t, Q)$ 计算相关性.相关分数向量 $A_t = [f(R_\tau, Q)]_{\tau=t}^{t-\omega}$ 由经过相容性函数计算的查询元素和键元素之间的一系列相关性组成.相容性函数通过softmax运算被解释为概率分布 $p(z | R_t, Q) = \text{softmax}(A_t)$ ,因此查询 $Q$ 与键之间的相关性可以表示为下列方程中定义的注意分数 $s$ , $s$ 的值表示不同重要性的电力消耗期望:

$$p(z = t | Q) = \frac{\exp(f(R_t, Q))}{\sum_{\tau=t}^{t-\omega} \exp(f(R_\tau, Q))}, \quad (6)$$

$$p(z = t | Q) = \frac{\exp(f(R_t, Q))}{\sum_{\tau=t}^{t-\omega} \exp(f(R_\tau, Q))} s_t =$$

$$\sum_{\tau=t}^{t-\omega} p(z = \tau | \mathbf{R}_t, \mathbf{Q}) = E_{t-p(z|\mathbf{R}_t, \mathbf{Q})}(\mathbf{R}_t). \quad (7)$$

注意力机制可以表示为深度学习层,注意力层位于功率谱和时间序列的建模步骤之间.注意力层适合对电力设备使用量的骤增进行建模,而传统的深度学习模型难以预测这种情况的电力消耗.自注意力是注意力机制的一种特殊情况:它用信号  $\mathbf{X}_t$  替换查询  $\mathbf{Q}$ .单头注意力机制直观地对卷积池化层编码的自身进行点积以获取时空特征向量间的协方差.

本文提出的多头注意力是注意力机制的延伸,它将多个注意力保持在一个窗口内,比单头注意力表现性能更好.图 2 展示了多头注意力机制的内部结构.本文不再为每个时间步长计算单个标量分数  $f(\mathbf{X}_t, \mathbf{X}_t)$ ,而是定义与  $\mathbf{X}_t$  长度相同向量的相容性函数  $f$  的输出.  $Z_k$  表示来自相容性函数的相关分数.由于将注意力向量的维数扩展为  $f(\mathbf{R}_t, \mathbf{Q})$ ,因此可以将每步  $t$  中电力属性元素或编码特征  $k$  的重要性权重向量  $\mathbf{p}_{kt}$  形式化:

$$\mathbf{p}_{kt} = p(Z_k = i | \mathbf{R}_t, \mathbf{Q}). \quad (8)$$

### 1.4 通道注意力机制

本文通过引入通道注意力机制,对卷积特征的通道之间的相互依赖关系进行建模,以提高网络的表示能力.通过这种机制,网络可以学习使用全局信息有选择地强调信息量大的特征,并抑制不太有用的特征.通道注意力机制通过模拟通道之间的相互依赖关系,自适应地重新校准通道方面的特征响应.通道注意力机制内部结构如图 3 所示.为了解决通

道依赖性的问题,首先对输入特征使用全局平均池化操作,把全局空间信息压缩成一个通道描述符,然后利用全连接层和激活函数进行通道间的压缩与扩张,旨在完全捕获通道相关性.选择 Sigmoid 激活函数作为门控机制,得到通道分数图,再与输入特征相融合,得到最终融合了通道间信息的输出特征.

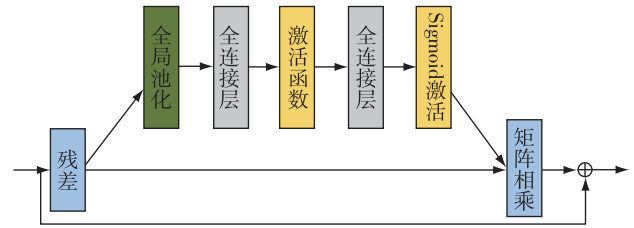


图 3 通道注意力机制的内部结构

Fig. 3 Structure of channel attention mechanism

通道注意力机制超参数设置如表 1 所示.经过上层的输入之后,本文使用 ResNet-50 作为主干网络,通道注意力机制模型分为 4 个模块.首先使用  $7 \times 7$  的卷积核,步长为 2 的方式对输入特征进行下采样,再使用  $3 \times 3$  的卷积核,步长为 2 的方式对全局特征进行最大池化;最后一个模块其他 3 个注意力模块基本相同,首先使用  $1 \times 1$  的卷积核融合通道特征,再使用  $3 \times 3$  的卷积核进行下采样,然后使用  $1 \times 1$  的卷积核进行通道特征的融合,并进行多次重复操作.每个注意力模块中包含一个小的全连接层,最后再进行一个简单的逐通道缩放操作,输出所要预测的目标参数.

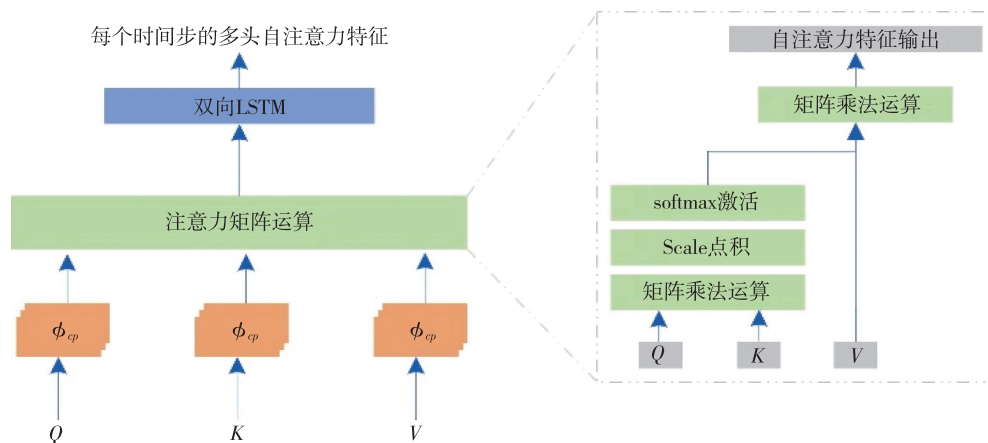


图 2 多头注意力机制的内部结构

Fig. 2 Structure of multi-headed attention mechanism



表1 通道注意力机制超参数设置

Table 1 Channel attention mechanism parameter setting

Layer name	Stride	SE-ResNet50
Layer 0	2	conv, 7×7, 64
	2	max pool, 3×3
Layer 1	1	conv, 1×1, 64
		conv, 3×3, 64
		conv, 1×1, 256
		SE:fc, [16, 256]
×3		
Layer 2	2	conv, 1×1, 128
		conv, 3×3, 128
		conv, 1×1, 512
		SE:fc, [32, 512]
×4		
Layer 3	2	conv, 1×1, 256
		conv, 3×3, 256
		conv, 1×1, 1 024
		SE:fc, [64, 1 024]
×6		
Layer 4	2	conv, 1×1, 512
		conv, 3×3, 512
		conv, 1×1, 2 048
		SE:fc, [128, 2 048]
×3		
global average pool, 1000-1 fc, softmax		

## 2 实验

本节介绍了具有多头注意力的 CNN-LSTM 如何通过预测误差的 10-折交叉验证来预测电力消耗和评估性能,然后与相关的深度学习模型进行定量比较.本文所有实验全部都基于 PyTorch 深度学习框架,版本号为 1.10.0,Python 版本为 3.8.12,显卡型号为 NVIDIA GeForce RTX 3070,显存共 8 GB;CPU 型号为 i7-11700F;计算机内存为 16 GB.

## 2.1 数据集

在 UCI 家庭用电数据集<sup>[27]</sup>上验证 CNN-LSTM 算法的有效性.该数据集收集了约 207 万个多通道传感器的数据,由 2 075 259 个时间序列和 7 个变量组成,电力属性包括总有功功率、总无功功率、电压、电流和另外 3 个子属性.数据被标准化,并在带有时滞参数的滑动窗口中进行处理.预测模型在时间分辨率条件下接收到这 7 个属性,并生成下一个时间步长的总有功功率.

## 2.2 模型实现细节

CNN-LSTM 的架构可以根据堆叠的卷积池化层和 LSTM 层的数量,以及卷积滤波器的数量、卷积核大小和 LSTM 中的隐藏层节点数量进行不同程度的修改.由于深度学习模型需要一个优化过程,因此有必要对超参数进行调整和优化.本文模型的超参数是经过对电力消耗统计数据的直观判断以及对迭代优化的深入研究确定的.表 1 展示了 CNN-LSTM 的超参数设置.图 4 展示了所提出的模型总体架构.

## 2.3 电力消耗预测性能比较

表 2 和图 5 比较了包括卷积神经网络和 LSTM 在内的机器学习模型的预测性能结果.本文模型对时空特征进行选择建模,与传统 CNN-LSTM 神经网络相比,误差降低了 22.72% (时间分辨率为 1 min).评价指标是基于欧氏空间误差测量的均方误差(MSE):

$$E_{MS} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \phi(X_i))^2. \quad (9)$$

表2 CNN-LSTM 超参数设置

Table 2 Setting of super parameters of CNN-LSTM

操作	卷积滤波器/节点的数量	卷积核	步长	激活函数	参数大小/个
一维卷积	64	2×1	1	tanh	192
TimeDistributed(一维最大池化)			2	tanh	0
TimeDistributed(多头注意力)				softmax	0
TimeDistributed(一维卷积)	64	2×1	1	tanh	8 256
TimeDistributed(一维最大池化)			2	tanh	0
TimeDistributed(多头注意力)				softmax	0
Dropout	0.5				0
LSTM	64			tanh	73 984
LSTM	64			tanh	33 024
Dropout	0.5				0
Dense	64			tanh	4 160
Dense	64			tanh	4 160
Dense	1			linear	65

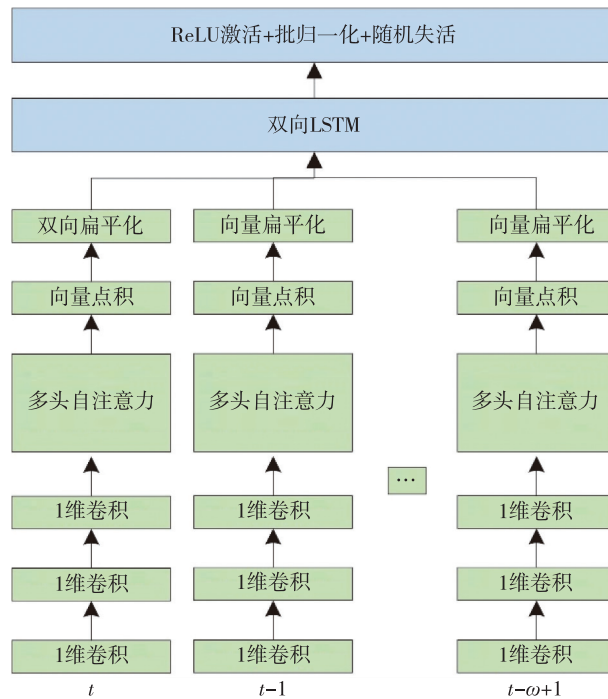


图4 基于多头注意力的 CNN-LSTM 结构示意图

Fig. 4 Structure diagram of CNN-LSTM based on multi-headed attention

表3 基于不同时间分辨率的不同建模方法的 MSE 指标比较

Table 3 Comparison of the MSE index of modeling methods based on different time resolutions

时间分辨率	LR	ARIMA	DT	RF	SVR	MLP	CNN (1D)	CNN (2D)	LSTM	CNN-LSTM (单头)	本文模型
1 min	0.084 4	0.083 8	0.102 6	0.080 2	0.079 7	0.100 3	0.074 8	0.066 6	0.074 1	0.066 0	0.051 0
15 min	0.251 5	0.242 8	0.423 4	0.397 2	0.322 8	0.237 0	0.219 7	0.212 3	0.220 8	0.208 5	0.180 7
30 min	0.305 0	0.299 2	0.519 9	0.395 1	0.361 9	0.281 3	0.270 3	0.268 4	0.277 3	0.259 2	0.229 5
45 min	0.332 1	0.343 1	0.480 4	0.431 5	0.424 7	0.320 8	0.320 5	0.318 3	0.322 0	0.313 3	0.277 8
1 h	0.339 8	0.325 2	0.525 9	0.434 4	0.405 9	0.307 2	0.288 5	0.286 5	0.290 3	0.280 3	0.263 1
1 d	0.108 3	0.098 0	0.189 1	0.119 0	0.131 1	0.113 4	0.109 4	0.106 9	0.112 9	0.101 3	0.091 3
1 周	0.062 4	0.061 6	0.070 6	0.061 7	0.062 0	0.044 1	0.036 9	0.033 3	0.038 7	0.032 8	0.028 9

进一步评估不同时间分辨率下的模型性能,其结果如表 2 所示(LR:线性回归;ARIMA:自回归综合移动平均;DT:决策树;RF:随机森林;SVR:支持向量回归;MLP:多层感知器;CNN(1D 为一维卷积,2D 为二维卷积);LSTM:长短时记忆网络).每个 MSE 值都是 10-折交叉验证的结果.预测误差在 45 min 和 1 h 的时间分辨率下最大,有时短时特征的丢失会对长时建模产生影响,从而导致网络性能下降.考虑到终端用户的长期行为在短时分辨率下表现为趋势,因此平滑策略是有效的.自回归综合移动平均(ARIMA)可以基于移动平均操作对时间序列的总体趋势进行建模.正如预期的那样,ARIMA 的优势在 1 周和 1 d 的长时周期内显现,而支持向量回归器

(SVR)和神经网络等非线性映射方法在 1 h 和 1 min 的短周期内减少了误差.与最新的机器学习方法和深度学习方法相比,本文提出的方法在所有时间分辨率下取得了最好的性能.

为了更深入地验证所提出模型的性能,本文还对所提出的多头注意力机制进行了消融实验,实验结果如图 6 所示.可以看出,相较于原始的 CNN-LSTM(单头),多头注意力机制能够显著地提高电力消耗预测的准确性.

### 3 结束语

本文提出一种基于多头注意力的深度学习模型来预测电力消耗.在解决电力消耗建模问题的基础

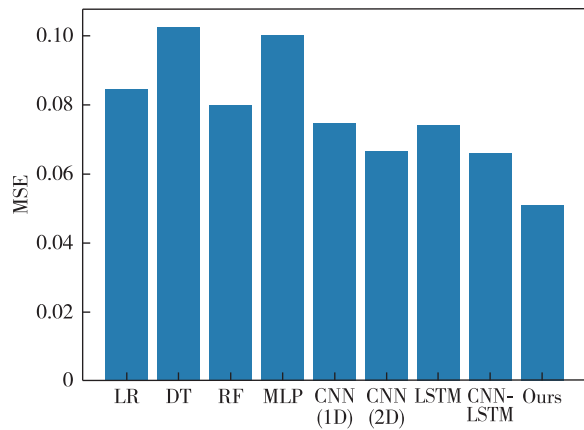


图5 不同算法的10-折交叉验证结果

Fig. 5 The 10-fold cross validation results of different algorithms

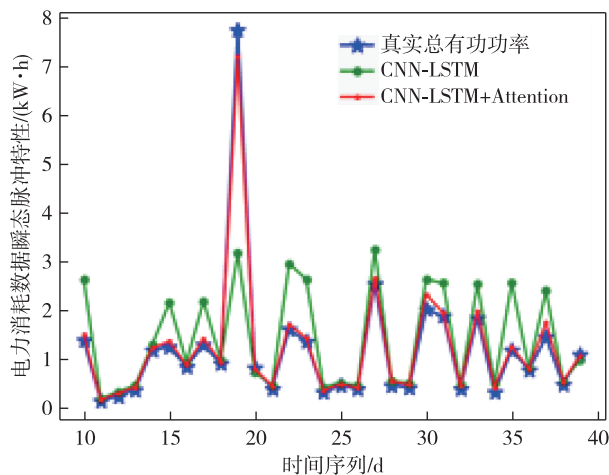


图6 多头注意力机制的消融实验结果

Fig. 6 Ablation experimental results of multi-headed attention mechanism

上,回顾了基于深度学习方法的电力消耗预测模型,提出了用于提取时空特征的 CNN-LSTM 模型和可学习加权的多头注意力模型,并在不同的时间条件下对模型进行了评估。

此外,长期行为之外的事件可以认为是可能发生的事件,而未能正确预测它们的情况可以归结为异常值。在未来的工作中,我们将通过聚合算法来解决这个问题,消除这种错误预测的影响,并利用混合方法提供的灵活性将其处理在可容忍的误差范围内。

## 参考文献

### References

[ 1 ] 陈冬洋,郑舟,吴永峰.基于时间序列的电力系统短期负荷预测问题分析[J].自动化应用,2017(11):

99-101  
 [ 2 ] Nejat P, Jomehzadeh F, Taheri M M, et al. A global review of energy consumption, CO<sub>2</sub> emissions and policy in the residential sector (with an overview of the top ten CO<sub>2</sub> emitting countries) [J]. Renewable and Sustainable Energy Reviews, 2015, 43: 843-862  
 [ 3 ] Chen D F, Zheng Z, Wu Y F. Analysis of power system short-term load forecasting based on time series [J]. Automation Application, 2017 (11): 99-101  
 [ 4 ] Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and machine learning forecasting methods: concerns and ways forward [J]. PLoS One, 2018, 13(3): e0194889  
 [ 5 ] González-Briones A, Hernández G, Corchado J M, et al. Machine learning models for electricity consumption forecasting: a review [C] // 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS). May 1-3, 2019, Riyadh, Saudi Arabia. IEEE, 2019: 1-6  
 [ 6 ] 张亚宁.基于 LSTM 模型的短期负荷预测[J].科技与创新,2018(12):18-21  
 [ 7 ] 陈卓,孙龙祥.基于深度学习 LSTM 网络的短期电力负荷预测方法[J].电子技术,2018,47(1):39-41  
 CHEN Zhuo, SUN Longxiang. Short-term electrical load forecasting based on deep learning LSTM networks [J]. Electronic Technology, 2018, 47(1): 39-41  
 [ 8 ] Rajagopalan V, Ray A. Symbolic time series analysis via wavelet-based partitioning [J]. Signal Processing, 2006, 86(11): 3309-3320  
 [ 9 ] Lin J, Khade R, Li Y. Rotation-invariant similarity in time series using bag-of-patterns representation [J]. Journal of Intelligent Information Systems, 2012, 39(2): 287-315  
 [ 10 ] Tso G K F, Yau K K W. Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks [J]. Energy, 2007, 32(9): 1761-1768  
 [ 11 ] Ekonomou L. Greek long-term energy consumption prediction using artificial neural networks [J]. Energy, 2010, 35(2): 512-517  
 [ 12 ] 罗嘉,王乐豪,涂姗姗,等.基于 LSTM-BLS 的突发气象灾害事件中公众情感倾向分析[J].南京信息工程大学学报(自然科学版),2021,13(4):477-483  
 LUO Jia, WANG Lehao, TU Shanshan, et al. Analysis of public sentiment tendency in sudden meteorological disasters based on LSTM-BLS [J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2021, 13(4): 477-483  
 [ 13 ] 王子岳,邵曦.基于 S-LSTM 模型利用‘槽值门’机制的说话人意图识别[J].南京信息工程大学学报(自然科学版),2019,11(6):751-756  
 WANG Ziyue, SHAO Xi. Speaker intention recognition based on S-LSTM model and slot-gate [J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2019, 11(6): 751-756  
 [ 14 ] Li C D, Ding Z X, Zhao D B, et al. Building energy consumption prediction: an extreme deep learning approach [J]. Energies, 2017, 10(10): 1525  
 [ 15 ] Fan C, Sun Y J, Zhao Y, et al. Deep learning-based feature engineering methods for improved building energy

- prediction[J]. *Applied Energy*, 2019, 240:35-45
- [16] Mocanu E, Nguyen P H, Gibescu M, et al. Deep learning for estimating building energy consumption [J]. *Sustainable Energy, Grids and Networks*, 2016, 6:91-99
- [17] Hinton G. Deep belief networks[J]. *Scholarpedia*, 2009, 4(5):5947
- [18] Kong W C, Dong Z Y, Jia Y W, et al. Short-term residential load forecasting based on LSTM recurrent neural network [J]. *IEEE Transactions on Smart Grid*, 2019, 10(1):841-851
- [19] Shi H, Xu M H, Li R. Deep learning for household load forecasting: a novel pooling deep RNN[J]. *IEEE Transactions on Smart Grid*, 2018, 9(5):5271-5280
- [20] Bouktif S, Fiaz A, Ouni A L, et al. Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: comparison with machine learning approaches[J]. *Energies*, 2018, 11(7):1636
- [21] 殷豪,董朕,孟安波.基于结合混沌纵横交叉的粒子群算法优化极限学习机的短期负荷预测[J]. *计算机应用研究*, 2018, 35(7):2088-2091  
YIN Hao, DONG Zhen, MENG Anbo. Short-term load forecasting based on extreme learning machine optimized by particle swarm optimization integrated with chaotic crisscross optimization[J]. *Application Research of Computers*, 2018, 35(7):2088-2091
- [22] 牟宏,王春义,顾洁,等.基于支持向量-分位数回归的短期负荷预测及风险分析[J]. *电力科学与技术学报*, 2016, 31(3):51-56  
MOU Hong, WANG Chunyi, GU Jie, et al. Short-term load forecasting and risk analysis based on support vector-quantile regression[J]. *Journal of Electric Power Science and Technology*, 2016, 31(3):51-56
- [23] 庞昊,高金峰,杜耀恒.基于时间卷积网络分位数回归的短期负荷概率密度预测方法[J]. *电网技术*, 2020, 44(4):1343-1350  
PANG Hao, GAO Jinfeng, DU Yaoheng. A short-term load probability density prediction based on quantile regression of time convolution network[J]. *Power System Technology*, 2020, 44(4):1343-1350
- [24] Taieb S B. Machine learning strategies for multi-step-ahead time series forecasting[D]. Brussels, Belgium: Free University of Brussels, 2014:75-86
- [25] Kim J Y, Cho S B. Electric energy consumption prediction by deep learning with state explainable autoencoder[J]. *Energies*, 2019, 12(4):739
- [26] Ronao C A, Cho S B. Human activity recognition with smartphone sensors using deep learning neural networks [J]. *Expert Systems with Applications*, 2016, 59:235-244
- [27] Hebrail G, Barard A. Individual household electric power consumption data set [DS/OL]. [2021-10-18]. [archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption](https://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption)

## Power load prediction based on multi-headed attentional convolutional network

ZHENG Zheng<sup>1</sup> TAN Lei<sup>2</sup> ZHOU Nan<sup>1</sup> HAN Junwei<sup>1</sup> GAO Jing<sup>2</sup> WENG Ligu<sup>3</sup>

<sup>1</sup> State Grid Henan Economics Research Institute, Zhengzhou 450052

<sup>2</sup> Henan Jiuyu Tenglong Information Engineering Co., Ltd., Zhengzhou 450052

<sup>3</sup> Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science & Technology, Nanjing 210044

**Abstract** Predicting residential energy consumption is tantamount to forecasting a multivariate time series. A specific window for several sensor signals can extract various features to forecast the energy consumption by using a prediction model. However, it is still a challenging task because of irregular patterns inside including hidden correlations between power attributes. In order to extract the complicated irregular energy patterns and selectively learn the spatiotemporal features to reduce the translational variance between energy attributes, we propose a deep learning model based on the multi-headed attention with the convolutional recurrent neural network. Compared with the simple time series model, the proposed model uses convolution and weighting mechanism to model the local correlation between power attributes and active power. It exploits the attention scores calculated with softmax and dot product operation in the network to model the transient and impulsive nature of energy demand, predicting the instantaneous pulse power consumption effectively. Experiments with the dataset of University of California, Irvine (UCI) household electric power consumption consisting of a total 2,075,259 time-series show that the proposed model greatly improves the prediction accuracy compared to the state-of-the-art deep learning models.

**Key words** deep learning; attention mechanism; time-series forecasting; load prediction