

<http://bhxb.buaa.edu.cn> jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2024.0056

基于迁移学习和 TCN-BiGRU 的短期负荷预测

刘杰¹, 周博文¹, 田明², 韩轲^{3,*}

(1. 哈尔滨理工大学 黑龙江省模式识别与信息感知重点实验室, 哈尔滨 150080; 2. 中国电信股份有限公司黑龙江分公司, 哈尔滨 150040; 3. 哈尔滨商业大学 计算机与信息工程学院, 哈尔滨 150028)

摘要: 电力负荷预测对于电力系统的稳定运行有着重要意义。传统短期预测算法常使用线性回归模型进行负荷预测, 无法捕捉到复杂的负荷波动, 导致预测结果准确性受到限制, 因此, 提出一种基于迁移学习 (TL) 的时间卷积网络-双向门控循环单元 (TCN-BiGRU) 模型。采用迁移学习策略将相关性高的信息迁移到实验模型中, 利用 K-medoids 聚类算法对数据进行聚类分析, 通过并行卷积策略提取 TCN 不同尺度的特征, 利用时间注意力 (TA) 捕获相关信息, 结合 BiGRU 进一步提取 TCN 训练输出的非线性特征, 使用动态多群粒子群优化 (DMS-PSO) 算法对网络训练的超参数寻找最佳的超参数组合。实验结果表明: 相对于门控循环单元 (GRU), 所提 TL-TCN-BiGRU 算法的平均绝对误差 (MAE) 降低了 38.6%, 均方根误差 (RMSE) 降低了 40.7%, 平均绝对百分比误差 (MAPE) 降低了 30.4%, R^2 提升了 5.3%。

关键词: 短期负荷预测; 迁移学习; 时间卷积网络; K-medoids 聚类; 融合网络

中图分类号: V247; TP391

文献标志码: A **文章编号:** 1001-5965(2026)04-0995-10

电力负荷预测是电力系统稳定运行的基础, 准确的电力负荷预测对于维持居民生活稳定、降低电网运营成本具有重要意义^[1]。电力负荷数据通常会呈现周期性、季节性变化的规律, 根据电力负荷预测时间长短的不同, 可以将其分为短期负荷预测、中期负荷预测、长期负荷预测^[2-3]。其中, 短期负荷预测的经典算法包括移动平均法、季节性指数法、自回归移动平均模型、神经网络模型等^[4], 现阶段主要通过统计模型^[5]、机器学习^[6]、深度学习^[7]、集成学习^[8]等算法对短期电力负荷数据进行预测分析。

基于深度学习算法的数据预测主要有卷积神经网络 (convolutional neural network, CNN)^[9]、循环神经网络 (recurrent neural network, RNN) 和长短期记忆 (long short term memory, LSTM) 网络^[10-12]。近年来, Bai 等^[13]提出了时间卷积网络 (temporal convolutional network, TCN), 该网络具有良好的可扩展性和平移不变性, 但对输入序列长度较为敏感。

Chung 等^[14]提出了门控循环单元 (gated recurrent unit, GRU), 后来又在其基础上进行改进, 提出双向门控循环单元 (bidirectional gated recurrent unit, BiGRU), 该网络可以很好地解决梯度消失问题^[15], 但对于某些特定的序列模式, 其可能会出现过拟合的问题。相较于单一负荷预测, 多元复合预测能够考虑不同特征间的耦合关系, 使其具有更高的预测表现力^[16-18]。邹智等^[19]提出了一种基于贝叶斯优化的 CNN-BiGRU 网络进行电力负荷预测, 模型通过对数据中的多特征进行筛选, 将匹配度较高的数据特征送入到网络中, 但该模型结构较复杂, 对结果产生的原因难以解释。Huo 等^[20]将 TCN 和 LSTM 网络结合, 与 TCN 算法相比, 提高了评价参数的预测精度。在深度学习中, 选择 CNN 内部的隐层数和激活函数等超参数直接影响网络模型的性能表现^[21], 为防止网络出现过拟合或欠拟合, 选择适合

收稿日期: 2024-01-23; 录用日期: 2024-03-26; 网络出版时间: 2024-04-07 11:22

网络出版地址: link.cnki.net/urlid/11.2625.V.20240403.1036.002

基金项目: 黑龙江省自然科学基金 (LH2020F009)

* 通信作者. E-mail: thruster@163.com

引用格式: 刘杰, 周博文, 田明, 等. 基于迁移学习和 TCN-BiGRU 的短期负荷预测 [J]. 北京航空航天大学学报, 2026, 52 (4): 995-1004.

LIU J, ZHOU B W, TIAN M, et al. Short-term load forecasting based on transfer learning and TCN-BiGRU [J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52 (4): 995-1004 (in Chinese).

的优化算法对 CNN 内部的超参数进行优化是非常关键的。Li 等^[22]使用 K-medoids 算法进行多线性时间序列预测,实验结果表明,算法能有效反映数据的多线性趋势模式,对时间序列分析具有重要意义,但在处理高维数据时会令簇中心的选择变得困难,导致算法性能下降。欧阳福莲等^[23]提出一种多尺度 CNN-BiLSTM-Attention 模型,对 CNN 层进行扩展,通过串行叠加、并行拼接的方式加大网络层数,实现更高精度的预测效果,但多层的结构可能会增加模型参数调整的难度。迁移学习(transfer learning, TL)是一种从原模型学习到内在规律转移到新问题上的过程^[24]。目前,将迁移学习应用于负荷预测领域的研究较少。Jin 等^[25]通过迁移学习的方式将 CNN-GRU 训练模型从一个广泛的数据集模型转移到一个较小的数据集上,实验结果显示,迁移后的模型性能得到了显著提升,但对数据质量要求较高且易出现过拟合。

上述文献多是基于 CNN 或 RNN 的算法实现,在此基础上,本文提出了一种 TL-TCN-BiGRU 电力负荷预测模型,用来进一步解决预测精度不高及模型性能下降等问题。利用迁移学习完成源域与目标域数据间的匹配,同时使用 K-medoids 聚类分析数据特征,采用并行拼接的方式对 TCN 进行扩展,提升网络对模型的依赖关系,并引入时间注意力(temporal attention, TA)增强 TCN 特征提取能力,利用 TCN-BiGRU 进一步提升网络性能。通过动态多群粒子群优化(dynamic multigroup particle swarm optimization, DMS-PSO)算法优化模型超参数。为验证模型性能,在公开数据集上进行大量实验并与其他模型进行对比,结果证明了本文模型的有效性。

1 基本算法

1.1 迁移学习

迁移学习主要针对数据集缺失或小样本数据集问题,在迁移时对源域数据与目标域数据进行相似度匹配,达到某一目标后完成迁移操作。在匹配过程中,为减小源域数据集和目标域数据集之间的分布差异,目前主要使用最大均值差异(maximum mean discrepancy, MMD)来衡量 2 个样本之间不同但相关的分布距离,其定义为

$$\text{MMD}(U, V) = \left\| \frac{1}{n} \sum_{i=1}^n \phi(u_i) - \frac{1}{m} \sum_{j=1}^m \phi(v_j) \right\|^2 \quad (1)$$

式中: u_i 和 v_j 为 2 个样本; $\phi(\cdot)$ 表示映射,用于把原变量映射到再生核希尔伯特空间中; m 和 n 为样本的取值范围; $\text{MMD}(U, V)$ 表示 U 和 V 样本间的最大均值差异。

如果 2 个样本间最大均值差异足够小,则认为

两者分布相同;反之,则认为两者分布不同。

1.2 皮尔逊相关系数

PCC 是用于衡量 2 个变量间线性关系强度的统计指标,其取值范围被限定在 $[-1, 1]$ 之间,结果绝对值接近于 1 时,表示两者之间的相关度高,计算式如下:

$$r(A, B) = \frac{\text{Cov}(A, B)}{\sqrt{\text{Var}[A] \text{Var}[B]}} \quad (2)$$

式中: $\text{Cov}(A, B)$ 表示 2 个变量之间的协方差; $\text{Var}[A]$ 表示变量 A 的方差, $\text{Var}[B]$ 表示变量 B 的方差; $r(A, B)$ 表示 2 个变量的关联度。

本文通过 PCC 绘制特征相关热力图,可更直观地看出不同影响因素间的相关性。

1.3 K-medoids 聚类

K-medoids 聚类是一种基于中心点的聚类算法,该算法选择簇中的一个实际样本作为中心点,而不是使用均值,公式表示为

$$c = \sum_{C_i} \sum_{P_i \in C_i} |P_i - C_i| \quad (3)$$

式中: C_i 为质心; P_i 为非质心; c 为每个非质心点到质心距离的和。

1.4 双向门控循环单元

BiGRU 通过使用 2 个 GRU 层来解决传播方向问题。前向 GRU 从开始到结束处理输入序列,而后向 GRU 以相反方向处理。BiGRU 相较于 GRU,具有双向建模能力、更好的长期依赖建模、更准确的序列标注及强大的特征提取能力等优点,其内部隐层单元计算式如下:

$$\begin{cases} \vec{h}_t = \text{GRU} \left(i_t, \vec{h}_{t-1} \right) \\ \overleftarrow{h}_t = \text{GRU} \left(i_t, \overleftarrow{h}_{t-1} \right) \\ h_t = f \left(W_{\vec{h}_t} \vec{h}_t + W_{\overleftarrow{h}_t} \overleftarrow{h}_t + b_t \right) \end{cases} \quad (4)$$

式中: i_t 为 t 时刻输入; \vec{h}_t 和 \overleftarrow{h}_t 分别为 t 时刻隐藏层前向和后向状态; $W_{\vec{h}_t}$ 和 $W_{\overleftarrow{h}_t}$ 分别为 t 时刻隐藏层前向和后向权重; b_t 为 t 时刻隐藏层状态的偏置。

1.5 动态多群粒子群优化算法

传统的 PSO 算法模仿鸟类的捕食过程,利用速度和位置来标记和更新信息,其更新公式如下:

$$v_i^{(t+1)} = \omega v_i^{(t)} + c_1 r_1 (p_{\text{best}} - z_i^{(t)}) + c_2 r_2 (g_{\text{best}} - z_i^{(t)}) \quad (5)$$

$$z_i^{(t+1)} = z_i^{(t)} + v_i^{(t+1)} \quad (6)$$

式中: p_{best} 为个体最优位置; g_{best} 为全局最优位置; c_1 、 c_2 为加速系数; ω 为惯性权重; $v_i^{(t)}$ 和 $z_i^{(t)}$ 分别为第 i 个粒子的速度和位置; r_1 、 r_2 为 $[0, 1]$ 内的随机变量。

从式(5)可以看出,惯性权重 ω 对粒子的收敛性能具有较大影响。当 ω 较大时,粒子全局搜索能力更强;当 ω 较小时,粒子局部搜索更敏锐。本文

在原有基础上对代码结构进行了优化, 采用 DMS-PSO 算法, 利用动态群组最优位置替换全局最优位置, 其通过引入多个可以动态配置自身的群来解决 PSO 寻找最佳解决方案的限制。改进后的 PSO 算法将式(5)更新为

$$v_i^{(t+1)} = \omega v_i^{(t)} + c_1 r_1 (p_{best} - z_i^{(t)}) + c_2 r_2 (l_{best} - z_i^{(t)}) \quad (7)$$

式中: l_{best} 为动态群组的最佳位置。

更新后的 $v_i^{(t+1)}$ 通过式 (6) 对粒子的位置进行更新, 若找到更好的位置, 则更新 p_{best} 和 l_{best} 。DMS-PSO 算法每经过 R 代, 则随机组合新粒子, R 称为重组周期, 本文实验设置为 10。本文实验设置 DMS-PSO 算法初始参数粒子数 $N=20$, 惯性权重 ω 范围为 $[0.4, 0.9]$, 加速系数 $c_1=c_2=2$ 。

DMS-PSO 算法的每个群体都专注于搜索空间的特定区域, 并保持自己的本地最佳和全局最佳。群体定期交换信息, 以共享知识并适应问题环境的变化, 防止陷入局部最优解。每个动态群组中的粒子通过计算适应度来确定搜寻的最优值, 其适应度目标函数为

$$F_{fix} = \frac{1}{H} \sum_{j=1}^H \sum_{t=1}^L (d_t^j - \hat{d}_t^j)^2 \quad (8)$$

式中: F_{fix} 为适应度值; H 为训练样本数; j 为迭代轮次; L 为训练终止时间; d_t^j 和 \hat{d}_t^j 分别为模型迭代时的真实输出和目标输出。

DMS-PSO 算法通过式 (8) 计算粒子适应度函数并更新, 直至满足终止条件后输出此时的最优解。

2 TL-TCN-BiGRU 模型

2.1 问题描述

假设某地区一年内的电力负荷采样序列为 $\{x_1, x_2, x_3, \dots, x_n\}$, 采样间隔为 x (单位: min), 每日的

采样点数为 $\tau=24(60/x)$, 将这些采样点构建为一个含有 H 个训练数据的样本集合 $D=\{(X^{(1)}, Y^{(1)}), (X^{(2)}, Y^{(2)}), \dots, (X^{(H)}, Y^{(H)})\}$ 。其中, $X=[X^{(1)}, X^{(2)}, \dots, X^{(H)}] \in \mathbf{R}^{d \times T}$ 为 T 时刻下包含历史负荷数据和相关影响因素等 d 维度数据信息, $Y=[Y^{(1)}, Y^{(2)}, \dots, Y^{(H)}] \in \mathbf{R}^{\tau}$ 为对应 X 未来 τ 个采样点的真实负荷值, 本文旨在通过构建一个模型 F , 对其中的关键信息进行提取, 来预测未来电力负荷值, 公式表示为

$$F(X, \theta) \rightarrow y \quad (9)$$

式中: θ 为模型中的参数; X 为历史数据信息; y 为未来 η 时刻的电力负荷值。

2.2 TA-TCN 结构设计

本文提出的 TA-TCN 相较 TCN 的主要优势在于引入了时间注意力机制, 可以更好地聚焦当前隐藏状态信息, 并重新计算时间序列信息向量中各隐藏状态的权重。信息向量通过使用编码器中所有神经元的隐藏状态计算得到, 采用注意力机制, 使用加权计算方式, 分别传递给编码器中的每个神经元。利用该结构, 在进行短期电力负荷数据预测时, 可以根据时间序列数据自适应地学习每个时间步的重要特征, 通过时间注意力来确定哪些特征对当前时间步的预测最为重要, 这使得该结构可以更好地处理时间序列中的噪声和异常值, 帮助模型更好地利用时间信息, 并处理长期依赖问题, 从而提高模型的表达能力和预测精度。此外, 通过并行拼接处理后的 TCN 可以捕获不同尺度的特征, 提高系统资源利用率, 增加系统的可扩展性, 从而提高模型对不同时间尺度的表达能力, 其结构设计如图 1 所示。图中, Kernel1、Kernel2、Kernel3 表示不同大小的卷积核, 其大小分别为 1、4、7, Kernel1 用于数据升维。原始的 TCN 先通过不同长度的卷积单元来提取不同时序尺度的特征, 再通过并行拼接

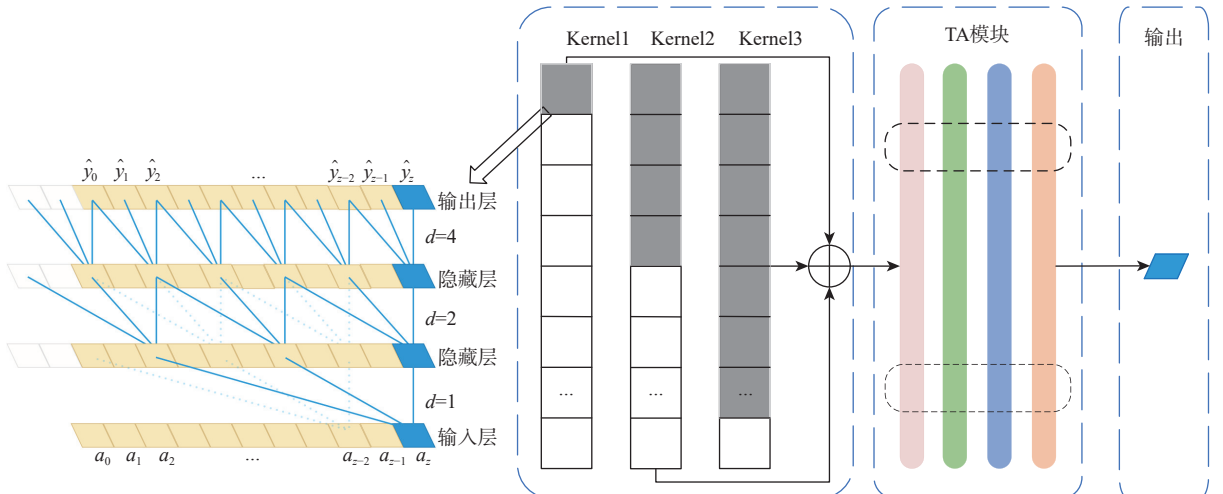


图 1 TA-TCN 结构设计

Fig. 1 TA-TCN structural design

策略对特征进行融合,并输送到改进的TA模块,其对接收到的多维特征进行信息提取,并将提取到的结果进行输出。

2.3 模型介绍

本文针对传统短期电力负荷预测算法在准确性、适应性和时效性等方面存在的问题,提出了一种TL-TCN-BiGRU电力负荷预测算法,算法整体框架如图2所示。首先,要对获取到的源域数据进行多特征分析,通过关联规则挖掘不同数据特征间的关系,并通过PCC对特征因子进行特征选择;其次,将与目标域相似的源域数据选为待迁移目标,

传递到输入样本并对其进行数据预处理,预处理时需要删除异常值、填入缺失值、量化数据以减少数据量纲不同对网络性能的影响;再次,通过K-medoids算法对数据进行聚类并确定聚类数,以更好地捕获输入数据的分布规律;然后,CNN将接收到的数据依次传入TA-TCN层、BiGRU层,通过TA-TCN结构捕获不同时间尺度的特征以提高计算效率,通过BiGRU结构更好地捕捉时间序列数据中的双向依赖关系,并通过DMS-PSO算法优化网络内部的超参数;最后,在训练过程结束后,模型会对其训练得到的参数和权重进行更新,并完成迭代,输出预测结果。

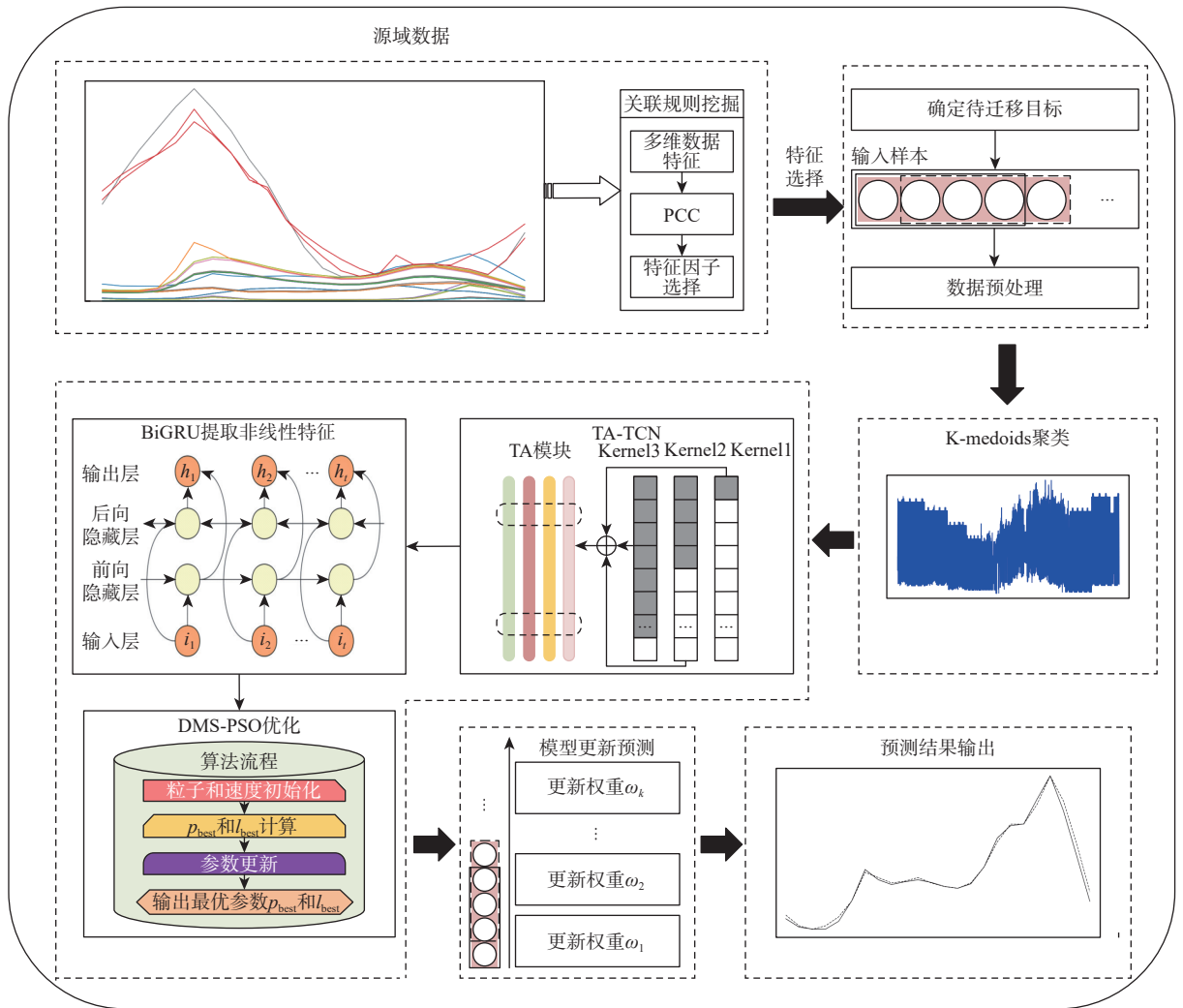


图2 TL-TCN-BiGRU模型整体框架

Fig. 2 Overall framework of TL-TCN-BiGRU model

3 实验

3.1 实验数据集及环境

数据集选用美国住宅负荷数据,包含美国各州若干地区一年内的住宅建筑电力消耗,根据建筑高低的不同分为高、中、低3个分段。实验中,将1月1日至5月31日的数据作为训练集,将6月

1日至6月30日的数据作为测试集。

实验源域数据集选择加利福尼亚州中等高度建筑下的每小时住宅负荷数据,包括照明设备、热水设备、暖通空调等用电器的电力消耗数据及住宅整体电力消耗数据,该州共包含73个地点的住宅负荷数据,选择合适地点对其负荷数据进行融合,并迁移到实验目标域中。实验目标域数据集选用洛杉矶市中等高度建筑下的每小时住宅负荷数据。

经式 (1) 计算最大均值差异, 得到 2 个数据集间分布距离为 0.023 8, 说明两者数据相似度较高。匹配完成后, 经最大最小归一化对数据量化处理。TCN 层数通过控制变量法确定为 4, 在每个 BiGRU 层后通过 Dropout 防止模型过拟合, 最终经全连接层对数据信息进行汇集后, 输出未来 1 d 内的负荷预测值。

实验采用 Windows10 操作系统, CPU 选择 Intel Core i7-7700HQ, 显卡选择 NVIDIA GeForce GTX 1050, 内存选择 16 GB RAM DDR4@2 400 MHz, 所提模型构建在 Python (3.9.12)、Tensorflow2.0 框架及 Keras 框架中运行。

3.2 评价指标

使用平均绝对误差 (mean absolute error, MAE)、均方根误差 (root mean square error, RMSE)、平均绝对百分比误差 (mean absolute percentage error, MAPE)、拟合优度 R^2 等指标来评估模型性能, 相应的计算公式如下:

$$e_{ma} = \frac{1}{n} \sum_{i=1}^n |y_i - \tilde{y}_i| \quad (10)$$

$$e_{rms} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2} \quad (11)$$

$$e_{map} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \tilde{y}_i}{y_i} \right| \quad (12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (13)$$

式中: n 为样本数量; y_i 为数据真实值; \tilde{y}_i 为数据预测值; \bar{y} 为数据真实值的平均值。

MAE、RMSE、MAPE 的数值越小, R^2 的结果越接近 1, 代表模型的预测准确度越高。

3.3 多特征分析

本节对目标域数据集中的多特征进行分析。通过 PCC 分析特征间的相关性, PCC 是一种常用且有效的统计量, 表征 2 个变量之间的关系。由于选择的住宅地点缺失天气数据, 对数据集其余特征 (如室内外照明设备、室内混杂设备等) 进行分析, 实验结果如图 3 所示, 图中, 数字表示皮尔逊相关系数。

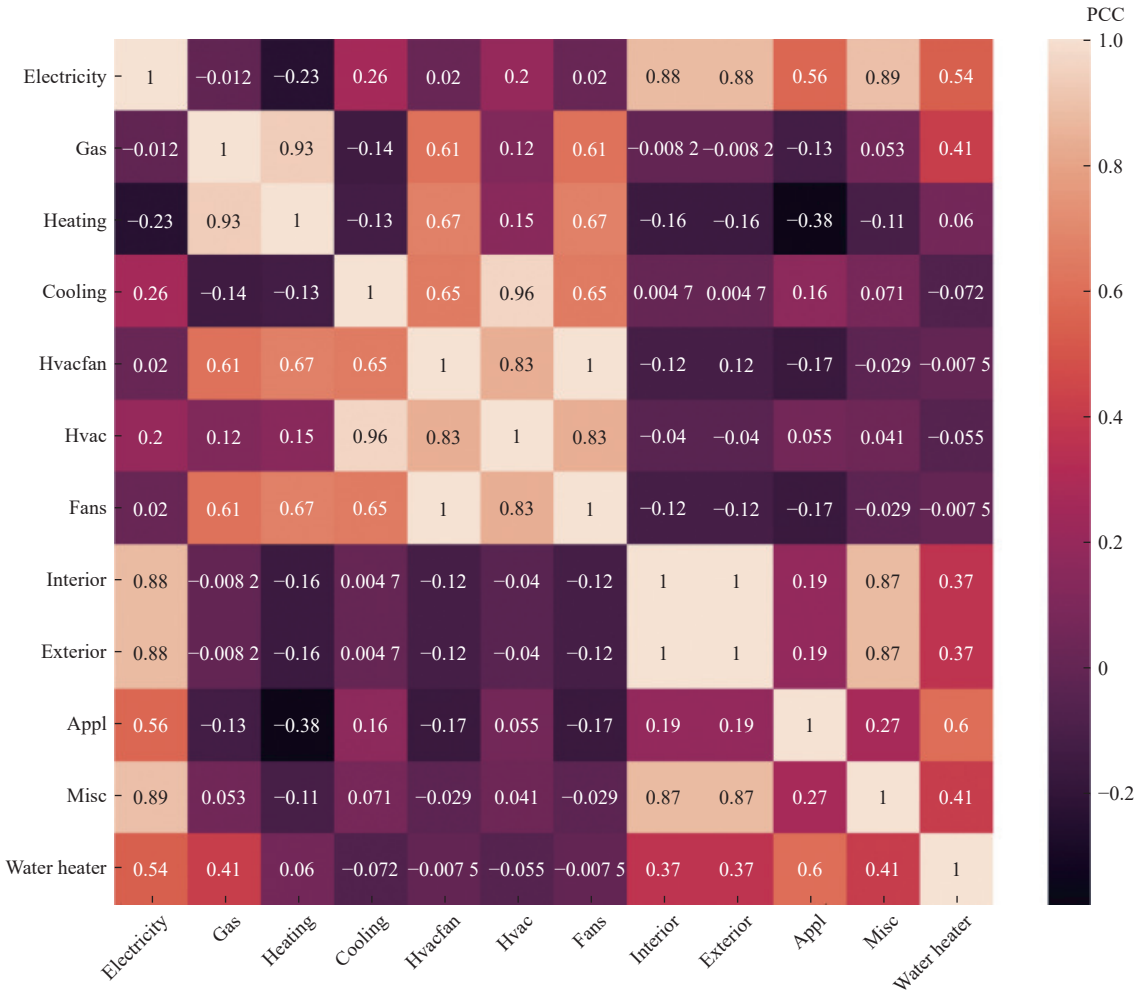


图 3 特征相关性热力图

Fig. 3 Characteristic correlation heat map

图3为特征相关性热力图,其数值越接近1,说明特征间相关性越强。图3中删除了部分缺失数据,观察第1列可以看出,住宅电力总负荷与室内外的照明设备及室内的混杂设备相关性较高,因此,考虑将这些特征与住宅电力总负荷特征进行融合并输入到CNN中进行训练。此外,实验还通过one-hot编码引入季节和日期因素,进一步检验不同因素对电力负荷特征的影响情况。

3.4 聚类分析

聚类分析是典型的无监督学习,用于对未知类别的样本进行划分,按照一定的规则将不相似的样本分为不同类簇,从而揭示样本之间内在的性质及相互之间的联系规律。通过误差平方和(sum of squared error, SSE)确定聚类数,评价公式如下:

$$I_{SSE} = \sum_{i=1}^k \sum_{J_i \in R_i} |e(J_i, K_i)|^2 \quad (14)$$

式中: R_i 为类别 i 中数据点集合; k 为簇的个数; J_i 为数据点; K_i 为每类数据的样本中心; $e()$ 为误差函数,衡量数据点和样本中心的欧氏距离。

实际应用中,应同时考虑两者的取值,选取SSE较小且下降速度较快的点对应的作为确定的聚类数。本文对输入数据进行测试分析,得到的聚类结果如图4所示。

由图4可知,当 k 为4时,图4(a)中拐点降幅较大,附近不存在突变拐点,且趋近于最小值,图4(b)中轮廓系数数值较高,同时考虑到过多的聚类数会增加算法的计算复杂度,因此,最适宜的聚类数为4个。对归一化后的数据进行聚类,结果如图5所示。

根据聚类结果将输入的数据信息划分为4类,在每个类别中将该类别下的历史负荷数据送入模型内部进行单独训练。通过对数据进行聚类划分,构建相似日集合,可以更好地提取数据中的相似信

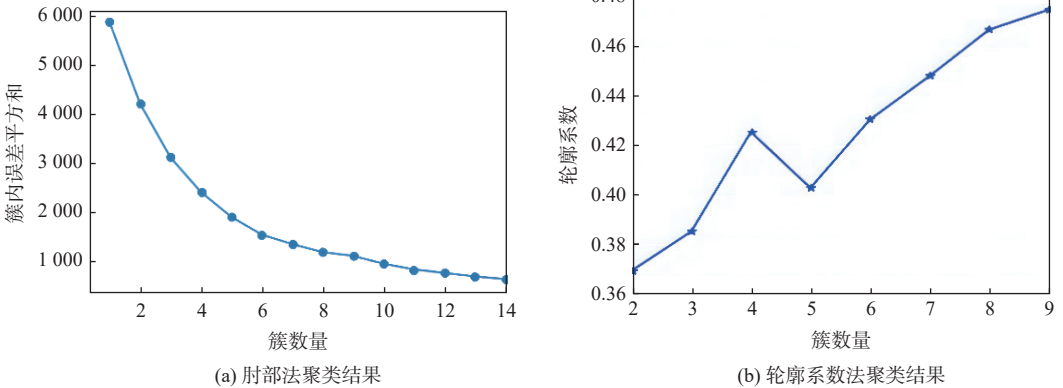


图4 不同簇数量下SSE和轮廓系数的变化曲线

Fig. 4 Curves of SSE and silhouette coefficient under different cluster numbers

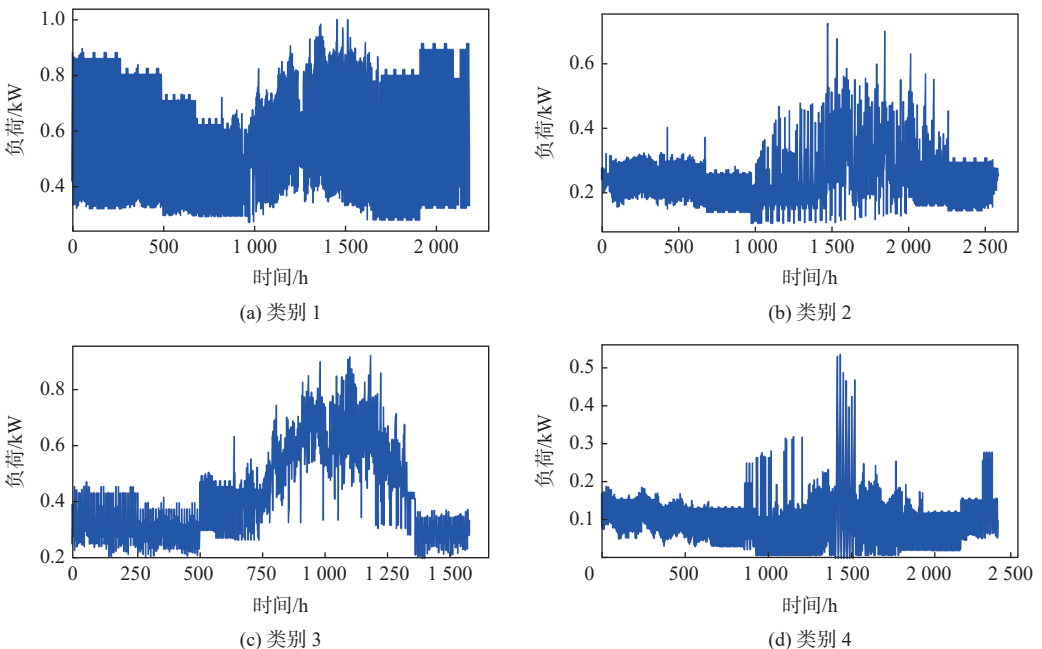


图5 聚类数为4时的负荷曲线

Fig. 5 Load profile for a clustering number of 4

息, 进而有效提升集合内数据预测的准确性。

3.5 实验结果与分析

3.5.1 实验参数确定

以历史 10 h 的电力负荷数据作为输入送入 TCN 层。TCN 层数设置为 4 层, TCN 改变模型尺寸大小输送到 Attention 层进行特征提取及注意力分配。Attention 层输出后送入 BiGRU 中。经实验测试, 4 层的 BiGRU 可以在学习网络特征的同时获得更好的拟合效果及更低的实验误差。本文选用 Adam 优化器, 激活函数选用 ReLU 函数, TCN 膨胀系数经对比实验选择 1、2、4、8, CNN 中的超参数, 如 BiGRU 隐藏层神经元数量、卷积核大小、Dropout、学习率、批处理大小、迭代次数均通过改进的 PSO 算法进行优化。

本文通过 DMS-PSO 算法对实验中相关的超参数进行优化, 经搜寻后得到的最优结果如表 1 所示。可以看出, 经 DMS-PSO 算法优化后, 模型具有更明确的参数选择, 可以更高效地利用计算资源, 大大提高了训练效率, 在实际运行效果上具有更好的表现。

表 1 DMS-PSO 算法寻优结果

Table 1 DMS-PSO algorithm optimisation results

寻优参数	范围	寻优结果
BiGRU 隐藏层神经元数量	[1,50]	31
卷积核大小	[3,5]	4
Dropout	[0,0.5]	0.01
学习率	[0.000 1,0.01]	0.001
批处理大小	[16,32]	20
迭代次数	[100,200]	121

3.5.2 有效性分析

为说明本文 K-medoids 聚类算法的有效性, 通过 DB 指数和 CH 指数对聚类效果进行评定, 其计算式分别为

$$I_{DB} = \frac{1}{g} \sum_{i=1}^g \max_{j=1, \dots, g, j \neq i} \frac{\alpha_i + \alpha_j}{\|q_i - q_j\|_2^2} \quad (15)$$

$$I_{CH} = \frac{\text{Tr}(B_g)}{\text{Tr}(W_g)} \cdot \frac{n-g}{g-1} \quad (16)$$

式中: α_i 和 α_j 分别为类别 i 和 j 下样本的平均距离;

$\|q_i - q_j\|_2^2$ 为类别 i 和 j 的类中心欧氏距离; n 为总样本数; g 为聚类数; $\text{Tr}(B_g)$ 为 g 类的类间离差矩阵的迹; $\text{Tr}(W_g)$ 为 g 类的类内离差矩阵的迹。

I_{DB} 指数越小, I_{CH} 指数越大, 说明聚类效果越好。将本文算法与 K-means、DBSCAN 进行效果对比, 结果如表 2 所示。

表 2 不同聚类算法指标对比

Table 2 Comparison of indicators of different clustering methods

聚类算法	DB指数	CH指数
DBSCAN	1.459	805.715
K-means	1.368	990.687
K-medoids	1.231	1 083.451

为说明本文 DMS-PSO 算法的优越性, 在使用 K-medoids 算法的基础上, 将其与遗传算法 (genetic algorithm, GA)、PSO 算法进行对比, 结果如表 3 所示。

表 3 不同优化算法效果对比

Table 3 Comparison of effect of different optimization algorithms

聚类算法	MAE/kW	MAPE/%
GA	0.092	0.702 3
PSO	0.075	0.523 6
DMS-PSO	0.035	0.418 3

从表 2 和表 3 可以看出采用 K-medoids 算法和 DMS-PSO 算法的优势, 为进一步说明本文模型各部分模块的有效性, 逐一对 TA 模块、BiGRU 模块、DMS-PSO 模块、K-medoids 模块、TL 模块的作用效果进行验证。实验以相同的评价指标 MAE、RMSE、MAPE、 R^2 作为判别依据, 采用单维负荷数据特征对测试集某日 24 h 电力负荷数据进行预测, 实验结果如表 4 所示。

实验测试选用测试集预测结果中 1 d 的数据, 从表 4 可以看出, 相较于结合 TA 及 BiGRU 的算法, 直接用 TCN 进行预测的结果精度更低, 说明设计的 TA 结构可以更好地关注数据的时间步长, 且 BiGRU 能够有效减弱数据的随机性和非线性干扰, 从而提高预测的精确度。结果显示, 本文提出的各

表 4 消融实验结果

Table 4 Ablation experiment results

TCN	TA结构	BiGRU	DMS-PSO	K-medoids聚类	TL模块	MAE/kW	RMSE/kW	MAPE/%	R^2
√						0.065	0.099	0.743 5	0.908
	√					0.062	0.091	0.717 4	0.922
	√	√				0.059	0.088	0.670 9	0.928
	√	√	√			0.052	0.080	0.579 5	0.940
	√	√	√	√		0.046	0.072	0.506 1	0.964
√	√	√	√	√	√	0.035	0.054	0.418 3	0.972

模块均对模型具有一定增益。本文模型与TCN相比, MAE降低46.2%, RMSE降低45.5%, MAPE降低43.7%, R^2 提升7.0%。本文提出的各模块都对模型整体效果的提升具有各自的优势,从而发挥模型整体的性能。

3.5.3 不同模型的预测结果

将本文模型与其他模型(BP、SVR、LSTM、GRU、TCN)进行效果对比,对不同模型采用相同的数据处理方式,实验结果如表5和图6所示。

表5 不同模型评价结果

Table 5 Evaluation results of different models

模型	MAE/kW	RMSE/kW	MAPE/%	R^2
BP	0.066	0.096	0.799 5	0.914
SVR	0.076	0.101	1.166 9	0.905
LSTM	0.058	0.097	0.624 4	0.912
GRU	0.057	0.091	0.600 9	0.923
TCN	0.065	0.099	0.743 5	0.908
本文模型	0.035	0.054	0.418 3	0.972

从图6可以看出,本文模型各项指标都具有最优的表现。结合表5可得,与次优模型GRU相比,本文指标MAE降低了38.6%, RMSE降低了40.7%, MAPE降低了30.4%, R^2 提升了5.3%。

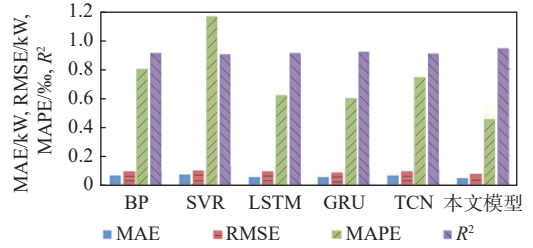
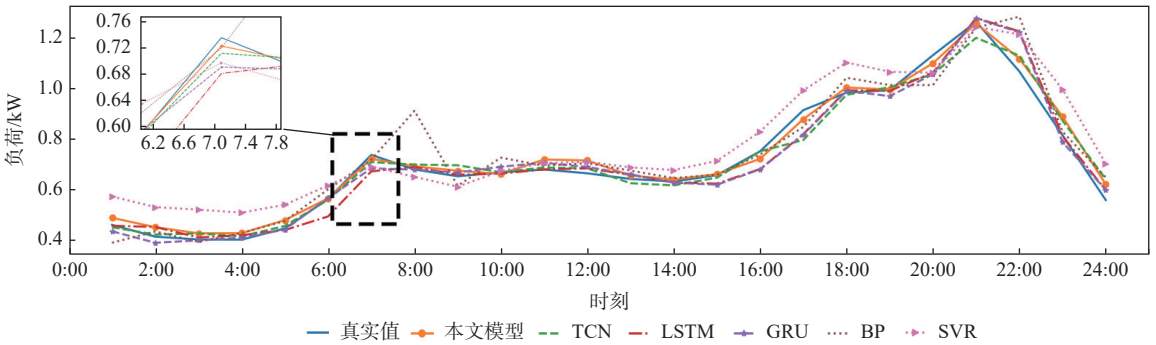


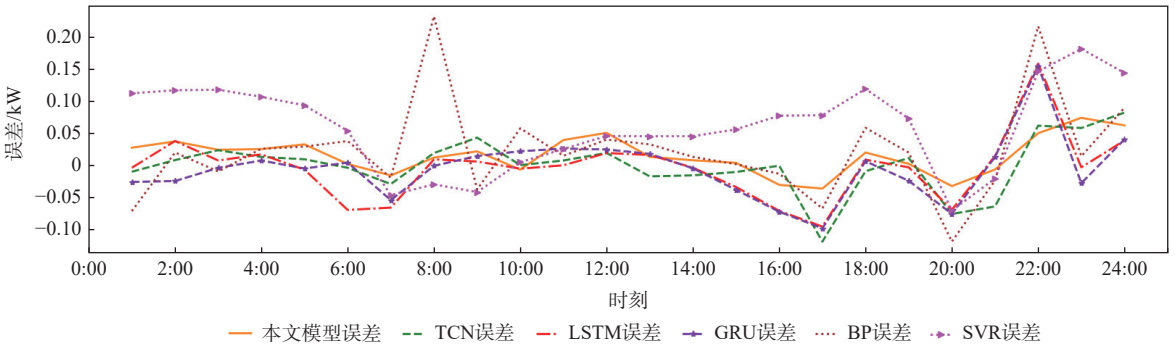
图6 不同模型评价结果

Fig. 6 Different model evaluation result graphs

此外,本文还对测试集预测结果中1 d的电力负荷数据进行提取,并与其他模型(BP、SVR、LSTM、GRU、TCN)在同一时间内的负荷分布曲线进行对比,如图7所示。图7(a)为电力数据在1 d内的波动情况。凌晨时刻,用户耗电量较低,早上耗电量增加,在7:00左右达到小高峰,又逐渐降低,傍晚时刻用户用电需求增多,导致实际用电量有所增加,预测曲线也随之上升。从图7(b)可以看出,本文模型预测结果平均误差为0.026 kW,相较其他模型最大平均误差减小了0.052 kW,与真实值之间数据偏离最小。结合图7可知,本文算法在整体预测方面更接近真实值,并且具有更高的预测精度和性能。



(a) 不同模型预测曲线



(b) 误差分布曲线

图7 不同模型24 h预测结果

Fig. 7 24 h prediction results of different models

4 结论

本文提出了一种基于迁移学习的TCN-BiGRU

电力负荷预测模型,在美国住宅负荷数据集上进行实验分析,得出以下结论:

- 1) 引入迁移学习算法,对与寻找的目标域数据

相匹配的源域数据进行模型迁移训练,可以解决数据稀缺问题,提高模型性能。

2) 通过并行拼接的TCN对输入的数据进行处理,可以提高模型的多尺度特征提取能力;利用时间注意力对特征进行权重分配具有实际效果;利用BiGRU可以解决非线性时序数据处理问题;利用DMS-PSO算法对网络内部的超参数进行优化,可以进一步提升模型的泛化性能。

3) 与传统的机器学习模型相比,实验结果显示,本文模型在预测时具有更低的误差损失和更高的拟合优度。

此外,本文实验仅在公开的电力负荷数据集上进行,在未来的工作中,会进一步对模型的结构和功能进行细化,考虑融入多步预测,并对不同领域的数据进行分析,提高模型的整体性能。

参考文献 (References)

- [1] 刘倩倩,刘钰山,温焯婷,等.基于PCC-LSTM模型的短期负荷预测方法[J].北京航空航天大学学报,2022,48(12):2529-2536.
LIU Q Q, LIU Y S, WEN Y T, et al. Short-term load forecasting method based on PCC-LSTM model[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(12): 2529-2536 (in Chinese).
- [2] LI X M, GUO X C, LIU L N, et al. A novel seasonal grey model for forecasting the quarterly natural gas production in China[J]. Energy Reports, 2022, 8: 9142-9157.
- [3] 王艳松,申晓阳,李强,等.基于PCA-GRD-LWR模型的海上油田中长期最大电力负荷预测[J].中国石油大学学报(自然科学版),2023,47(2):129-135.
WANG Y S, SHEN X Y, LI Q, et al. Forecasting of medium and long-term maximum power load for offshore oilfields based on PCA-GRD-LWR model[J]. Journal of China University of Petroleum (Edition of Natural Science), 2023, 47(2): 129-135(in Chinese).
- [4] 李丹,孙光帆,缪书唯,等.基于多维时序信息融合的短期电力负荷预测方法[J].中国电机工程学报,2023,43(S1):94-106.
LI D, SUN G F, MIAO S W, et al. A short-term power load forecasting method based on multidimensional temporal information fusion[J]. Proceedings of the CSEE, 2023, 43(S1): 94-106(in Chinese).
- [5] ZHANG R X, ZHU Z Y, YUAN M, et al. Regional residential short-term load-interval forecasting based on SSA-LSTM and load consumption consistency analysis[J]. Energies, 2023, 16(24): 8062.
- [6] BACANIN N, STOEAN C, ZIVKOVIC M, et al. On the benefits of using metaheuristics in the hyperparameter tuning of deep learning models for energy load forecasting[J]. Energies, 2023, 16(3): 1434.
- [7] AL-JAMIMI H A, BINMAKHASHEN G M, WORKU M Y, et al. Advancements in household load forecasting: deep learning model with hyperparameter optimization[J]. Electronics, 2023, 12(24): 4909.
- [8] WANG J Z, WANG K, LI Z W, et al. A multitask integrated deep-learning probabilistic prediction for load forecasting[J]. IEEE Transactions on Power Systems, 2024, 39(1): 1240-1250.
- [9] TRAN T N. Grid search of convolutional neural network model in the case of load forecasting[J]. Archives of Electrical Engineering, 2021, 70(1): 25-36.
- [10] ESKANDARI H, IMANI M, MOGHADDAM M P. Convolutional and recurrent neural network based model for short-term load forecasting[J]. Electric Power Systems Research, 2021, 195: 107173.
- [11] CAICEDO-VIVAS J S, ALFONSO-MORALES W. Short-term load forecasting using an LSTM neural network for a grid operator[J]. Energies, 2023, 16(23): 7878.
- [12] 李艳波,尹轸,陈俊硕,等.结合改进残差网络和Bi-LSTM的短期电力负荷预测[J].哈尔滨工业大学学报,2023,55(8):79-86.
LI Y B, YIN P, CHEN J S, et al. Short-term power load forecasting based on combination of residual network and Bi-LSTM[J]. Journal of Harbin Institute of Technology, 2023, 55(8): 79-86(in Chinese).
- [13] BAI S J, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[EB/OL]. (2018-04-19)[2024-01-10]. <https://arxiv.org/abs/1803.01271>.
- [14] CHUNG J, ÇAGLAR G, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[EB/OL]. (2014-12-11)[2024-01-10]. <https://arxiv.org/abs/1412.3555>.
- [15] XU H S, FAN G L, KUANG G F, et al. Construction and application of short-term and mid-term power system load forecasting model based on hybrid deep learning[J]. IEEE Access, 2023, 11: 37494-37507.
- [16] 吴晨,姚菁,薛贵元,等.基于MMoE多任务学习和长短时记忆网络的综合能源系统负荷预测[J].电力自动化设备,2022,42(7):33-39.
WU C, YAO J, XUE G Y, et al. Load forecasting of integrated energy system based on MMoE multi-task learning and LSTM[J]. Electric Power Automation Equipment, 2022, 42(7): 33-39(in Chinese).
- [17] 杨国华,郑豪丰,张鸿皓,等.基于Holt-Winters指数平滑和时间卷积网络的短期负荷预测[J].电力系统自动化,2022,46(6):73-82.
YANG G H, ZHENG H F, ZHANG H H, et al. Short-term load forecasting based on Holt-Winters exponential smoothing and temporal convolutional network[J]. Automation of Electric Power Systems, 2022, 46(6): 73-82(in Chinese).
- [18] 叶林,李奕霖,裴铭,等.寒潮天气小样本条件下的短期风电功率组合预测[J].中国电机工程学报,2023,43(2):543-555.
YE L, LI Y L, PEI M, et al. Combined approach for short-term wind power forecasting under cold weather with small sample[J]. Proceedings of the CSEE, 2023, 43(2): 543-555(in Chinese).
- [19] 邹智,吴铁洲,张晓星,等.基于贝叶斯优化CNN-BiGRU混合神经网络的短期负荷预测[J].高压技术,2022,48(10):3935-3945.
ZOU Z, WU T Z, ZHANG X X, et al. Short-term load forecasting based on Bayesian optimized CNN-BiGRU hybrid neural network[J]. High Voltage Technology, 2022, 48(10): 3935-3945(in Chinese).
- [20] HUO F C, CHEN Y, REN W J, et al. Prediction of reservoir key parameters in 'sweet spot' on the basis of particle swarm optimization to TCN-LSTM network[J]. Journal of Petroleum Science and Engineering, 2022, 214: 110544.
- [21] 苏伟,肖小龙,史明明,等.周期规律增强的多视角短期电力负荷预测[J].北京航空航天大学学报,2024,50(2):477-486.

- SU W, XIAO X L, SHI M M, et al. Periodic pattern-enhanced multi-view short-term load prediction[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(2): 477-486(in Chinese).
- [22] LI F, WANG C F. Develop a multi-linear-trend fuzzy information granule based short-term time series forecasting model with K-medoids clustering[J]. Information Sciences, 2023, 629: 358-375.
- [23] 欧阳福莲, 王俊, 周杭霞. 基于改进迁移学习和多尺度 CNN-BiLSTM-Attention 的短期电力负荷预测方法[J]. 电力系统保护与控制, 2023, 51(2): 132-140.
- OUYANG F L, WANG J, ZHOU H X. Short-term power load forecasting method based on improved hierarchical transfer learning and multi-scale CNN-BiLSTM-Attention[J]. Power System Protection and Control, 2023, 51(2): 132-140(in Chinese).
- [24] HU Q H, ZHANG R J, ZHOU Y C. Transfer learning for short-term wind speed prediction with deep neural networks[J]. Renewable Energy, 2016, 85: 83-95.
- [25] JIN Y W, ACQUAH M A, SEO M, et al. Short-term electric load prediction using transfer learning with interval estimate adjustment[J]. Energy and Buildings, 2022, 258: 111846.

Short-term load forecasting based on transfer learning and TCN-BiGRU

LIU Jie¹, ZHOU Bowen¹, TIAN Ming², HAN Ke^{3,*}

(1. Heilongjiang Province Key Laboratory of Pattern Recognition and Information Perception, Harbin University of Science and Technology, Harbin 150080, China;

2. China Telecom Heilongjiang Branch, Harbin 150040, China;

3. School of Computer and Information Engineering, Harbin University of Commerce, Harbin 150028, China)

Abstract: Electricity load forecasting is of great significance to the stable operation of power systems. For load forecasting, traditional short-term forecasting techniques frequently employ linear regression models, which have low forecasting accuracy due to the models' inability to incorporate complicated load changes. A temporal convolutional network-bidirectional gated recurrent unit (TCN-BiGRU) model based on transfer learning (TL) is proposed. Highly relevant information is moved into the experimental model using a transfer learning strategy; the data is clustered and analyzed using a K-medoids clustering algorithm; features at various TCN scales are extracted using a parallel convolution strategy; pertinent information is captured using temporal attention (TA); and the TCN training is further extracted using a BiGRU. The non-linear features of the output are further extracted using the dynamic multigroup particle swarm optimization (DMS-PSO) algorithm to optimize and tune the hyperparameters of the network training in order to find the best combination of hyperparameters. The experimental results show that the proposed TL-TCN-BiGRU algorithm reduces mean absolute error (MAE) by 38.6%, root mean square error (RMSE) by 40.7%, mean absolute percentage error (MAPE) by 30.4%, and R^2 by 5.3% relative to the gated recurrent unit (GRU).

Keywords: short-term load forecasting; transfer learning; temporal convolutional network; K-medoids clustering; fusion network