[文章编号] 1003－4684(2026)01-0037-06

基于 PSO-BiGRU-Attention 的短期负荷预测

李新新，付 波

（湖北工业大学电气与电子工程学院，湖北 武汉 430068）

[摘 要] 短期负荷预测是电力调度部门制定日内调度计划和发电计划的重要依据，也是确保电网安全可靠运行的重要保障。旨在搭建基于粒子群算法-双向门控循环单元-注意力机制的短期负荷预测模型，深度挖掘数据间的隐藏关系，突出关键因素的影响，并解决传统网络调参难的问题。首先，对数据进行缺失值和归一化处理；其次，采用 BiGRU 网络挖掘时域数据过去和未来之间的复杂关系；最后，通过 PSO 算法对 BiGRU 网络的参数进行调整，并结合 Attention 机制分配权重，以突出关键特征。为了评估模型的性能，采用某两地的用电负荷数据进行多组对比实验。实验结果表明，所提模型表现更为优异，具有良好的预测效果。

[关键词] 双向门控循环单元；粒子群算法；注意力机制；短期负荷预测

[中图分类号] TM715 [文献标识码] A

负荷预测对维持电网系统稳定运行至关重要。准确的负荷预测不仅可以大幅节省运维成本，提高供电效率，还能为输配电做出合理的未来规划[1]。短期负荷预测的目标是根据历史负荷（考虑天气、温度和其他因素）预测未来几小时或几天的负荷值[2]。进行短期负荷预测往往会得到很大的直接经济效益，避免资源浪费。随着分布式电源和智能电表在智能电网中的应用，对短期负荷预测的精度和效率提出了新的挑战[3]。

电力系统短期负荷预测模型主要分为传统模型、机器学习模型和人工智能模型[4]。传统的预测方法主要有时间序列法、回归分析法和状态空间法[5]。这些预测方法虽然具有计算量小、线性预测简单以及精度高等特点，但是由于现代电力系统规模庞大，传统的预测方法难以处理大量复杂的非线性负荷数据，满足不了现代负荷预测需求。而机器学习模型如灰色预测只适合近似于指数增长的预测；支持向量机（SVR）虽然可以提取非线性特征、避免过度拟合，但是对于异常值的处理以及参数的设置存在很多困难[6]。

人工神经网络（artificial neural networks，ANN）在处理负荷预测中的时间序列非线性以及复杂关系方面显示出了良好的能力[7]。递归神经网络（recurrent neural network，RNN）是一种特殊的人工神经网络，可以用于挖掘数据中的时序和语义信息，适用于处理具有序列特性的数据，近年来在负荷预测领域广受关注[8]。但是传统的递归神经网络在训练过程中存在梯度爆炸和梯度消失的缺点[9]。文献[10]针对传统神经网络的不足，提出了长短时记忆网络（long short-term memory，LSTM），通过门机制改变隐含层的计算方法将短期记忆与长期记忆结合起来，但其网络结构过于复杂与冗余。为此，Cho 等在 2014 年[11]提出了门控循环单元（gated recurrent unit，GRU）对 LSTM 进行改进，其结构更简单，训练速度更快。然而，随着电网规模的不断扩大，现代电力系统受诸多因素影响，单一模型预测在面对复杂电力负荷数据时往往会出现随机性过高的问题，导致训练效率不高，以及场景适应性低等问题[12]。

组合模型可以实现不同技术之间的优势互补[13]，是提高预测精度的最佳方法之一。文献[14]提出了一种 CNN-BIGRU-ATTENTION 混合模型。该模型利用 CNN 提取出较为重要的特征，引入 ATTENTION 机制优化经 BIGRU 分析后的重要时序特征，从而提高预测的精度。文献[15]采用 BiLSTM 解决锂电池数据的时域特征，然后用 PSO 优化网络模型的参数，并引入 Attention 机制解决信息过载问题，具有较好的预测效果。

综上所述，本文综合考虑深度学习算法对复杂数据的处理能力，以及神经网络对时间序列数据信

[收稿日期] 2023－03－28

[基金项目] 国家自然科学基金(51879111)

[第一作者] 李新新(1994－)，男，湖北孝感人，湖北工业大学硕士研究生，研究方向为发电设备性能劣化趋势预测。

[通信作者] 付 波(1973－)，男，湖北武汉人，湖北工业大学教授，研究方向为智能控制、模式识别。

息挖掘方面的优良性能，提出一种基于 PSO-BiGRU-Attention 的短期负荷预测方法。其中，PSO 算法用来寻找 BiGRU 网络中的最优参数，BiGRU 用来深度提取数据中的时空特性以及快速进行信息读取，最后结合 Attention 机制，突出关键特征。

1 理论来源

1.1 粒子群优化算法

PSO 来源于鸟群寻找最佳觅食区域的过程[16]，是一种基于群体搜索的优化算法。通过群体中个体的信息共享，整个群体的运动将在问题求解空间中从无序演变为有序，从而获得最优解。

PSO 的核心是通过迭代寻求最优解。在每次迭代中，粒子通过跟踪个体最优解和群体最优解两个极值来自我更新[17]。在找到这两个最优解后，粒子通过以下公式更新其速度和位置。

$$v_i = v_i + c_i \times \mathrm{rand}() \times (\mathrm{pbest}_i - x_i) + c_2 \times \mathrm{rand}() \times (\mathrm{gbest}_i - x_i) \tag{1}$$

$$x_i = x_i + v_i \tag{2}$$

式中，v_i 为记忆项；$c_i \times \mathrm{rand}() \times (\mathrm{pbest}_i - x_i)$ 为自身认知项；$c_2 \times \mathrm{rand}() \times (\mathrm{gbest}_i - x_i)$ 为群体认知项。PSO 的流程如图 1 所示。

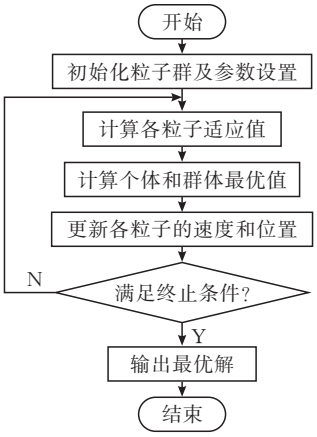

图 1 PSO 算法流程

1.2 门控循环神经网络

LSTM 用来解决 RNN 存在的梯度消失问题，可以轻松学习长期依赖的关系[18]。GRU 作为 LSTM 的一种特殊的变体，只含有两个门控结构，与 LSTM 相比，GRU 参数更少，训练速度更快，用时更短。如图 2 所示，GRU 包含重置门和更新门，前者控制要遗忘多少过去的信息，后者控制就要保留多少之前的信息到当前时刻[19]。计算公式如下：

$$\begin{cases} z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \\ r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \\ \tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \\ h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \end{cases} \tag{3}$$

式中，z_t 和 r_t 分别为更新门和复位门；W_z、W_r、W 为训练参数；h_{t-1} 为前一时刻隐藏层神经元输出；$\tilde{h}_t$ 为当前时刻的候选激活状态。

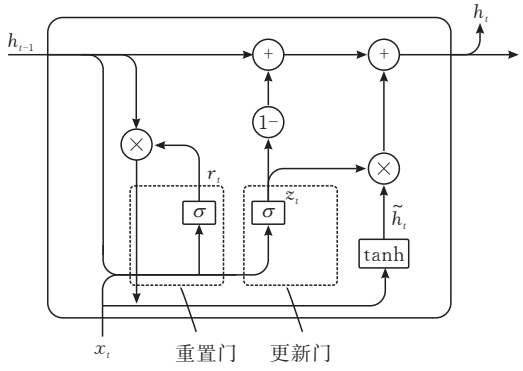

图 2 GRU 神经元结构

经典的 GRU 网络结构中，信息总是从前往后输出的。但是，在电力负荷预测时，要考虑过去电力负荷与未来电力负荷的影响[20]，将其与当前电力负荷联系起来，对预测的准确性更有利。BiGRU 网络由正反两个 GRU 组成，在每一时刻，输出会同时由两个相反的 GRU 共同决定，可以实现对时间序列的双向训练。结构如图 3 所示，公式如下：

$$\overrightarrow{h}_t = \mathrm{GRU}(x_t, \overrightarrow{h}_{t-1}) \tag{4}$$

$$\overleftarrow{h}_t = \mathrm{GRU}(x_t, \overleftarrow{h}_{t-1}) \tag{5}$$

$$h_t = w_t \overrightarrow{h}_t + v_t \overleftarrow{h}_t + b_t \tag{6}$$

式中，$\overrightarrow{h}$ 和 $\overleftarrow{h}_t$ 分别为 t 时刻前向 / 后向传播的隐含信息；w_t 和 v_t 分别表示 t 时刻 BiGRU 所对应的前向/反向隐层状态所对应的权重；b_t 表示 t 时刻的隐藏层状态所对应的偏置量。

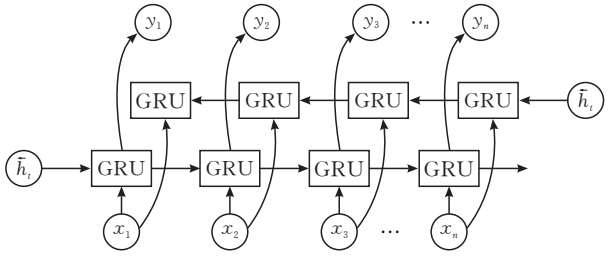

图 3 BIGRU 网络结构

1.3 注意力机制

注意力机制(attention mechanism，A)是机器学习中一种处理数据的方法，源于对人类视觉的研究。在某些特殊时刻，人们会将注意力集中于某些特定领域，减少或忽略对其他领域的关注，以获取更多有用信息，抑制或摒弃无用信息，其核心思想是通过巧妙合理的分配，忽略无用信息并放大有用信息[21]。其原理类比于人类的注意力，通过赋予特征向量不同的权重，聚焦关键特征，从而得到更好的效果，其结构如图 4 所示。

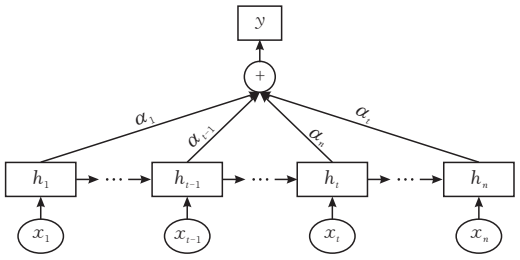

图 4　注意力机制结构

2　基于 PSO-BiGRU-Attention 的负荷预测模型

基于 PSO-BiGRU-Attention 的负荷预测模型层级结构如图 5 所示，主要由输入层、PSO 层、隐藏层、Attention 层和输出层五部分构成。

输入层对原始电力负荷数据预处理，将数据划分为训练集和测试集；PSO 层训练并优化 BiGRU 神经网络结构的超参数，得到精度更高的预测模型；隐藏层采用 BiGRU 神经网络，对数据正反两方面提取各分量间的共性信息，提高预测精度；Attention 层根据 BiGRU 神经网络输入的特征计算不同特征的权重，突出重要特征；输出层最终输出预测结果。

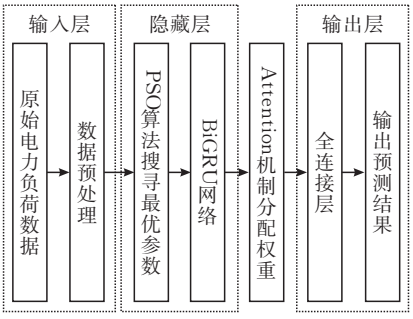

图 5　PSO-BiGRU-Attention 模型结构

2.1　数据预处理

由于原始的电力负荷数据具有随机性、波动性和周期性等特点[22]，直接进行拟合会出现较大误差，因此有必要对历史数据进行预处理，提高输入数据的整体质量，为后续数据训练创造良好条件。

2.1.1　缺失值处理　样本数据的缺失在很大程度上限制了模型的预测精度，本文采用 K 邻近算法(K-nearest neighbor，KNN)进行缺失值填充。其原理是通过识别数据集中空间相似或距离相近的 k 个样本，然后利用 k 邻域的均值对缺失值进行插补[23]。其欧氏距离公式如下：

$$d(x,y)=\sum_{k=1}^{n}\sqrt{(x_k-y_k)^2} \tag{7}$$

式中，x_k、y_k 为该点的横、纵坐标。

2.1.2　归一化处理　为消除不同评价指标间的量纲差异影响，通常会对数据进行归一化处理，将它们统一标准化[24]，其公式如下：

$$x_1=\frac{x-\min(x)}{\max(x)-\min(x)} \tag{8}$$

式中，x_1 为归一化后的数据；x 为原始数据。该函数将原始数据线性化转为[0，1]内。

2.2　评价指标

为客观公正地评价 PSO-BiGRU-Attention 模型的预测效果，方便与其他模型进行对比。选取均方误差(MSE)、均方根误差(RMSE)、平均绝对误差(MAE)、平均绝对百分比误差(MAPE)、决定系数(R^2)、相关系数(R)作为评价指标。

$$\mathrm{MSE}=\frac{1}{n}\sum_{i=1}^{n}(r(t)-c(t))^2 \tag{9}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(r(t)-c(t))^2} \tag{10}$$

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}|r(t)-c(t)| \tag{11}$$

$$\mathrm{MAPE}=\frac{100\%}{n}\sum_{i=1}^{n}\frac{|r(t)-c(t)|}{r(t)} \tag{12}$$

$$R^2=1-\frac{\sum_{i=1}^{n}(r(t)-c(t))^2}{\sum_{i=1}^{n}(r(t)-\bar{r}(t))^2} \tag{13}$$

$$R=\frac{\frac{1}{n}\sum_{i=1}^{n}(r(t)-\bar{r}(t))(c(t)-\bar{c}(t))}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(r(t)-\bar{r}(t))^2}\times\sqrt{\frac{1}{n}\sum_{i=}^{n}(c(t)-\bar{c}(t))^2}} \tag{14}$$

其中，n 为数据个数；t 为时间；$c(t)$ 为模型输出值；$r(t)$ 为真实值；$\bar{c}(t)$ 和 $\bar{r}(t)$ 分别为模型输出值和真实值的均值。

3　算例分析

3.1　A 地区用电负荷数据集对比分析

为验证模型预测准确性，本研究采用 A 地 2016 年某月电力负荷数据进行分析。数据采样间隔为每小时一个数据点，共计 30 天 720 个样本。将全部数据划分为训练集与测试集：前 29 天数据作为训练集用于模型参数训练，剩余 1 天数据作为测试集用于日负荷预测验证。原始数据波形如图 6 所示。

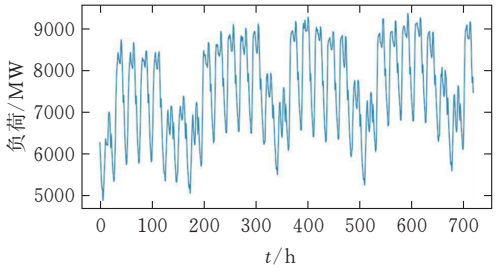

图 6　A 地区 2016 年某月原始数据

通过图 6 波形图，可看出电力负荷数据存在一定的波动性和缺失现象。为降低数据缺失对分析结果的影响，对原始数据进行了插值处理，以更准确地反映其变化趋势，为后续分析提供更可靠的数据基础。

为了验证模型的有效性，本文采用 ANN、LSTM、GRU 和 Attention、PSO 算法优化后的模型进行精度对比，以及 PSO 和 Attention 共同优化的模型进行对比试验。实验结果如表 1 所示。

表 1 不同模型的日负荷预测精度比较

MODEL	MSE	RMSE	MAE	MAPE	R^2	R
ANN	98923.23	314.521	229.365	3.063	0.896	0.958
LSTM	73976.60	271.986	213.875	2.818	0.922	0.962
GRU	72956.66	270.105	208.968	2.768	0.923	0.964
BiGRU-A	54152.89	232.708	183.627	2.391	0.943	0.975
PSO-LSTM	52215.10	228.506	185.037	2.444	0.945	0.973
PSO-GRU	42709.36	206.662	160.163	2.101	0.955	0.980
PSO-BiGRU-A	16350.08	127.867	91.768	1.192	0.983	0.992

由表 1 可以看出，PSO-BiGRU-A 模型日负荷精度更高，其中决定系数 R^2 达到了 0.983，相关系数 R 达到了 0.992，较其他 6 种模型更接近于 1；MSE、RMSE、MAE、MAPE 较其他 6 种模型更低，分别为 16350.08、127.867、91.768、1.192。因此相较之下，本文所提模型具有更高的预测精度。

为了更直观地展示模型的拟合效果，绘制了前 29 天各模型训练后的验证图，如图 7 所示。从图中可以看出，随着训练时间的推移，各模型的预测值在峰值和低谷处大致接近真实值。

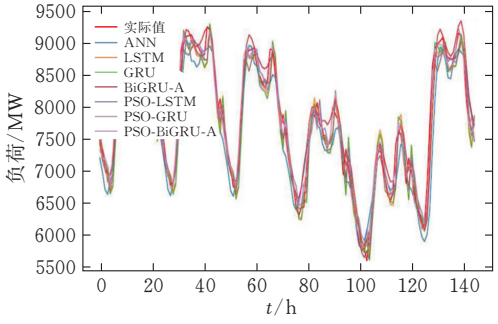

图 7 训练模型对比

图 8 展示了第 30 天 24 h 采样后的模型对比图。从该图中可以看出，PSO-BiGRU-A 在负荷预测方面的表现最接近真实值，预测精度更高。在负荷预测的低谷处，ANN、GRU、LSTM、PSO-GRU 和 PSO-LSTM 均会出现明显的波动，而 BiGRU-Attention 在高峰处会出现较大的偏差。相比于其余六条曲线，PSO-BiGRU-A 预测的曲线更加平滑，且在波峰波谷处更接近真实值。因此，可以得出结论，本文提出的模型具有更好的拟合效果。

为进一步说明 PSO-BiGRU-A 模型的性能，各模型在测试集上的误差箱线图如图 9 所示。可以看出，PSO-BiGRU-A 模型不仅误差中位线更接近于 0，而且箱体更窄，说明其波动性更小，表明该模型较其他 6 种模型误差更小。

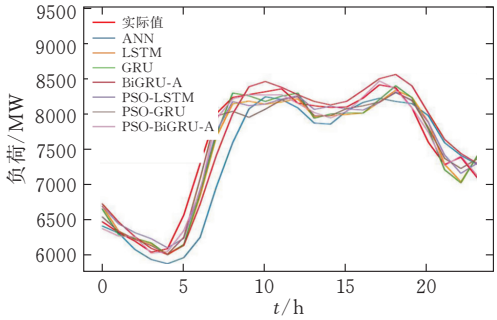

图 8 预测模型对比结果

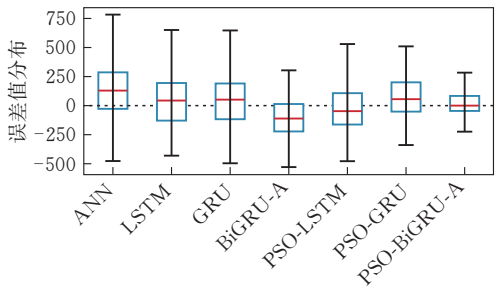

图 9 不同模型的误差箱线图

3.2 B 地区用电负荷数据集对比分析

为了进一步验证模型的鲁棒性，采用 B 地区某月的用电负荷数据集进行验证，处理方法同 A 地数据集。原始数据的波形如图 10 所示。

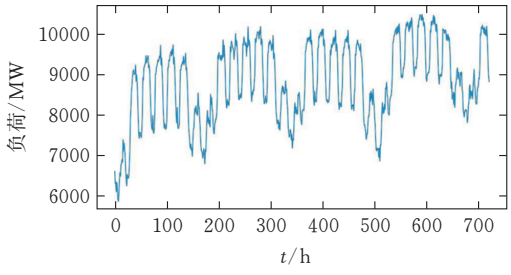

图 10 B 地区某月原始数据

由表 2 可以看出，本文所提出的日负荷预测模型的精度更高。决定系数 R^2 为 0.952，相关系数 R 为 0.977，均优于其他 6 种模型。同时，本模型的 MSE、RMSE、MAE、MAPE 均低于其他 6 种模型，

分别为 26916.95、164.064、125.355、1.369。因此，相较于其他几种模型，本文模型具有更高的精度。

表 2　不同模型的日负荷预测精度比较

MODEL	MSE	RMSE	MAE	MAPE	R^2	R
ANN	75601.19	274.957	218.709	2.421	0.866	0.952
LSTM	58434.00	241.732	190.123	2.033	0.896	0.955
GRU	48966.92	221.285	163.300	1.772	0.913	0.956
BiGRU-A	32988.23	181.627	128.259	1.395	0.942	0.972
PSO-LSTM	45533.26	213.385	170.102	1.877	0.919	0.968
PSO-GRU	38945.48	197.346	156.690	1.721	0.931	0.971
PSO-BiGRU-A	26916.95	164.064	125.355	1.369	0.952	0.977

图 11 展示了前 29 天各模型在训练后的验证结果。在训练期间，除了 ANN 和 LSTM 模型在最低值和最高值时与真实值之间的误差较大之外，其余模型逐渐趋向真实值。图 12 则是第 30 天进行 24 h 采样后的模型对比图。由图可知 ANN、LSTM、PSO-GRU 和实际值之间存在较大误差；GRU 模型的预测结果比真实值提前；PSO-LSTM 模型的波形偏移较大，误差也比较显著；而在低谷时，BiGRU-A 模型的误差较大。综上所述，本文提出的模型不仅在线性拟合时更平滑，而且与真实值之间的误差更小，这表明该模型具有更好的预测效果，相对于其他 6 种模型，更接近真实值。

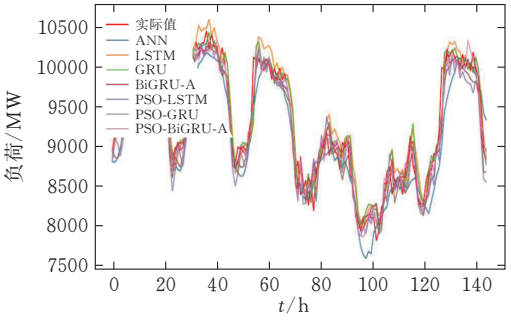

图 11　训练模型对比

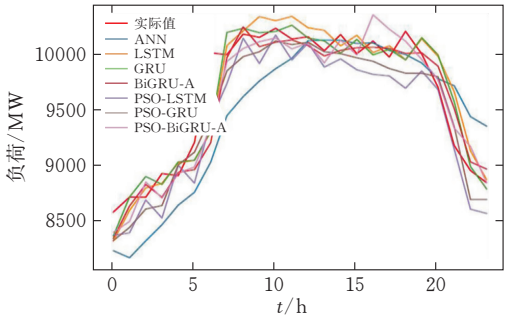

图 12　预测模型对比结果

为了更直观地展示各模型的预测效果，采用误差分布箱线图来反映各模型预测性能上的差异，如图 13 所示。从图中可以看出，BiGRU-A 的中位线接近于 0，但上下边缘极差较大，而 PSO-BiGRU-A 的中位线也接近于 0，且上下边缘极差较小。结合表 2 中给出的各项评价指标可以发现，基于 PSO-BiGRU-A 模型具有较高的预测精度。

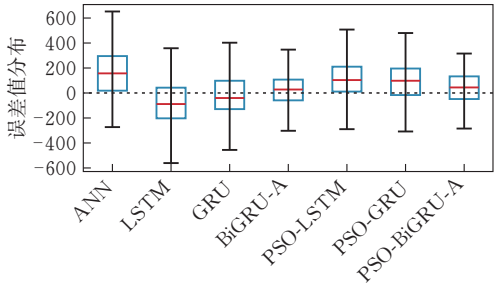

图 13　不同模型的误差箱线图

4　结论

本文提出了一种基于 PSO-BiGRU-Attention 的短期负荷预测方法，并通过实际案例验证了该模型具有优越的预测效果。主要结论如下：

1）PSO 算法能够有效优化 BiGRU 模型参数，提高负荷预测精度；同时，BiGRU 网络可融合时序数据中过去与未来状态之间的复杂关联，实现特征深层提取，进一步提升预测准确度。

2）在模型中引入 Attention 机制，通过对不同特征赋予差异化权重以聚焦关键信息，从而增强预测效果。相较于传统机器学习与深度学习方法，该模型能更好地捕捉数据的时空特征和相关性，提高预测精度和效率，能够有效地处理电力系统负荷预测中的非线性、动态和复杂特性。

[参　考　文　献]

[1] MADHUKUMAR M, SEBASTIAN A, LIANG X, et al. Regression model-based short-term load forecasting for university campus load[J]. IEEE Access, 2022, 10: 8891-8905.

[2] ZhANG M, YU Z, XU Z. Short-term load forecasting using recurrent neural networks with inputattention mechanism and hidden connection mechanism[J]. IEEE Access, 2020, 8: 186514-186529.

[3] AZEEM A, ISMAIL I, JAMEEL S M, et al. Electrical load forecasting models for different generation modalities: a review[J]. IEEE Access, 2021, 9: 142239-142263.

[4] DENG Z, WANG B, XU Y, et al. Multi-scale convo-

lutional neural network with time-cognition for multi-step short-term load forecasting[J]. IEEE Access, 2019, 7: 88058-88071.
[5] YAO X, FU X, ZONG C. Short-term load forecasting method based on feature preference strategy and light-gbm-xgboost[J]. IEEE Access, 2022, 10: 75257-75268.
[6] 龚田李慧.基于 MDS 和 PSO-GRU 神经网络的短期负荷预测研究[D]. 武汉:湖北工业大学,2021.
[7] 余阳.基于深度学习的电力系统短期负荷预测研究[D]. 南昌:南昌大学, 2021.
[8] WANG S, SUN Y, WANG J, et al. Very Short-term Prediction for Wind Power Based on BiLSTM-Attention[C]// 2021 IEEE Sustainable Power and Energy Conference (iSPEC). IEEE, 2021: 292-296.
[9] 葛琼宇,王致杰.基于 ISSA-ELM 的短期电力负荷预测研究[J].国外电子测量技术,2022,41(04):119-125.
[10] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural computation, 1997, 9(08): 1735-1780.
[11] 周国兵.最少门结构的循环神经网络及其应用[D]. 南京:南京大学,2016.
[12] 李孟洲.基于组合模型的短期电力负荷预测研究[D].郑州:郑州大学,2021.
[13] 乐昊哲.基于神经网络混合模型的短期电力负荷预测算法研究[D].南昌:南昌大学,2021.
[14] 方娜,余俊杰,李俊晓,等.基于 CNN-BIGRU-ATTENTION 的短期电力负荷预测[J].计算机仿真,2022,39(02):40-44.
[15] 张永,辛宇琪,钱启政,等.PSO 优化的 BiLSTM-Attention 网络的锂电池健康状态评估[J].控制工程,2022,29(02):287-293.
[16] 赵晋斌,张建平,毛玲,等.基于 PSO-Soft attention 双向 LSTM 算法的光伏发电量预测研究[J].智慧电力,2022,50(03):1-7.
[17] 吴松梅,蒋建东,燕跃豪,等.基于 VMD-PSO-多核极限学习机的短期负荷预测[J].电力系统及其自动化学报,2022,34(05):18-25.
[18] QIN J, ZHANG Y, FAN S, et al. Multi-task short-term reactive and active load forecasting method based on attention-LSTM model[J]. International Journal of Electrical Power & Energy Systems, 2022, 135: 107517.
[19] 刘国海, 孙文卿, 吴振飞, 等. 基于 Attention-GRU 的短期光伏发电功率预测[J]. 太阳能学报, 2022, 43(02): 226.
[20] 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].高电压技术,2022,48(10):3935-3945.
[21] 赵兵,王增平,纪维佳,等.基于注意力机制的 CNN-GRU 短期电力负荷预测方法[J].电网技术,2019,43(12):4370-4376.
[22] ZHANG M, YU Z, XU Z. Short-term load forecasting using recurrent neural networks with input attention mechanism and hidden connection mechanism[J]. IEEE Access, 2020, 8: 186514-186529.
[23] 朱金冲.基于神经网络的属性关联建模与缺失值填补[D]. 大连:大连理工大学,2021.
[24] 胡文波.基于 CSWOA 优化 BiGRU 融合时间模式注意力模型的短期电力负荷预测[D].广州:广东工业大学,2022.

Short-term Load Forecasting Based on PSO-BiGRU-Attention

LI Xinxin, FU Bo

(*School of Electrical and Electronic Engineering, Hubei Univ. of Tech., Wuhan* 430068, *China*)

Abstract: Short-term load forecasting is an important basis for the daily scheduling and power generation planning of the power dispatch department, and also an important guarantee for ensuring the safe and reliable operation of the power grid. This paper aims to build a short-term load forecasting model based on the Particle Swarm Optimization (PSO) optimized Bidirectional Gate Recurrent Unit (BiGRU) Attention mechanism, which deeply explores the hidden relationships between data, highlights the impact of key factors, and solves the problem of difficult tuning of traditional networks. First, the missing values and normalization processing of the data are carried out. Second, the BiGRU network is used to mine the complex relationships between past and future temporal data. Finally, the parameters of the BiGRU network are adjusted by the PSO algorithm, and combined with the Attention mechanism to allocate weights to highlight key features. In order to evaluate the performance of the model, multiple comparative experiments were conducted using the electricity load data of two locations. The experimental results show that the proposed model in this paper performs better and has good prediction effect.

Keywords: bidirectional gated recurrent unit; particle swarm optimization algorithm; attention mechanism; short-term load forecasting

[责任编校: 赵 瞳]