

# 基于XGBoost特征提取的热电联产发电功率预测

常庆<sup>1</sup>, 罗龙峰<sup>2</sup>

<sup>1</sup>安徽理工大学机械工程学院, 安徽 淮南

<sup>2</sup>浙江工业大学工商管理学院, 浙江 杭州

收稿日期: 2022年9月27日; 录用日期: 2022年10月14日; 发布日期: 2022年10月25日

## 摘要

精准的短期电力负载预测对电力系统的调度与控制、安全与防御等方面具有重要意义。随着电厂设备的日益复杂化, 依靠专家经验或人工提取特征也变得愈发困难。结合极限梯度提升(Extreme Gradient Boosting, XGBoost)在提取特征和处理高维、非线性数据等方面的优势, 提出一种基于XGBoost特征提取的融入注意力机制的深度卷积长短期记忆网络(CNN-LSTM-A)模型预测方法。首先, 针对短期负荷预测, 其电力负荷以及影响因素历史数据的随机性、波动性和不确定性, 使用XGBoost对历史序列数据进行特征选择。将处理后的数据通过卷积层进行更深入的特征提取。然后, 通过长短期记忆层进行拟合, 并采取注意力机制对权值进行优化, 减少历史信息的丢失并加强重要信息的影响, 最终完成对功率的精准预测。利用浙江某电厂的实测数据进行实验, 并与传统常用的经典模型进行对比, 实验结果表明, 所提方法预测精度更高, 验证了该模型的精准性与可行性。

## 关键词

热电联产, 极限梯度提升, 注意力机制, 卷积神经网络, 长短期记忆网络, 功率预测

# Power Forecast of Cogeneration Based on Feature Extraction Using XGBoost

Qing Chang<sup>1</sup>, Longfeng Luo<sup>2</sup>

<sup>1</sup>College of Mechanical Engineering, Anhui University of Science and Technology, Huainan Anhui

<sup>2</sup>College of Business Administration College, Zhejiang University of Technology, Hangzhou Zhejiang

Received: Sep. 27<sup>th</sup>, 2022; accepted: Oct. 14<sup>th</sup>, 2022; published: Oct. 25<sup>th</sup>, 2022

## Abstract

Accurate short-term power load prediction is of great significance to power system scheduling and control, security and defense. With the increasing complexity of power plant equipment, it is increasingly difficult to rely on expert experience or manual extraction of features. Combining with Extreme Gradient Boosting (XGBoost) advantages in feature extraction and processing of high-dimensional and nonlinear data, a deep convolutional long and short-term memory network with attention mechanism (CNN-LSTM-AM) model prediction method based on XGBoost feature extraction was proposed. First of all, for short-term load forecasting, the randomness, volatility and uncertainty of historical data of power load and influencing factors, XGBoost is used to select the characteristics of historical sequence data. Further feature extraction is carried out on the processed data through the convolutional layer. Then, the long and short-term memory layer is fitted, and the attention mechanism is adopted to optimize the weights, reduce the loss of historical information and strengthen the influence of important information, and finally complete the accurate prediction of power. The measured data from a power plant in Zhejiang province are used for experiments and compared with the traditional classical model. Experimental results show that the proposed method has higher prediction accuracy, which verifies the accuracy and feasibility of the model.

## Keywords

Cogeneration, Extreme Gradient Boosting, Attention Mechanism, Convolution Neural Network, Long and Short-Term Memory Network, Power Forecast

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

提高短期用电负荷预测的准确性与预测精度, 对于电力系统的安全、稳定和经济运转具有非常重要的意义[1]。根据国际能源署发布的《2019 年世界能源展望》[2], 未来几十年, 住宅部门的对电力供应量需求将持续快速上升, 仅次于工业部门。在此背景下, 通过对电力用户对于汽轮机功率需求的准确预测, 决策部门可以制定合理的用电计划, 确保居民用电供应的安全。估计不当可能导致严重的社会和经济损失。

热电联产是电厂在生产电能的同时, 利用蒸汽轮机中做完功的蒸汽提供供热产品的工艺过程。相较于传统的热电分产, 热电联产具有节约燃料、污染排放少、能源综合利用效率高、提高能源供应安全以及增加电力供应等综合效益。因产出能源的能量形式不同, 在我国火力发电厂可分为凝汽式发电厂和热电厂。其中, 前者仅产出电能然后供给用户, 并且采用单元发电机组。而热电厂则多了另一个选择: 母管制热电机组。如今, 国内绝大多数热电厂采用单元制形式, 但仍有部分热电厂的运行方式为母管制。在母管制运行方式下, 锅炉和汽轮机发电机组之间由于主蒸汽母管的存在, 使其并无联系。各机组之间可以通过相互协调满足实际情况的负荷调度[3]。浙江某热电厂采用的是 3 台锅炉加 2 台汽轮机的母管制运行方式, 对不同锅炉和汽轮机进行负荷调度调节, 需要提前准确预测出汽轮机组功率的值以及变化趋势, 进而充分有效地利用汽轮机组的全部负荷, 避免资源浪费。

目前, 统计学方法、人工智能方法以及它们之间相互组合的方法已经成为热电厂负载预测的三种主

流技术[4]。每种方法都具备了各自适用的时间尺度大小和数据范围。统计学方法主要包括时间序列法、相关分析法和卡尔曼滤波等, 人工智能方法主要包括支持向量机法(Support Vector Machine, SVM) [5]、专家系统法和人工神经网络法等, 组合方法就是基于上述两种方法进行组合。

常峰铭等[6]为了更好地提取输入信息中数据的特性, 通过使用卷积神经网络(Convolutional Neural Network, CNN)的模型对其进行了负荷预测。得出了相较于其它的神经网络, CNN 对于非线性的序列数据更加敏感, 可以更好地挖掘和分析学习序列变化规律。但是当负荷数据值出现较大震荡变化且平稳性较低时, 单一的 CNN 模型就很难较好地掌握序列变化规律。栗然等[7]提出了基于长短期神经网络(Long Short-Term Memory, LSTM)的短期电能负荷预测分析方法, 并且充分说明了 LSTM 对于数据的时序特征提取的优异能力及方法的实际应用可行性。刘亚瑛等[8]提出了一种基于聚类经验模态分解的 CNN 和 LSTM 相互结合的超短期电力负荷预测模型, 解决了 LSTM 因输入过长导致预测精度准确率降低的问题, 具有更高的预测精度。

LSTM、RNN 等方法存在由于输入序列过长, 容易导致丢失序列信息、难以建模数据间结构信息的问题, 从而在一定程度上会直接影响负荷预测的计算精度。针对 LSTM 等方法目前存在的问题, 本文提出基于注意力机制(Attention Mechanism) [9]的 CNN-LSTM 模型, 对 LSTM 隐含状态赋予不同的概率权重, 减少重要历史信息的遗漏并加强对重要信息的影响, 最终达到提高系统负载预测值的精确性。最后通过实验证明, 将本文所提方法和经典方法进行对比, 得出本文方法在预测时可以具备较高的准确性。

## 2. 基于 XGBoost 特征提取的发电功率预测

### 2.1. XGBoost 原理

提取电力负荷和其影响因素历史数据的深层特征对于短期电力负荷预测具有非常重大的意义。XGBoost 是基于决策树的集成机器学习算法, 它以梯度提升(Gradient Boost)为框架, 在多数情况下挑选模型进行训练时, XGBoost 模型是一个非常热门的模型。但其实在前面特征选择部分, 基于 XGBoost 进行特征筛选也大有可为。XGBoost 根据结构分数( $L_{split}$ )的增益情况计算出某个特征的某个分割点。某个特征的重要性, 就是这个特征在所有树中出现的次数之和。

XGBoost 是由迭代梯度决策树(Gradient Boosting Decision Tree, GBDT)发展而来, 同样是利用加法模型与前向分步算法实现学习的优化过程, 但与 GBDT 是有区别的。主要区别包括以下几点: 1) 目标函数: XGBoost 的损失函数添加了正则化惩罚项, 使用正则化用以控制模型的复杂度; 2) 优化方法: GBDT 在优化时只使用了一阶导数信息, XGBoost 在优化时使用了一、二阶导数信息; 3) 稀疏数据处理: XGBoost 对稀疏数据进行了处理, 通过学习模型自动选择最优的缺失值默认切分方向; 4) 防止过拟合: XGBoost 除了增加了正则项来防止过拟合, 还支持列采样的方式来防止过拟合; 5) 效率: XGBoost 可以在最短时间用更少的计算资源得到更好的结果。

XGBoost 是一种提升树模型, 其算法思想是将许多弱分类器集成在一起形成一个强分类器。通俗的理解是不断地进行特征分裂来生长一棵树, 每次添加一个树, 其实是学习一个新函数, 去拟合上一次的残差。XGBoost 的主要目标就是通过找到这样一个新的函数去优化目标函数, 使得最终结果足够小。下面是对目标函数的推导化简过程。

对于给定的  $n$  个样本,  $m$  个特征的数据集  $D = \{(x_i, y_i)\} (|D| = n, x_i \in \mathbb{R}^m, y_i \in \mathbb{R})$ , 树集模型使用  $K$  个函数相加来预测输出。

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (1)$$

式中,  $q$  表示每个树的结构,  $F = \{f(x) = w_{q(x)}\} (q: \mathbb{R}^m \rightarrow T, w \in \mathbb{R}^T)$  是回归树空间。这些树将每一个样本映射到相应的叶索引。  $T$  是树的叶子节点数。每一个  $f_k$  对应一个独立的树结构  $q$  和叶权值  $w$ 。与决策树不同, 每棵回归树在每个叶子上都包含一个连续的权值, 我们使用  $w_i$  来表示第  $i$  个叶子上的权值。对于一个给定的例子, 我们将使用树中的决策规则(由  $q$  给出)将其分类到叶子节点中, 并通过将相应叶子节点(由  $w$  给出)的分数相加来计算最终的预测。为了学习模型中使用的函数集, 将最小化以下目标函数。

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \tag{2}$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|W\|^2$$

式(2)中,  $l$  是一个可微的凸损失函数, 用来衡量预测值与真实值之间的差值。第二项  $\Omega$  为正则化惩罚项。增加正则化项有助于平滑最终学习的权值, 避免过度拟合。直观地说, 正则化目标倾向于选择使用简单预测函数的模型。当正则化参数设置为零时, 目标回落到传统的梯度树增强。

式(2)中的树集成模型将函数作为参数, 无法在欧几里德空间中使用传统的优化方法进行优化。相反, 以加法的方式训练模型。形式上, 设  $\hat{y}_i$  是在第  $t$  次迭代时对第  $i$  个样本的预测, 我们需要添加  $f_t$  来最小化以下目标函数。

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_{t-1} + f_t(x_i)) + \Omega(f_t) \tag{3}$$

式(3)意味着根据式(2)加入了对模型改进效果最大的  $f_t$ , 一般情况下可以用二阶导数近似来快速优化目标函数[10]。

$$L^{(t)} \approx \sum_{i=1}^n \left[ l(y_i, \hat{y}_{t-1}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{4}$$

式(4)中,  $g_i = \partial_{\hat{y}_{t-1}} l(y_i, \hat{y}_{t-1})$ 、 $h_i = \partial_{\hat{y}_{t-1}}^2 l(y_i, \hat{y}_{t-1})$  是目标函数一阶、二阶导数。在第  $t$  步, 去掉常数项, 得到以下简化目标函数。

$$\tilde{L}^{(t)} = \sum_{i=1}^n \left[ g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{5}$$

定义  $I_j = \{i | q(x_i) = j\}$  为叶子节点  $j$  的样本数据集, 可以将式(5)如下展开。

$$\begin{aligned} \tilde{L}^{(t)} &= \sum_{i=1}^n \left[ g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \\ &= \sum_{j=1}^T \left[ \left( \sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left( \sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \end{aligned} \tag{6}$$

对于一个固定的树结构  $q(x)$ , 可以通过  $\tilde{L}^{(t)}$  对  $w_j$  求导后令其等于零, 即得到叶子节点  $j$  的最佳权重  $w_j^*$ 。

$$w_j^* = \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{7}$$

将  $w_j^*$  带回  $\tilde{L}^{(t)}$ , 由此可以计算出目标函数的如下最优值。

$$\tilde{L}^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^T \frac{\left( \sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \tag{8}$$

式(8)可以作为一个评分函数来衡量树结构  $q$  的质量, 这个评分类似于评价决策树的混乱程度评分。

在决策树的生长过程中, 一个非常关键的问题是如何确定叶子的节点的最优切分点。XGBoost 支持两种分裂节点的方法: 1) 贪婪算法; 2) 近似算法。

在计算机计算资源足够的情况下, 贪婪算法可以得到最优解。因此, 本文采用贪婪算法找寻分裂节点。假设  $I_L$  和  $I_R$  是拆分后左右节点的实例集。设  $I = I_L \cup I_R$ , 对于目标函数来说, 拆分后的收益为:

$$L_{split} = \frac{1}{2} \left[ \frac{\left( \sum_{i \in I_L} g_i \right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left( \sum_{i \in I_R} g_i \right)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{\left( \sum_{i \in I} g_i \right)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \quad (9)$$

观察分裂后的收益, 我们会发现节点划分不一定会使得结果变好, 因为我们有一个引入新叶子的惩罚项, 也就是说, 引入的分割带来的增益如果小于一个阈值的时候, 我们可以剪掉这个分割。该公式通常在实践中用于评估候选的分裂节点。

除了以上提到的正则化以外, XGBoost 还使用了两种额外的技术来进一步防止过拟合。第一种是由 Friedman [11] 提出的 shrinkage 策略(权重衰减, 类似于 LSTM)。在每一步树的提升之后, 新增了一个  $\eta$  因子的权重。把整个训练过程看作一个时间序列, 离当前树时间点越靠前的权重对当前权重的累加影响越小, 这个衰减就是由  $\eta$  控制的。第二种是特征降采样[12] [13]的方法来避免过拟合, 只是这里的降采样使用的是列降采样(与随机森林做法一样), 它的另外一个好处是可以方便加速并行化。XGBoost 做出的创新还有提出了加权分位数略图, 证明了在求解对样本点合理切割的问题上, 可以用样本的特征值和其二阶导数的关系更好地划分数据分布范围。

## 2.2. CNN 原理结构

目前的 CNN [14]一般具有局部连接、权值共享和汇聚这三个结构上的特性。利用这三个特性大幅度降低模型参数的数量, 加快训练速度, 提高泛化性能。它在处理图像和视频分析[15]的各种任务中有着显著的优势(如人脸识别[16]、图像处理等[17])。卷积神经网络分为一维卷积[18]、二维卷积和三维卷积, 每一类都有适用的场景。其中, 一维卷积神经网络主要用于时间序列数据。以下介绍其中最常见卷积神经网络框架结构。多个神经元被包含于一个特征表面内, 多个特征表面共同构成一个卷积层。卷积核将每个神经元和上一层空间中的局部特征表面区域互相地连接了起来, 卷积核神经网络的卷积层就是通过这样的卷积操作提取从上一层输入到下一层空间中的局部特征, 以卷积核或滤波核(Kernel)步长为单位纵向依次与数据进行卷积操作, 得到简化数据。然后将简化后的数据进行最大池化操作, 输出特征域的最大值。Dropout 操作将部分神经元按照一定的概率从神经网络暂时地舍弃, 降低神经网络复杂度, 提高训练效率, 防止了过拟合。当神经网络投入使用时, 被舍弃的神经元再次连接。随着计算机计算能力的提升, 池化操作和 Dropout 操作可以舍弃。

## 2.3. LSTM 原理结构

LSTM [19]是一种循环神经网络的重要变体。通过引入门控机制来实现对信息进行累加和处理, 能够有效地解决简单循环神经网络的梯度爆炸或消失问题。为了改善循环神经网络的长程依赖问题, LSTM 作为一种特殊的 RNN 结构, 在以往的各种研究中已经证明了它在建模长期依赖关系方面的稳定性和强大能力。LSTM 的主要创新之处在于引入了表征长期记忆的细胞态  $c_t$ 。因为细胞态的容量是有限的, 所以要选择性的遗忘信息。此时刻细胞态  $c_t$  的记忆信息由两项组成: 一是上一时刻的细胞态  $c_{t-1}$  乘以遗忘门, 这个乘积项表示留存在此时刻细胞态  $c_t$  中对于过去信息的记忆, 二是候选态  $\tilde{c}_t$  与输入门的乘积项, 表征此时刻的记忆信息。其中, 候选态是由当前输入  $x_t$  和上一时刻的记忆体  $h_{t-1}$  (表征短期记忆)归纳得出。此

时刻的细胞态  $c_t$  再经过输出门得到此时此刻的记忆体  $h_t$ 。当有多层循环网络时, 第二层的网络的输入  $x_t$  即为上一层循环网络的输出  $h_t$ 。本文采用的 LSTM 公式, 关键方程如下所示:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (10)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (11)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (12)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (13)$$

$$h_t = o_t \odot \tanh(c_t) \quad (14)$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (15)$$

式(1)~(6)中,  $i_t$ 、 $f_t$  和  $o_t$  分别是输入门、遗忘门和输出门。 $\odot$  为向量元素乘积,  $\tilde{c}_t$  是通过非线性函数得到的等待存入长期记忆的候选状态。 $\sigma$  是 sigmoid 激活函数, 使得门限的范围在 0 到 1 之间。 $W_*, U_*, b_*$  为待学习的网络参数, 其中,  $* \in \{i, f, o, c\}$ 。

LSTM 层对卷积层的输出映射(长×宽的矩阵)提取时间特征后, 将多维输入变量与目标预测值建立映射关系。

## 2.4. Attention 机制原理

Attention 机制主要是一种模仿人脑对注意力的信息资源利用和分配方式的机制, 人脑在某些特殊的时刻可能会将注意力主要集中到自己所需要关注的一些重点地方, 减少甚至完全忽略对其他地方区域的关注, 以便获取人们更多的所需要和关注的重点区域细节或者信息, 抑制其他冗余或无用的信息, 其中, 最为主要和核心的思想就是巧妙地调动和改变了人们对这些信息的感知和注意力, 忽略了这些冗余或无用的信息并放大所需信息, 从而大大提高了模型的设计准确性。注意力机制能够有效改进 LSTM 因为输入的时间序列过长而导致遗漏序列信息的问题, 同时, 采用随机分配概率的方法取代了原始随机分配权值的方法。

本文所提混合模型中全连接层将 LSTM 提取的特征输入到 Attention 层中处理, Attention 层是一个融合了 Attention Mechanism 的全连接层。其根据所有输入的信息被划分为各自所需要的注意力权值, 不再必须把所有的信息都直接传输到神经网络, 大大节约了计算时间和资源。以下介绍了注意力机制计算的两个步骤。首先, 将所有输入的信息统一地计算出每个输入信息的注意力分布。其次, 通过注意力分布来计算出每个输入信息的加权平均[20]。计算公式如下所示:

$$\begin{aligned} \alpha_n &= p(z = n | X, q) \\ &= \text{softmax}(s(x_n, q)) \\ &= \exp(s(x_n, q)) / \left( \sum_{j=1}^N \exp(s(x_j, q)) \right) \end{aligned} \quad (16)$$

式中,  $\alpha_n$  为注意力分布, 即在给定的  $q$  和  $X$  下, 选择出哪一个向量的机率;  $z \in [1, N]$  为注意力变量, 表示被选择信息的索引位置;  $X = [x_1, \dots, x_n] \in R^{D \times N}$  表示  $N$  组输入信息, 每组信息的维度为  $D$ ; 我们将要引进一个查询向量  $q$ , 意指和任务相关的表示, 用来从  $N$  个输入向量中挑选出和某一个特殊的任务的相关信息。其表征输入向量与特定任务相关的信息; softmax 函数将计算值转化为 [0, 1] 内的概率分布;  $s(x_n, q)$  为注意力打分函数, 本文采用的是点积模型。可由下式计算:

$$s(x_n, q) = x_n^T q \quad (17)$$

将 LSTM 层的输出和经 Attention 操作后的权值一并输入 Flatten (压平)层进行维数的削减, 转化为一维向量, 经过输出层输出预测序列结果。

该模型在传统的 LSTM 上引入了卷积层, 提高了特征挖掘的效率, 降低了网络的参数复杂度; 引入注意力机制, 充分考虑数据纵向维度之间的相关程度, 优化权值分配, 能够获得更好的预测效果。

## 2.5. CNN-LSTM-A 模型

本文提出的模型结构如图 1 所示, 主要分为输入层、CNN 层、LSTM 层、Attention 层和输出层。历史功率数据作为输入通过 CNN 提取特征, LSTM 层和 Attention 层从所提取特征学习功率内部变化规律从而实现准确的预测, 最后输出层输出结果。

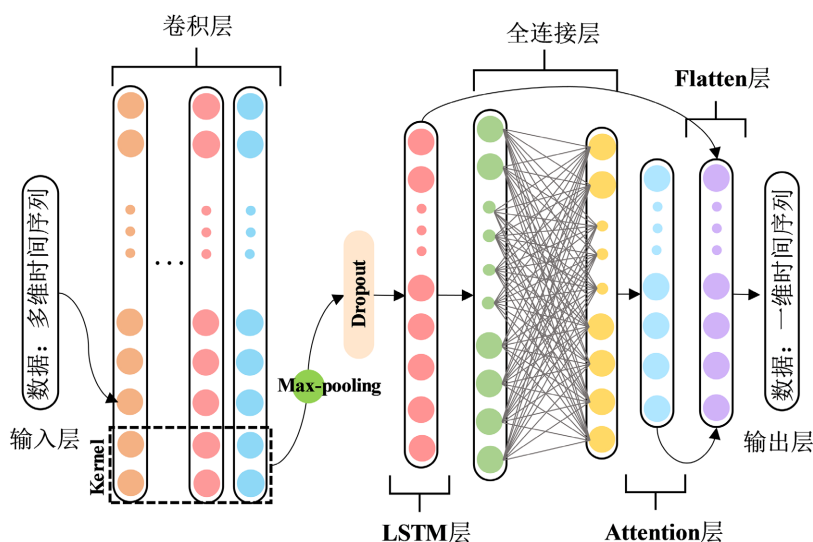


Figure 1. Hybrid deep learning model structure diagram  
图 1. 混合深度学习模型结构图

在混合深度学习模型原理的基础上, 本文功率预测的具体流程为:

### 1) 数据获取及预处理

以热电厂历史运行的数据作为源域, 以实时检测系统采集的数据作为目标域。将两部分的数据进行数据预处理后, 构建可喂进预测模型的数据集。

### 2) 源域模型离线训练

在训练过程中, 将源域数据集喂进已经搭建好的模型中, 模型通过反向传播不断优化更新模型参数来得到最优的神经网络模型。当模型收敛时, 保存模型的训练参数。

### 3) 目标域功率在线预测

在线预测过程中, 将第二步保存的最优模型参数同步更新到目标域模型中。将实时采集的数据集输入到模型中, 通过模型预测下一时段的功率趋势, 最后输出预测值。

## 3. 实验研究

### 3.1. 实验数据集

为了验证所提模型的有效性, 本文采用某泰爱斯环保能源有限公司 2 台 CB30-13.24/3.5/0.981 抽汽背压型汽轮机组日常运行数据, 采样周期为 1 分钟, 采集数据参数共有 250 个, 采集 2020 年 5 月 29 日至

2021年5月29日的数据。汽轮机组为超高压高温单缸抽汽背压机,设计最大进汽量为304.5 t/h,最大抽汽量为80 t/h。

热电机组正常运行时,常常按照“以热定电”原则确定热电负荷。即根据用户需求确定机组供热负荷,再根据供热负荷确定机组的供电负荷。本母管制机组也是根据“以热定电”原则来进行负荷调度。在冬天采暖季工况下,负荷变化以一天为周期呈周期性变化。取某一天的负荷变化如图2所示。1号机组和2号机组的排汽流量除去厂内自用的蒸汽流量即为对外供热总流量。晚上供热需求低,因此1号机组停运,只运行2号机组(供热蒸汽不足部分由主蒸汽直接减温减压供热);白天负荷需求大,1号机组也投入使用,在峰值期间满负荷运行。2号机组全天都以稳定的负荷运行。按照“以热定电”原则,机组的供电负荷变化也呈相似的规律,如图3所示。

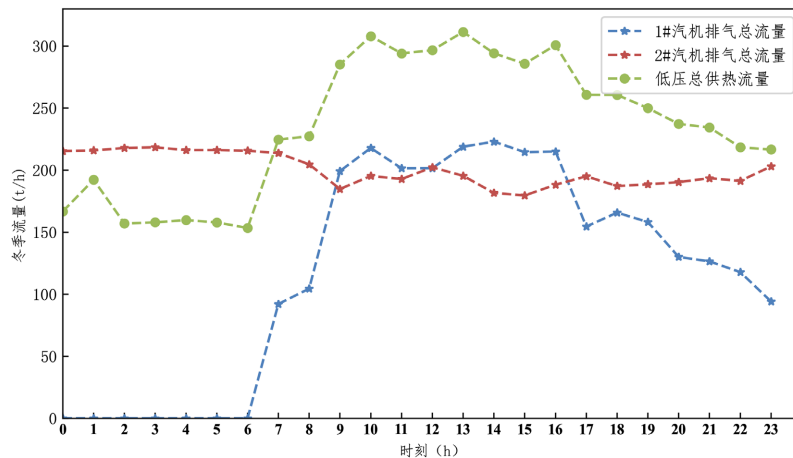


Figure 2. Curve of low pressure heating flow rate with time in heating season (in a day)  
图2. 采暖季机组低压供热流量随时间变化曲线(一天之内)

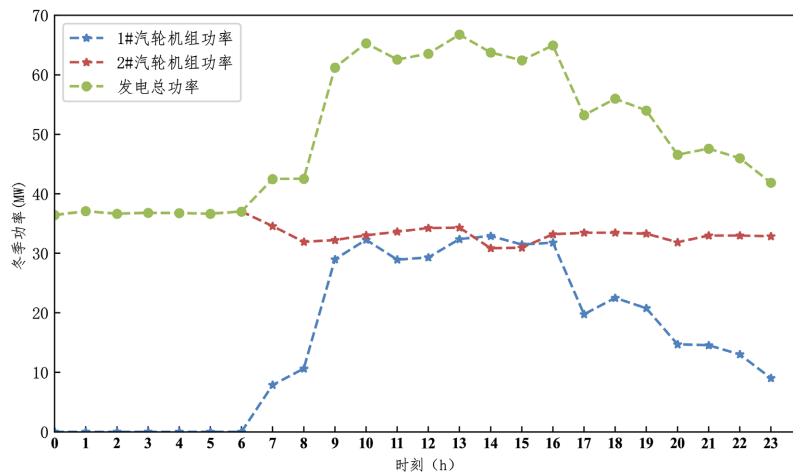


Figure 3. Curve of generating power over time in heating season (in a day)  
图3. 采暖季机组发电功率随时间变化曲线(一天之内)

而在夏季非采暖季工况下,机组的供热负荷变化规律大致相同,在中午12时负荷存在一个波动,但是总供热流量有明显的下降。因此机组的供电负荷也发生了变化。其变化情况如图4、图5所示。

采暖季与非采暖季的供热负荷需求波动较大,从而导致发电负荷有较为显著的差距。因此,准确的热电负荷预测对确定热电联产机组发电计划、控制购电成本具有重要的意义。



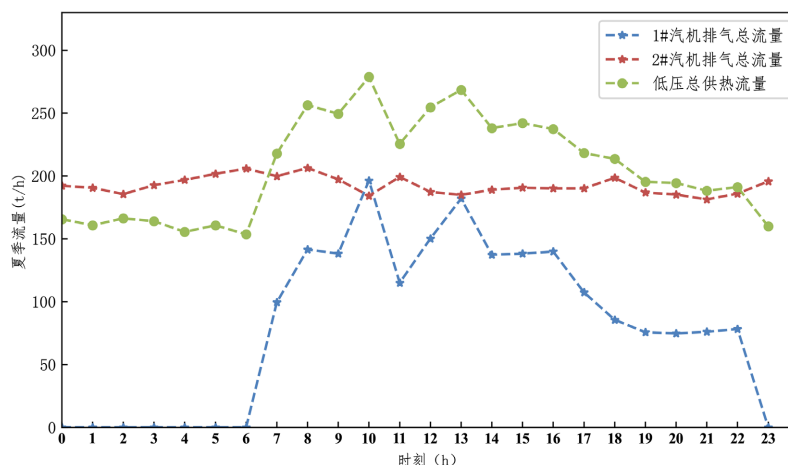


Figure 4. Curve of low pressure heating flow rate with time in non-heating season (in a day)  
 图 4. 非采暖季机组低压供热流量随时间变化曲线(一天之内)

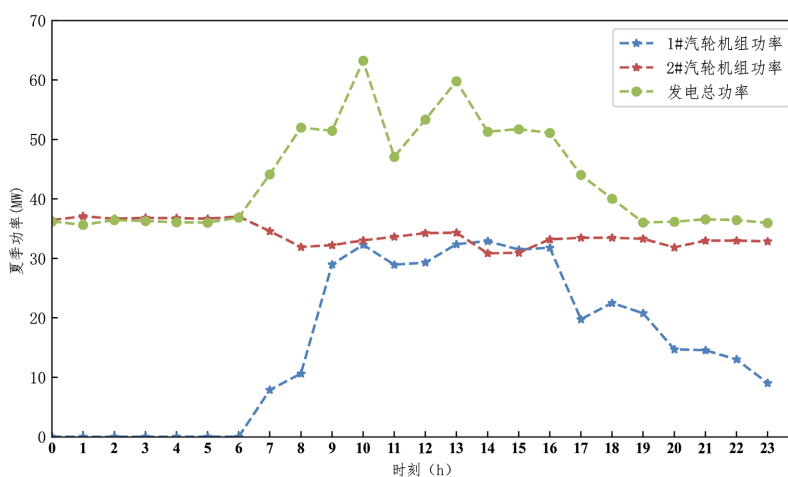


Figure 5. Curve of low pressure heating flow rate with time in non-heating season (in a day)  
 图 5. 非采暖季机组发电功率随时间变化曲线(一天之内)

### 3.2. 评价效果指标

为了验证预测模型的预测性能, 本文选取了四种常用的评价指标, 分别是 R 方值( $R^2$  score)、均方根误差(RMSE)、平均绝对百分比误差(MAPE)和平均绝对误差(MAE), 公式如下:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (18)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (19)$$

$$MAPE = \left( \frac{\sum_{i=1}^n |(y_i - \hat{y}_i)/y_i|}{n} \right) * 100\% \quad (20)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (21)$$

式(18)~(21)中,  $n$  为样本个数,  $y_i$  为实际值,  $\hat{y}_i$  为预测值,  $\bar{y}^2$  为实际值的平均值。RMSE、MAPE、MAE 的值越趋近于 0,  $R^2$  的值越趋于 1, 则预测效果越好。

### 3.3. 基于 XGBoost 的功率影响因素特征提取

实验所采集的数据维度为 250, 远远超过深度学习输入数据的合适维度, 需要对数据进行降维处理, 降低学习任务的难度。模型的输入特征只需要与汽轮机功率有关的相关特征, 即需要对原始数据集进行特征选择。本文采用物理机理分析和 XGBoost 算法来衡量各个特征与功率的相关程度。

首先, 运用物理机理分析, 即根据汽轮机实际运行机理, 排除对汽机功率无直接相关影响的属性。如中压减温减压器电动压力调节阀阀位、电动给水泵进水压力等。

最后, 使用 XGBoost 算法来衡量各属性与目标特征的相关程度。调参对 XGBoost 的效果来说至关重要。其中, 最为重要的三个参数分别为学习率(learning\_rate)、最大深度(max\_depth)和树的棵树(n\_estimators)。通过控制变量法, 先设定 learning\_rate 的值为 0.08, max\_depth 的值为 21, 然后, 通过 9 次调整 n\_estimators 的取值, 得到 9 组数据, 如表 1 所示。

Table 1. N\_estimators correspondence error and score

表 1. N\_estimators 调整的对应误差与得分

n_estimators	RMSE	$R^2$ score
150	0.0094748	0.999999423
200	0.0094693	0.999999424
250	0.0094687	0.999999424
255	0.0094686	0.999999424
260	0.0094685	0.999999424
265	0.0094685	0.999999424
270	0.0094685	0.999999424
290	0.0094685	0.999999424
300	0.0094685	0.999999424

基于这组数据可得到关系图, 如图 6 所示。

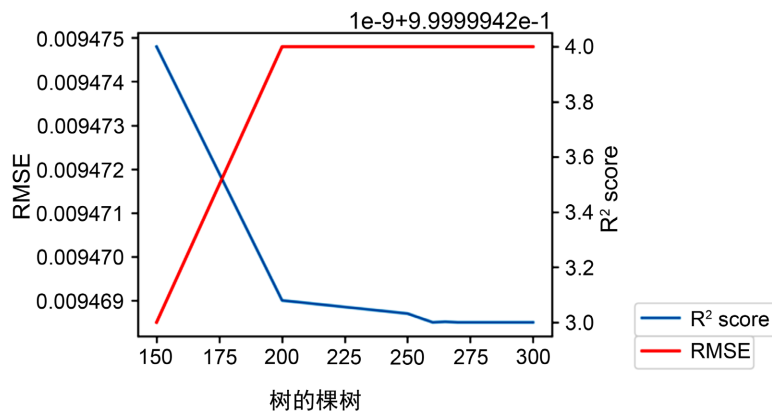


Figure 6. N\_estimators parameter tuning diagram

图 6. N\_estimators 调参关系图

由表 1 和图 6 可知, 当 learning\_rate 和 max\_depth 固定不变, n\_estimators 取值为 260 时, RMSE 出现收敛的第一个拐点且 R<sup>2</sup>score 已经收敛。故, 260 为 n\_estimators 的最佳取值。

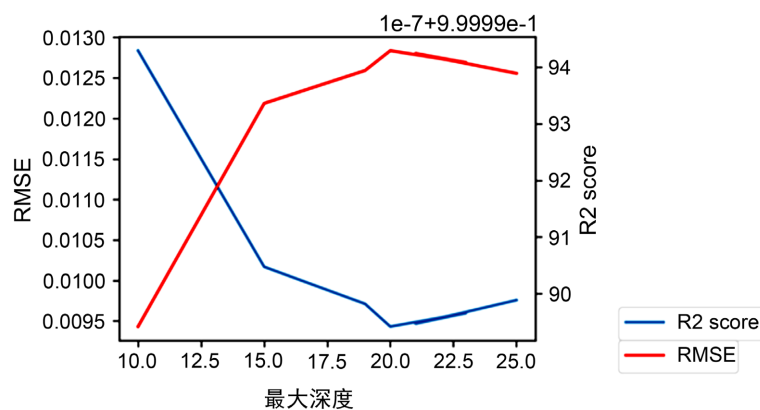
同理可得, 固定 learning\_rate 和 n\_estimators 的取值, 调整 max\_depth 的取值可得对应的误差与得分, 如表 2 所示。

**Table 2.** Max\_depth correspondence error and score

**表 2.** Max\_depth 调整的对应误差与得分

max_depth	RMSE	R <sup>2</sup> score
10	0.012836066	0.999998942
15	0.010169256	0.999999336
19	0.009712356	0.999999394
20	0.009432302	0.999999429
21	0.009468482	0.999999424
23	0.009598712	0.999999408
25	0.009757911	0.999999389

由这组数据可得关系图, 如图 7 所示。



**Figure 7.** Max\_depth parameter tuning diagram

**图 7.** Max\_depth 调参关系图

由表 2 和图 7 可知, 当 n\_estimators 和 learning\_rate 固定不变, max\_depth 取值为 20 时, RMSE 取得最小值, R<sup>2</sup>score 取得最大值。故, 20 为 max\_depth 的最佳取值。

同理亦得, 固定 max\_depth 和 n\_estimators 的取值, 调整 learning\_rate 得取值可得对应误差与得分, 如表 3 所示。

**Table 3.** Learning\_rate correspondence error and score

**表 3.** Learning\_rate 调整的对应误差与得分

learning_rate	RMSE	R <sup>2</sup> score
0.02	0.100070676	0.999935703
0.03	0.008977868	0.999999482
0.04	0.006271166	0.999999747

Continued

0.05	0.007835362	0.999999606
0.06	0.008827865	0.9999995
0.07	0.00935316	0.999999438
0.08	0.009432302	0.999999429
0.09	0.010331611	0.999999315
0.1	0.010718653	0.999999262

由这组数据可得关系图, 如图 8 所示。

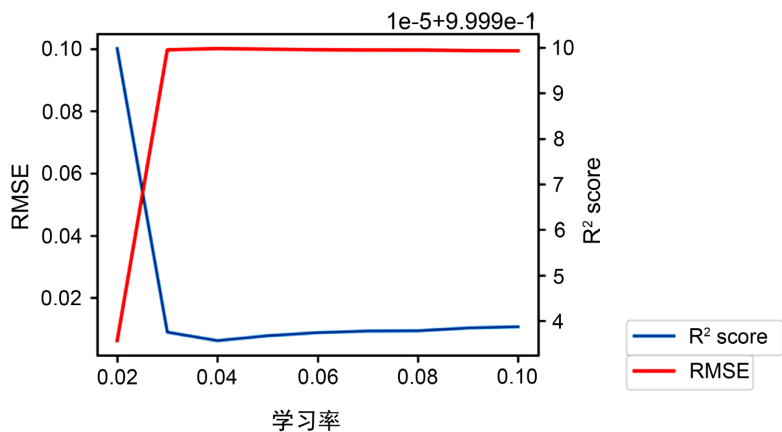


Figure 8. Learning\_rate parameter tuning diagram  
图 8. Learning\_rate 调参关系图

由表 3 和图 8 可知, 当  $n\_estimators$  和  $max\_depth$  固定不变,  $learning\_rate$  取值为 0.04 时, RMSE 出现收敛得第一个拐点并且  $R^2$  score 已经收敛。故, 0.04 为  $learning\_rate$  得最佳取值。

将最佳参数确定后, 对原始电力负荷数据进行特征选择, 可得特征重要性排名和权重值的关系图, 如图 9 所示。

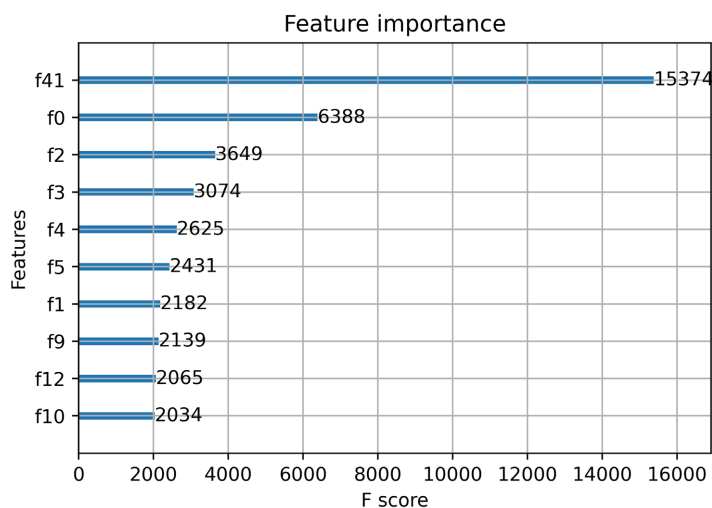


Figure 9. Feature importance ranking and weight diagram  
图 9. 特征重要排名和权重图

图中显示的数值为各个特征的影响权重。从高到低分别代表：锅炉出口蒸汽压力、发电总负荷、供电总负荷、炉给水温度等。

为了避免维数灾难，本文选择了上述 10 个相关特征作为混合深度学习模型的输入参数。

### 3.4. 数据预处理

原始数据集的采集周期为一分钟，本文将预测给定一段时间序列的下一个小时的功率变化，故，每 60 个数据点为周期提取原数据，得到以小时为单位的衍生数据集。

因为特征数据之间不在同一个数量级，会导致求解目标函数最优解的迭代次数增多。为了解决这一问题并且加快随机梯度下降算法搜索到最优解的过程以及加快模型的训练，需要将数据进行归一化处理，把数据特征映射到[0, 1]区间内，去除宏观维度的显性特征。本文采用的是最小最大值归一化(Min-Max Normalization)，计算公式如下：

$$X'_i = (X_i - X_{\min}) / (X_{\max} - X_i) \quad (22)$$

式中， $X'_i$  是归一化之后的数据； $X_i$  为原始数据； $X_{\max}, X_{\min}$  分别代表原始数据中的最大值和最小值。

### 3.5. 模型建立和参数设置

本文提出的混合深度学习模型是基于 CNN-LSTM-A 网络的热电联产汽轮机组功率预测，其模型具体结构参数如表 4。

Table 4. Model structure parameters

表 4. 模型结构参数

层号	层类别	神经元	备注
1	输入层	40 × 11	
2	卷积层	40 × 64	核 1 × 6/1
3	池化层	40 × 64	最大池化 1 × 6/2
4	Dropout 层	40 × 64	保留率为 0.8
5	LSTM 层	40 × 128	
6	全连接层	40 × 128	
7	Attention 层	40 × 128	
8	Flatten 层	5120 × 1	
12	输出层	1 × 1	Softmax

本文以 40 个数据点为时间步长，预测目标数据，原始时序数据共有 11 个维度(40 × 11)经输入层输入，通过卷积层提取数据特征、最大池化层降采样以及 Dropout 层舍弃 20% 的神经元，转化为 40 × 1 的向量输入 LSTM 层。LSTM 层的输出通过全连接层整合输入 Attention 层计算注意力分布，然后进行权值分配优化，优化后的权值连同 LSTM 的输出一并输入 Flatten 层压平为一维向量，最终在输出层以每 40 个时间步长为—组输出预测结果。

### 3.6. 模型训练

本文建立三组模型，每组模型都分别以汽轮机组 1 数据训练、汽轮机组 2 数据验证和汽轮机组 2 数据训练、汽轮机组 1 数据验证，并选择 LSTM、CNN-LSTM 为基模型，验证 CNN-LSTM-A 模型功率的

预测能力和优越性。

对于整个 CNN-LSTM-A 网络功率预测模型的训练, 本文采用按时间展开的误差反向传播算法 (BPTT), 即对神经网络按时间顺序展开为一个深层网络, 然后采用误差反向传播(back propagation, BP) 算法对展开后的网络进行训练。为了控制网络的学习率, 预防梯度消失、收敛速度慢等问题, 本文使用 Adam 优化算法[21]更新网络参数, 初始学习率设置为 0.01; 在 LSTM 层前引入 Dropout 正则化方法, 避免过度拟合训练数据, 保留率设置为 0.8。神经网络训练参数设置为: 最大迭代次数 epoch 为 50, 小批量大小 Batch size 为 1024; 训练方式: CPU, 并且样本集按照 4:1 的比例分为训练集和测试集。

## 4. 实验结果与分析

### 4.1. 训练过程评价

如图 10 为数据在归一化后送入模型训练迭代次数为 50 次时, 三个模型均方误差(MSE)的变化过程。

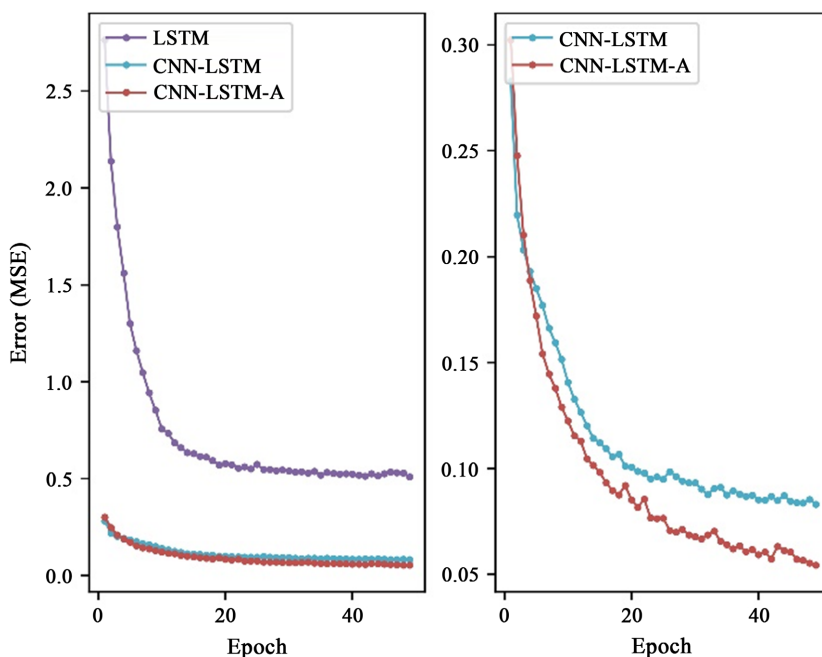


Figure 10. Model MSE change

图 10. 模型 MSE 变化

可见传统的 LSTM 在迭代了 50 次之后没有明显的收敛, 而 CNN-LSTM 和 CNN-LSTM-A 在训练初期很快就达到收敛。所以传统的 LSTM 在预测误差上要显著劣于添加了 CNN 结构的两种模型。并且在图中也可以直观地看到, 本文提出的方法添加了 Attention 机制, 根据输入信息的注意力优化权值分配, 其收敛速度要更快, 误差值要更低。

图 11 为数据经过 CNN、LSTM 层输出后再作为输入矩阵进入 Attention 层后, 注意力对输入矩阵进行权值分配的热力图。输入的矩阵形状为  $40 \times 128$ , 40 为模型中设定的时间步, 代表时序信息; 128 是输入的 10 维特征经过 CNN、LSTM 层后维度变化为 128, 代表输入特征信息。

可以看出, Attention 层对输入矩阵的权值进行了分配。其根据所有输入的信息被划分为各自所需要的注意力权值, 图中颜色越高亮的部分代表其影响因子越大, 不再必须把所有的信息都直接传输到神经网络, 大大节约了计算时间和资源。

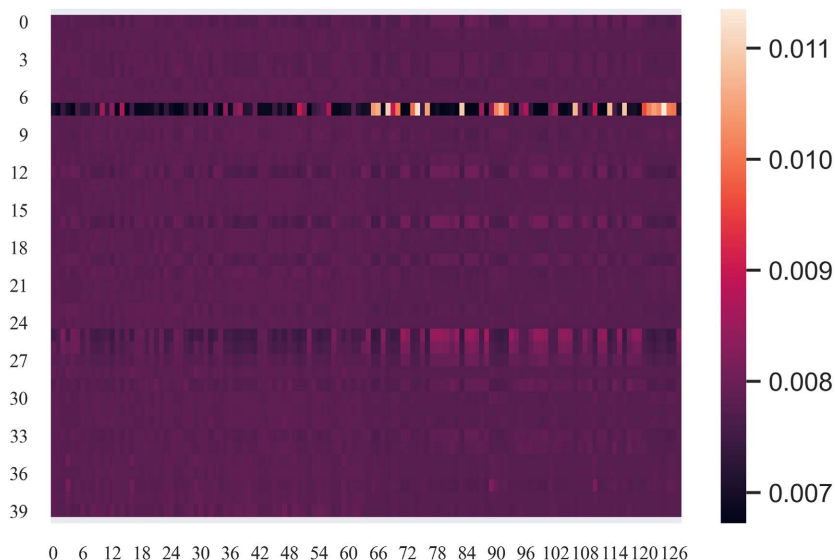


Figure 11. Weight distribution heat map

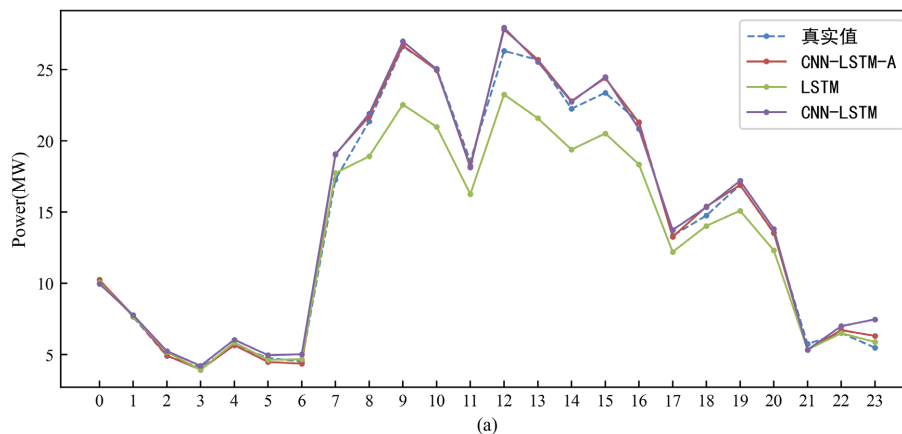
图 11. 权值分配热力图

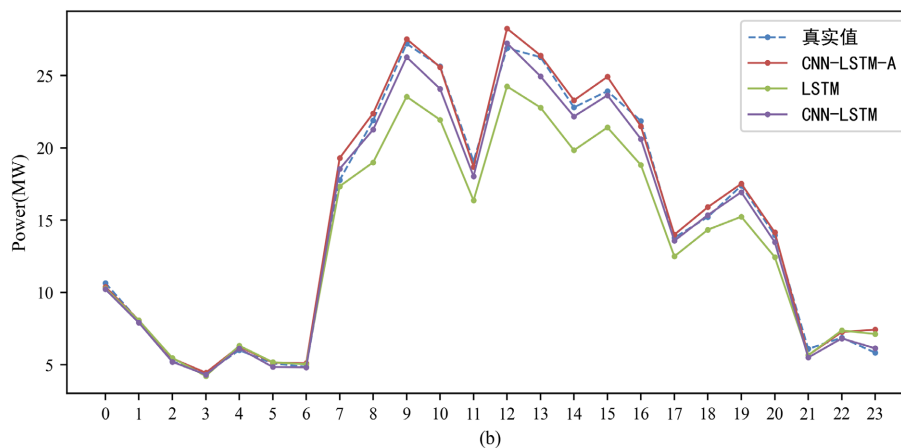
#### 4.2. 多模型预测结果对比

以小时为单位预测汽轮机组的功率变化。三个模型预测结果与真实值对比如图 12 所示。由预测结果可知, 本文提出的 CNN-LSTM-A 负荷预测模型的预测值与真实值贴合紧密, 误差较小, 而 LSTM 模型预测结果则与真实值有着较大差距, 效果最差。对每个汽轮机组每个模型分别进行 10 次验证, 预测误差平均值如图 13 所示。

本文提出的 CNN-LSTM-A 负荷预测模型在对两台汽轮机组进行功率预测时, 其 RMSE 均最小, MAPE 均小于 3.5%, 具有极高的准确率。

在 RMSE 衡量方面, CNN-LSTM-A 应用于两台汽轮机组功率预测较 LSTM 分别提升 1.815MW、1.57MW, 较 CNN-LSTM 分别提升 0.066MW、0.026MW; 在 MAPE 衡量方面, CNN-LSTM-A 较 LSTM 预测的两台汽轮机分别提升 7.3%、5.7%, 较 CNN-LSTM 分别提升 1.3%、0.6%; 在 MAE 衡量方面 CNN-LSTM-A 较 LSTM 预测的两台汽轮机分别提升 1.423MW、1.115MW, 较 CNN-LSTM 分别提升 0.075MW、0.053MW。对比结果表明, CNN-LSTM-A 混合模型预测结果不仅更准确, 在实际电力生产中具有更加稳定的应用效果。

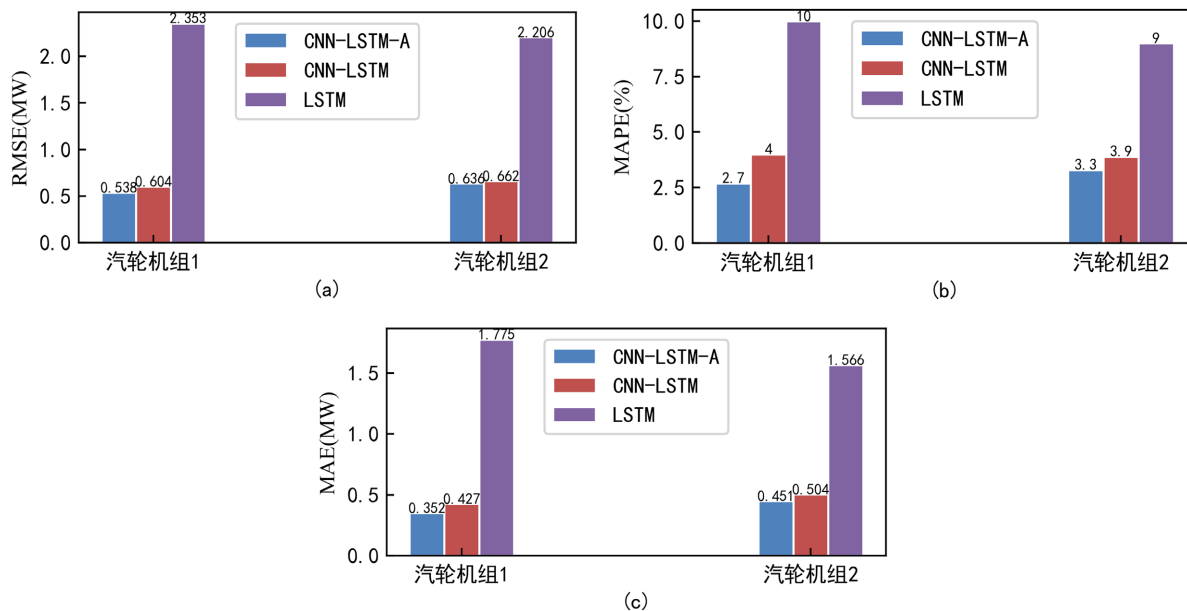




(a) 1号汽轮机组; (b) 2号汽轮机组

Figure 12. Comparison of model prediction results

图 12. 预测模型结果对比



(a) 均方根误差; (b) 平均绝对百分比误差; (c) 平均绝对误差方差

Figure 13. Prediction error

图 13. 预测误差

### 4.3. 基于混合模型的汽轮机功率预测

本文将训练效果最佳的源域模型参数保存, 分别以 2 台汽轮机组的在线检测数据为目标域数据导入模型, 自 2021 年 5 月 30 日开始预测, 预测 12 天的负荷, 并与真实值进行对比, 两台汽轮机组的预测结果及其误差如图 14。

由预测结果可知, 基于 CNN-LSTM-A 的热电厂汽轮机功率预测的结果同真实数据有着高度的重合, 其误差分布集中在 0 的两侧, 均值大于 0, 表明模型的预测结果相对于真实数据较大。在应用于功率调度时, 采用偏大的预测值, 汽轮机组功率有一定余量, 相较于预测值偏小时各汽轮机组分配时实际功率不足的问题来说更加安全可靠。



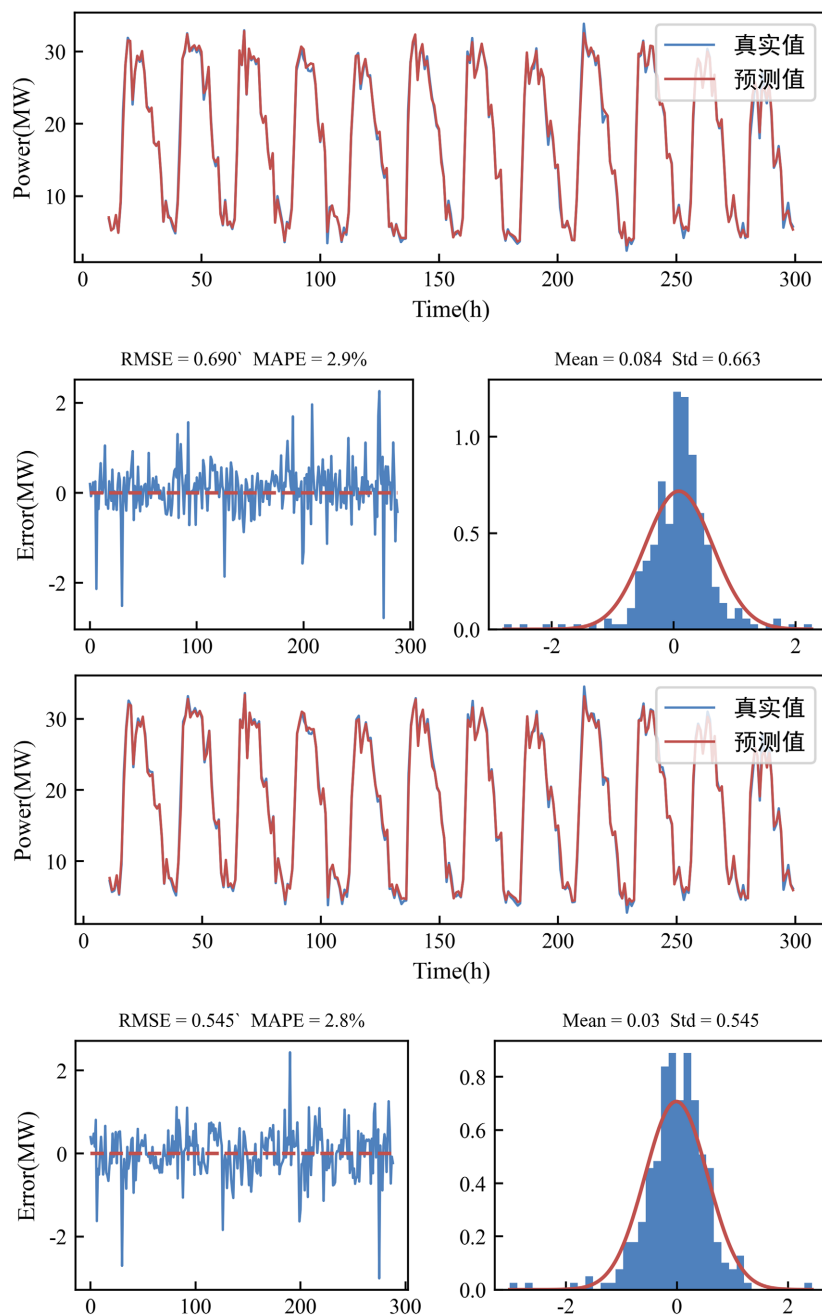


Figure 14. Power forecast and error of steam turbine  
 图 14. 汽轮机组功率预测及其误差

## 5. 结论

1) 基于 CNN-LSTM-A 的汽轮机组功率预测模型, 相较于传统的 CNN-LSTM, 添加了 Attention 机制, 使模型能够更精准地分配权重。其模型的误差收敛速度更快, 误差值更低, 能够精准地预测未来各汽轮机组的功率, 平均绝对百分比误差均小于 3.5%。

2) 相对于传统深度学习模型, CNN-LSTM-A 模型具有更高的鲁棒性和泛化能力。即使在复杂的多变量影响因素下, 依然可以有着很高的功率预测精度; 在均方根误差衡量方面, CNN-LSTM-A 应用于两

台汽轮机组功率预测较 LSTM 分别提升 1.815MW、1.57MW, 较 CNN-LSTM 分别提升 0.066MW、0.026MW; 在平均绝对百分比误差衡量方面, CNN-LSTM-A 较 LSTM 分别提升 7.3%、5.7%, 较 CNN-LSTM 分别提升 1.3%、0.6%; 本文将深度学习算法应用于热电联产的功率预测中, 促进了人工智能技术在电力行业的应用。设备运行维护人员可根据模型的预测结果进行后续的功率调度优化。

## 参考文献

- [1] 关哲. 热电厂汽轮机运行节能降耗研究[J]. 电气技术与经济, 2021(2): 15-17.
- [2] IEA (International Energy Agency) (2019) World Energy Outlook 2019. <https://www.iea.org/reports/world-energy-outlook-2019>
- [3] 郑伟, 姚远, 代邦武, 常玉清, 孙晓辉. 基于数据挖掘的燃煤机组厂级负荷经济调度[J]. 热力发电, 2021, 50(7): 78-83.
- [4] 栗然, 马涛, 张潇, 回旭, 刘英培, 尹晓钢. 基于卷积长短期记忆神经网络的短期风功率预测[J]. 太阳能学报, 2021, 42(6): 304-311.
- [5] 张航, 李洪双. 结构优化设计的改进交叉熵支持向量机方法[J]. 西北工业大学学报, 2018, 36(3): 448-455.
- [6] 常峰铭. 基于深度学习的智能电网短期电力负荷预测研究[D]: [硕士学位论文]. 湘潭: 湘潭大学, 2019.
- [7] 栗然, 孙帆, 丁星, 韩怡, 刘英培, 严敬汝. 考虑多能时空耦合的用户级综合能源系统超短期负荷预测方法[J]. 电网技术, 2020, 44(11): 4121-4134.
- [8] 刘亚琿, 赵倩. 基于聚类经验模态分解的 CNN-LSTM 超短期电力负荷预测[J]. 电网技术, 2021, 45(11): 4444-4451.
- [9] 杨振舰, 尚佳美, 张众维, 张艳, 刘树东. 基于残差注意力机制的图像去雾算法[J]. 西北工业大学学报, 2021, 39(4): 901-908.
- [10] Friedman, J., Hastie, T. and Tibshirani, R. (2000) Additive Logistic Regression: A Statistical View of Boosting. *Annals of Statistics*, **28**, 337-407. <https://doi.org/10.1214/aos/1016218223>
- [11] Friedman, J. (2002) Stochastic Gradient Boosting. *Computational Statistics & Data Analysis*, **38**, 367-378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- [12] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/A:1010933404324>
- [13] Friedman, J.H. and Popescu, B.E. (2003) Importance Sampled Learning Ensembles. *Journal of Machine Learning Research*, **9**, 305, 1-32.
- [14] 张科, 苏雨, 王靖宇, 王霁宇, 张彦华. 基于融合特征以及卷积神经网络的环境声音分类系统研究[J]. 西北工业大学学报, 2020, 38(1): 162-169.
- [15] 杨格兰, 邓晓军, 刘琮. 基于深度时空域卷积神经网络的表情识别模型[J]. 中南大学学报(自然科学版), 2016, 47(7): 2311-2319.
- [16] 董秋成, 吴爱国, 董娜, 冯伟. 用于卷积神经网络图像预处理的目标中心化算法[J]. 中南大学学报(自然科学版), 2019, 50(3): 579-586.
- [17] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) Imagenet Classification with Deep Convolutional Neural Networks. *26th Annual Conference on Neural Information Processing Systems, Lake Tahoe*, 3-6 December 2012, 1097-1105.
- [18] 谢清林, 陶功权, 温泽峰. 基于一维卷积神经网络的地铁钢轨波磨识别方法[J]. 中南大学学报(自然科学版), 2021, 52(4): 1371-1379.
- [19] 刘畅, 陈雯柏. 一种基于 MSDCNN-LSTM 的设备 RUL 预测方法[J]. 西北工业大学学报, 2021, 39(2): 407-413.
- [20] Yang, Y.R., Xiong, Q.Y., Wu, C., Zou, Q.H., Yu, Y., Yi, H.L. and Gao, M. (2021) A Study on Water Quality Prediction by a Hybrid CNN-LSTM Model with Attention Mechanism. *Environmental Science and Pollution Research International*, **28**, 55129-55193. <https://doi.org/10.1007/s11356-021-14687-8>
- [21] 张珂, 杨歆豪, 张嘉慧, 倪俊芳. 基于高次指数平滑动态边界限制的深度学习优化算法[J]. 信息与控制, 2021, 50(6): 685-693.