

# 基于随机森林与长短期记忆网络的电力负荷预测方法

董彦军<sup>1</sup>, 王晓甜<sup>2</sup>, 马红明<sup>2</sup>, 王立斌<sup>1</sup>, 李梦宇<sup>2</sup>, 岳凡丁<sup>3</sup>, 袁欢<sup>3\*</sup>

(1. 国网河北省电力有限公司, 河北省 石家庄市 050081;

2. 国网河北省电力有限公司营销服务中心, 河北省 石家庄市 050081;

3. 西安交通大学电气工程学院, 陕西省 西安市 710049)

## Power Load Forecasting Method Based on Random Forest and Long Short-term Memory

DONG Yanjun<sup>1</sup>, WANG Xiaotian<sup>2</sup>, MA Hongming<sup>2</sup>, WANG Libin<sup>1</sup>, LI Mengyu<sup>2</sup>, YUE Fanding<sup>3</sup>, YUAN Huan<sup>3\*</sup>

(1. State Grid Hebei Electric Power Co., Ltd., Shijiazhuang 050081, Hebei Province, China;

2. Marketing Service Center of State Grid Hebei Electric Power Co., Ltd., Shijiazhuang 050081, Hebei Province, China;

3. School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, Shaanxi Province, China)

**Abstract:** Power load has the characteristics of non-linearity and timing. In order to dig deeper into the importance of characteristic variables for power load forecasting and obtain higher accuracy of power load forecasting, in this paper, a hybrid load forecasting model based on random forest (RF) algorithm and long-term and short-term memory (LSTM) neural network model is applied to load forecasting. First, according to the time and date factors and climate factors, a high-dimensional characteristic data set is established as the input of the random forest model, and then the important features selected by the random forest algorithm and combined with historical load as the input of the LSTM network model. The particle swarm optimization algorithm is used to optimize the parameters of the LSTM network model to get the final RF-LSTM hybrid model and the load forecasting results. This method is applied to predict the power load of a certain station in Hebei. The results show that the hybrid prediction model proposed in this paper has better prediction accuracy than random forest model, LSTM network model and BP neural network without characteristic variable screening.

**Keywords:** random forest; long short-term memory; power load forecasting; hybrid model

**摘要:** 电力负荷具有非线性及时序性的特点, 为了深入研究各特征变量对于电力负荷预测的重要性, 进而获得更高的

电力负荷预测精度, 提出了基于随机森林 (random forest, RF) 算法及长短期记忆网络 (long short-term memory, LSTM) 的混合负荷预测模型。首先根据时间日期因素及气候因素建立高维特征数据集作为随机森林模型的输入, 通过随机森林算法筛选出重要特征量, 并使其与历史负荷结合作为LSTM模型的输入, 经过粒子群算法对LSTM模型进行参数寻优后得到RF-LSTM混合模型及负荷预测结果。使用该方法对河北电网某台区的电力负荷进行预测, 结果表明该混合模型的预测精度比未经特征变量筛选的传统单一的随机森林算法、LSTM模型以及BP神经网络更为理想。

**关键词:** 随机森林; 长短期记忆网络; 电力负荷预测; 混合模型

## 0 引言

电力系统以向各用户提供达到质量标准的电能为主要任务, 以满足社会各类负荷的用电需求, 其稳定运行需要实时动态平衡发电量与负荷变化。然而当前电能的大量存储难以实现, 负荷波动具有明显的非线性及随机性<sup>[1]</sup>, 因此需要对电力负荷进行准确预测, 从而合理分配用电负荷, 保证电网经济稳定运行<sup>[2]</sup>。

非线性和时序性是电力负荷的两大特点<sup>[3]</sup>。对于电力系统负荷的预测, 国内外研究方法主要分为两类: 传统方法与新型人工智能方法<sup>[4]</sup>。传统方法以时间序列法为代表, 如傅里叶展开法<sup>[5]</sup>、多元线性回归法<sup>[6]</sup>等。这些方法具有充分考虑电力负荷数据的时序性、计算速度快等优点, 然而其数据回归能力较弱,

基金项目: 国网河北省电力有限公司科技项目 (SGHEDK00DYJS1900303); 国家自然科学基金项目 (51877170)。

Science and Technology Foundation of State Grid Hebei Electric Power Co., Ltd. (SGHEDK00DYJS1900303); National Science Foundation of China (51877170).

且要求数据时间序列具有较好的平稳性,因此无法对具有非线性关系的数据进行准确预测。新型人工智能方法则能够较好地拟合非线性数据,文献[7-8]使用较为普遍的反向传播(back propagation, BP)神经网络进行负荷预测,但BP神经网络的学习能力相对较差,预测精度有待提高。文献[9-10]使用了模糊推理法,但该方法的计算速度过慢且精度较低。文献[11-12]使用了支持向量回归(support vector regression, SVR)算法预测负荷。文献[13]通过决策树进行预测。然而这些算法都没有考虑电力负荷的时序性,在预测中需要人为添加时间特征才能在一定程度上保证预测的精度。长短期记忆网络(long short-term memory, LSTM)兼顾了数据的时序性和非线性,训练时间短且预测精度高,因此被广泛应用于电力负荷的预测。

此外,随着电力大数据的发展及能源互联网的建设不断深入,海量的电力数据为负荷预测提供了坚实基础。电力负荷受时间、日期及天气等众多因素影响,但在负荷预测的特征集之中,并非特征因素越多预测精度就越高,过多的特征量会增加预测模型的复杂度,降低预测精度。目前关于负荷预测的研究大多未对特征因素进行筛选,或仅采用主观选择的方式确定预测特征集中的特征因素,而对特征因素的选择会直接影响预测结果。因此,为了避免特征量过多及人为主观筛选特征量对负荷预测精度造成的不利影响,需要通过科学合理的方法对特征量进行筛选。随机森林(random forest, RF)算法作为一种鲁棒性高、学习能力强的智能分类算法,具备度量变量重要性的能力,能够分析复杂且相互作用的特征,因此被广泛应用于高维度数据特征的选择。为实现电力负荷的准确预测,本文提出一种基于随机森林算法和长短期记忆网络的混合模型负荷预测方法,即RF-LSTM混合模型。该模型结合了随机森林和LSTM各自的特点,先将天气因素、日期因素等高维特征集输入随机森林模型进行重要性评估,筛选出重要特征变量后输入LSTM预测模型进行负荷预测,从而兼顾负荷数据的非线性和时序性,提高负荷预测精度<sup>[14-16]</sup>。结合河北电网某台区的实际数据,利用该混合模型进行负荷预测,结果表明,本文提出的模型可以有效提高负荷预测的精度,降低预测误差。

## 1 电力负荷预测特征集

特征集对模型的预测结果有着决定性的影响,高

精度的负荷预测结果要以合理的特征集作为前提。影响电力负荷的特征因素非常多,就目前研究分析结果来看,主要的影响因素除历史负荷外还有时间日期因素、气候因素等<sup>[17]</sup>。然而模型预测精度并非绝对与特征因素数量呈正相关,模型输入的特征量维数过多时易造成模型结构复杂,甚至降低预测精度。因此本文提出在对负荷正式预测前先对高维特征量进行筛选。首先构建高维特征集从而为后续筛选特征量提供充足的备选,具体如下。

1) 时间日期因素:节假日、双休日及时刻等时间因素对电力负荷的影响较大,因此构建包含月、日、是否为节假日、是否为工作日、周日期、当日的具体小时共6维时间日期特征。

2) 气候因素:不同的温度、湿度和天气对于电力负荷均有一定影响。考虑气候因素,构建包含风力、湿度、当日最低温度、当日最高温度、露点温度、天气类型、空气质量、日照时数共8维气候特征。

本文构建的包括时间日期因素、气候因素在内共14维特征向量如表1所示,为后续特征向量筛选提供特征集。

表1 预测特征集  
Table 1 Feature set of prediction data

影响因素	特征	特征描述
时间日期因素	月	1~12代表1月至12月
	日	1~31代表1号至31号
	周日期	1~7代表周一至周日
	工作日	0代表工作日, 1代表非工作日
	节假日	0代表非节假日, 1代表节假日
	当日小时	1~24代表1点至24点
气候因素	最高温度	当日温度的最高值, °C
	露点温度	当日露点温度, °C
	最低温度	当日温度的最低值, °C
	湿度	当日相对湿度, %
	天气类型	1代表晴天或多云, 2代表小雨或小雪, 3代表大雨、大雪、暴雨
	空气质量	空气质量指数(AQI)
	风力	1~5表示1级至5级
	日照时数	1~24表示当日日照小时数

## 2 RF-LSTM混合模型

### 2.1 随机森林分类原理

随机森林算法是对传统决策树的继承和改进,能

够分析复杂且相互作用的特征, 在处理存在缺失值的数据时学习速度较快并具有较高的鲁棒性。此外, 随机森林算法中的变量重要性度量作为其重要特点可以用于高维度数据特征的选择, 近年来在分类、特征选择等问题中得到了广泛的应用。本文采用的随机森林算法本质上是包含多个分类回归树的组合分类器, 利用随机重采样 (bootstrap) 技术和节点随机分裂技术构建了若干决策树, 在随机重采样的过程中, 部分未被选中的样本被称为袋外 (out of bag, OOB) 数据。使用OOB数据对随机森林模型进行评估可得到OOB误差, 本研究通过分析OOB误差从而得到各特征变量的重要性, 其原理可理解为: 当OOB数据自变量发生轻微扰动时, OOB误差增加幅度越大则该变量越重要。因此, OOB误差可以用于定量评价特征变量的重要性, 进而对高维特征数据进行选择<sup>[18-20]</sup>。

随机森林算法选择变量的流程 (伪代码) 如下。

1) 假设随机森林算法中共有 $k$ 棵树, 各特征变量分别为 $x_1, x_2, \dots, x_n$ , for  $i = 1 : k$

①针对其中的每棵树都通过随机重采样从数据集 $N$ 中有放回地随机抽取一定规模的数据作为样本, 构成样本训练子集 $N_i$ , 未被抽取的数据则构成 $b$ 个OOB数据。

②在 $N_i$ 中重复步骤a—c, 每次循环中使决策树按最大限度生长, 不对其剪枝, 得到决策树 $T_i$ 。

a) 假设共输入 $M$ 个特征属性, 随机抽取其中的 $m$ 个属性作为当前决策树分裂的属性集。

b) 从 $m$ 个特征变量中选择最佳的变量 $j$ 和切分点 $s$ 得到 $\theta_i(j, s)$ 。

c) 将该节点按照 $\theta_i(j, s)$ 切分成2个子节点。

end for

2) 生成 $k$ 棵决策树构成随机森林时, 对每棵决策树 $T_i$ 对应的 $b$ 个OOB数据进行投票, 从而获得OOB数据中每个样本的投票分数为

$$s_1, s_2, \dots, s_b \quad (1)$$

3) 随机改变OOB数据样本中各特征变量 $x_i$ 的数值从而生成新的OOB数据测试样本, 并通过随机森林对新OOB数据投票, 得到

$$\begin{bmatrix} s_{11} & s_{12} & \dots & s_{1b} \\ s_{21} & s_{22} & \dots & s_{2b} \\ \dots & \dots & \dots & \dots \\ s_{k1} & s_{k2} & \dots & s_{kb} \end{bmatrix} \quad (2)$$

4) 求特征变量 $x_i$ 的重要性评分

$$c_i = \sum_{j=1}^b \frac{(s_j - s_{ij})}{b} \quad (3)$$

式中:  $s_j$ 和 $s_{ij}$ 分别表示变量改变前后第 $i$ 棵树的OOB误差率;  $c_i$ 代表各特征向量对于分类过程的贡献大小, 可以衡量各特征的独立分类能力, 因此有助于在分类过程中确定各特征的重要性, 为选择特征提供依据。随机森林算法模型流程如图1所示。

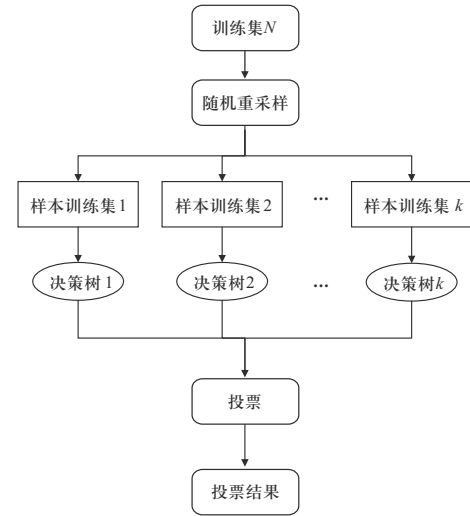


图1 随机森林算法流程图

Fig. 1 Flow chart of random forest algorithm

## 2.2 LSTM预测模型

LSTM作为循环神经网络 (recurrent neural network, RNN) 的继承和发展, 经Graves<sup>[21]</sup>改进后, 解决了RNN模型在训练过程中经常出现的问题——梯度消亡。此外, 面对时序性及非线性较强的电力负荷数据时, LSTM能够很好地掌握时间序列所依赖的信息, 因此在负荷预测领域得到了很好的应用和发展<sup>[22-23]</sup>。LSTM基本单元模型如图2所示。

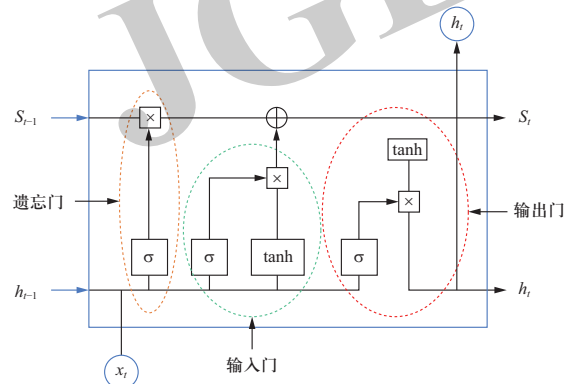


图2 LSTM模型图

Fig. 2 Diagram of LSTM model



一个LSTM单元由输入门、遗忘门、输出门和记忆单元构成,其核心记忆单元用于描述该单元当前的状态<sup>[24]</sup>。模型中的3个控制门分别与乘法单元连接,实现对该单元的输入、记忆单元和输出的控制。记忆单元的遗忘部分由记忆单元 $S_t$ 、遗忘门的输入 $x_t$ 以及中间输出 $h_{t-1}$ 共同决定,记忆单元的保留向量则由输入门中的 $x_t$ 分别经过 $\tanh$ 函数和 $\text{sigmoid}$ 函数变换后的结果共同决定,更新过后的 $S_t$ 和输出 $o_t$ 共同决定中间输出 $h_t$ ,其计算公式如式(5)一式(10)所示<sup>[25]</sup>。

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \quad (5)$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \quad (6)$$

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \quad (7)$$

$$g_t = \phi(W_{gx}x_t + W_{gh}h_{t-1} + b_g) \quad (8)$$

$$S_t = g_t \odot i_t + S_{t-1} \odot f_t \quad (9)$$

$$h_t = \phi(S_t) \odot o_t \quad (10)$$

式中: $\sigma$ 和 $\phi$ 分别表示sigmoid函数和 $\tanh$ 函数; $i_t$ 、 $o_t$ 、 $f_t$ 、 $h_t$ 、 $g_t$ 和 $S_t$ 分别表示输入门、输出门、遗忘门、中间输出节点、中间输入节点以及状态单元; $W_{ix}$ 、 $W_{ih}$ 、 $W_{ox}$ 、 $W_{oh}$ 、 $W_{fx}$ 、 $W_{fh}$ 、 $W_{gx}$ 以及 $W_{gh}$ 分别表示输入 $x_t$ 和中间输出 $h_{t-1}$ 在与对应门相乘时的矩阵权重; $\odot$ 表示两向量中的元素按位相乘; $b_i$ 、 $b_o$ 、 $b_f$ 及 $b_g$ 分别表示各对应门的修正向量。

根据对LSTM模型原理的分析,本实验以负荷预测值 $Y_{\text{pred}}$ 为目标建模变量,确定负荷预测的步骤如下:

1) 将随机森林算法筛选后的重要时间日期因素及气候因素特征向量与历史负荷数据结合作为特征建立预测特征集 $X_L = \{x_1, x_2, \dots, x_7, y\}$ ,其中 $x_1 \sim x_7$ 为随机森林算法所筛选的7个重要特征量, $y$ 为历史负荷数据。

2) 根据图3构造LSTM网络模型,将特征集 $X_L$ 作为输入层,中间为隐藏层,最后输出负荷预测值 $Y_{\text{pred}}$ 。

根据数据集的特征确定LSTM网络模型的网络隐藏层层数及每层包含的记忆单元数,并通过粒子群(particle swarm optimization, PSO)算法对训练次数等参数进行寻优。其中模型的激活函数及训练时的损失函数分别为式(11)和式(12)。

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (11)$$

$$L(Y_{\text{pred}}, f(X_L)) = (Y_{\text{pred}} - f(X_L))^2 \quad (12)$$

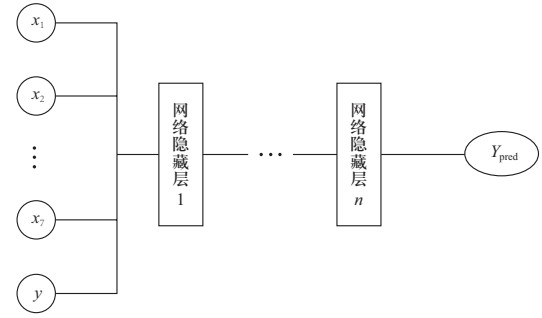


图3 LSTM网络模型

Fig. 3 LSTM network model

### 2.3 RF-LSTM混合预测模型

本文提出的RF-LSTM网络混合预测模型的流程如图4所示,其中包括对负荷及相关数据的采集和预处理、随机森林模型及LSTM模型的构建以及电力负荷的预测。

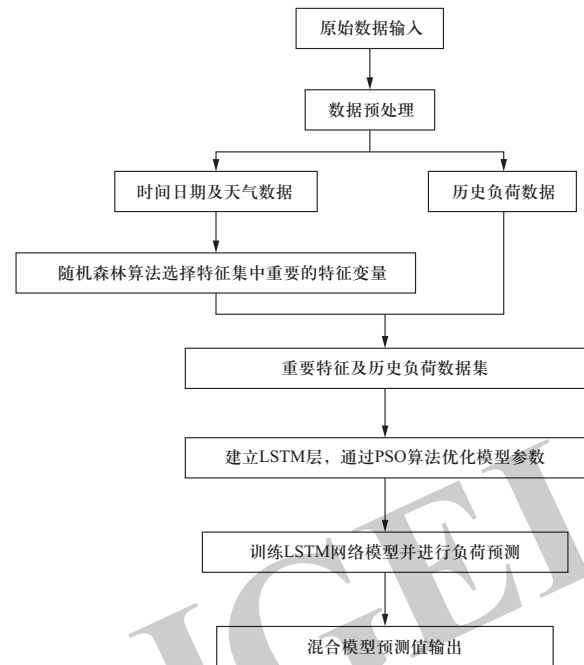


图4 RF-LSTM混合模型预测流程图

Fig. 4 Process of prediction of RF-LSTM hybrid model

1) 相关数据的采集和预处理。以河北电网某台区电力负荷数据,以及表1所示的可能对负荷造成影响的环境因素和时间日期因素等共15个特征因素构成数据集。对于部分缺失或异常的负荷数据采用线性插值法进行填补,并对所有特征数据进行归一化处理。

2) 重要特征变量选择。利用随机森林算法对14个时间日期及天气因素特征变量进行重要性排序,选

择最重要的7个特征作为进一步负荷预测的特征变量输入。

3) LSTM网络模型构建及负荷预测。将步骤2)中由随机森林算法筛选得到的7个重要特征变量及历史负荷结合构成8维LSTM网络模型的训练集 $X_L = \{x_1, x_2, \dots, x_7, y\}$ 。构建LSTM网络模型, 将模型参数初始化并通过粒子群算法优化, 训练得到负荷预测结果 $Y_{\text{pred}}$ 。

## 2.4 模型评价指标

根据国家电网负荷预测结果评价指标, 本实验设置平均绝对百分比误差 $E_{\text{MAPE}}$ 、均方根误差 $E_{\text{RMSE}}$ 以及预测精度 $A_F$ 三项评价指标, 具体表达式如式(13)一式(15)所示。

$$E_{\text{MAPE}} = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_{\text{act}}(i) - Y_{\text{pred}}(i)}{Y_{\text{act}}(i)} \right| \quad (13)$$

$$E_{\text{RMSE}} = \sqrt{\frac{\sum_{i=1}^n (Y_{\text{act}}(i) - Y_{\text{pred}}(i))^2}{n}} \quad (14)$$

$$A_F = \left( 1 - \frac{|Y_{\text{act}}(i) - Y_{\text{pred}}(i)|}{Y_{\text{act}}(i)} \right) \times 100\% \quad (15)$$

式中:  $n$ 为总预测次数;  $Y_{\text{act}}(i)$ 和 $Y_{\text{pred}}(i)$ 分别为负荷真实值和负荷预测值。

## 3 算例分析

本文使用河北电网某台区2020年3月至5月的电力负荷数据对所提出的RF-LSTM混合预测模型进行验证。根据如表1所示构建的预测特征, 通过随机森林算法进行重要性筛选, 与历史负荷数据结合作为特征向量, 并按照0.9:0.1的比例将数据集划分为训练集和测试集。在相同条件下分别使用LSTM模型、随机森林模型、BP神经网络模型以及RF-LSTM混合预测模型进行训练及负荷预测。

### 3.1 模型数据归一化

为消除各特征之间不同量纲的影响并使预测模型尽快收敛, 在确定特征集后需要对原始数据进行归一化处理, 使各特征数据经线性变换后都在 $[0, 1]$ 内:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (16)$$

式中:  $x$ 和 $x^*$ 分别为归一化前后的值;  $x_{\max}$ 为各特征样

本数据的最大值;  $x_{\min}$ 为各特征样本数据的最小值。

### 3.2 模型构建及参数确定

电力负荷易受时间、日期及天气环境等因素影响, 且具有非平稳、非线性等特点, 在预测电力负荷的过程中并非参与预测的特征变量越多, 预测越准确。本文通过随机森林算法对影响电力负荷的多个特征变量的重要性进行评估并排序, 选择对负荷影响较大的变量参与后续预测。将表1所示14维特征向量经数据预处理后, 作为重要特征备选, 经过随机森林算法重要性评估后的结果如图5所示。

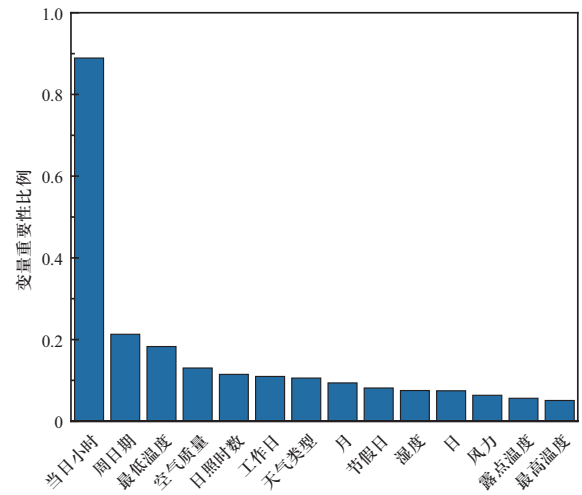


图5 随机森林特征变量重要性分析结果

Fig. 5 Results of the importance of characteristic variables based on random forest

由图5可知, 将对于电力负荷存在影响的14个特征变量按重要性排序, 剔除重要性比例小于0.1的特征变量, 选择当日小时、周日期、最低温度、空气质量、日照时数、工作日、天气类型等变量作为后续负荷预测模型的输入, 使得负荷预测中参与的特征向量由14维降至7维。

将由随机森林算法筛选得到的7维重要特征向量与历史负荷数据结合, 该8维特征向量作为LSTM网络模型的输入, 以1维电力负荷预测结果作为LSTM网络模型的输出, 构建LSTM网络模型。如图3所示, 首先确定LSTM网络模型的结构, 即确定网络隐藏层数和每个隐藏层的记忆单元数, 本文使用穷举搜索法研究二者对LSTM网络模型预测精度的影响。分别设置LSTM网络模型具有1、2、3个网络隐藏层, 并设置每层记忆单元数在 $[40, 80]$ 范围内递增, 计算不同模型结

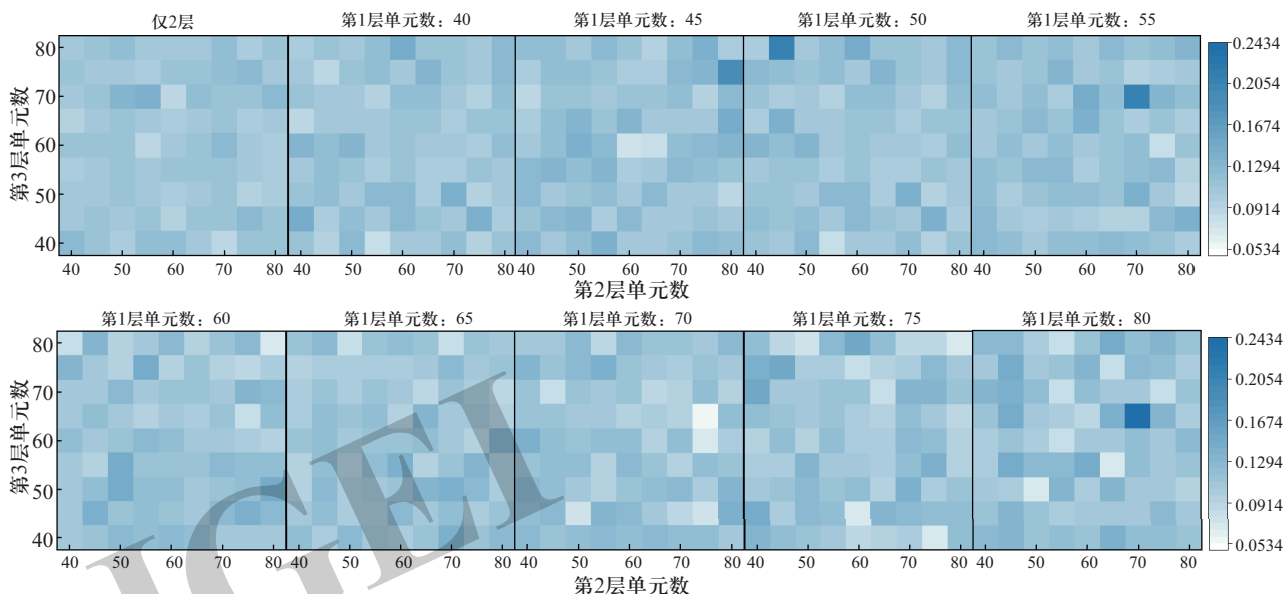


图6 多网络隐藏层不同LSTM网络结构的 $E_{MAPE}$ 比较

Fig. 6 Comparison of  $E_{MAPE}$  of different LSTM network structures of multi-hidden layers

构下负荷预测的平均绝对百分比误差 $E_{MAPE}$ ，得到单网络隐藏层实验结果（表2）和多网络隐藏层实验结果（图6）。

表2 单网络隐藏层不同记忆单元数 $E_{MAPE}$ 比较

Table 2 Comparison of  $E_{MAPE}$  for different number of memory unites of single hidden layer

记忆单元个数	$E_{MAPE}$
40	0.126 2
45	0.102 1
50	0.125 3
55	0.101 8
60	0.110 4
65	0.122 1
70	0.109 8
75	0.115 1
80	0.107 8

综合表2、图6的结果可知，当LSTM具有3个网络隐藏层且各层记忆单元数分别为70、75、65时，LSTM模型的预测精度可以达到最高。

除了模型的结构以外，训练次数的选择也会影响预测精度，过多或过少的训练次数对于LSTM模型的性能都是不利的。训练次数不足会导致模型无法达到

最佳收敛，训练次数过多则会导致训练所耗费的时间过长而预测精度并未得到明显的提升。此外，影响LSTM网络预测精度的主要参数还有初始学习速率（initial learn rate）、学习速率下降时的迭代数（learn rate drop period）及学习速率下降因子（learn rate drop factor），因此还需要确定这3个参数的最佳选择。需要确定的参数较多，不适宜再使用穷举搜索法，因此采用粒子群算法进行参数寻优<sup>[26]</sup>。设粒子群算法中最大迭代次数为100，种群规模为20，适应度函数为不同参数下负荷预测的平均绝对百分比误差 $E_{MAPE}$ ，寻优结果如图7所示。最终确定LSTM网络模型的关键参数如表3所示。

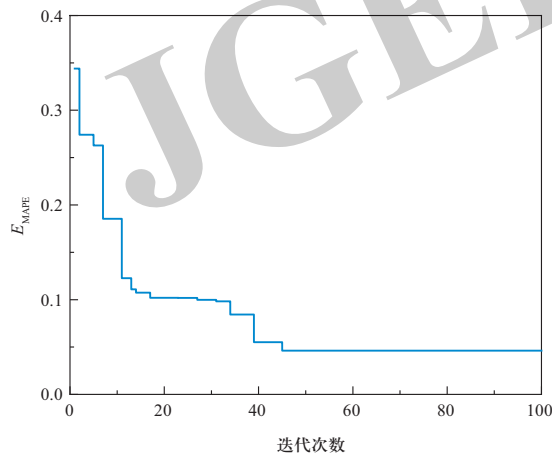


图7 基于粒子群算法的LSTM模型参数寻优

Fig. 7 Parameter optimization of LSTM network based on PSO

表 3 LSTM网络关键参数

Table 3 Key parameters of LSTM network

相关参数	数值
网络隐藏层数	3
LSTM各层单元数	70、75、65
最大迭代数 (max epochs)	250
初始学习速率 (initial learn rate)	0.025
学习速率下降时迭代数 (learn rate drop period)	150
学习速率下降因子 (learn rate drop factor)	0.5

### 3.3 实验结果分析

LSTM网络和随机森林模型构建完毕后, 对本文提出的RF-LSTM模型进行训练及负荷的预测工作, 计算模型的均方根误差 $E_{RMSE}$ 、平均绝对百分比误差 $E_{MAPE}$ 以及预测精度 $A_F$ 三项评价指标, 并将预测结果分别与未经特征量筛选的单一LSTM模型、随机森林算法、BP神经网络模型进行对比。其中, 单一LSTM模型、随机森林算法及BP神经网络模型均使用PSO算法进行参数优化, 各模型参数优化结果分别如表4—表6所示, 表7为各模型预测结果的对比情况, 可看出RF-LSTM混合模型的预测结果明显更优。相比较于传统高维特征量输入的LSTM算法, RF-LSTM混合模型预测结果的均方根误差减少了12.89%, 预测精度提高了0.83个百分点, 这表明通过随机森林算法对高维时间日期及天气等变量进行重要性筛选后模型预测效果更优, 验证了RF-LSTM混合预测模型的可行性。此外, RF-LSTM混合模型预测结果与输入为高维特征量的随机森林算法以及BP神经网络相比, RF-LSTM混合模型预测结果的均方根误差分别减少了19.44%和27.27%, 预测精度分别提高了1.10和2.22个百分点, 验证了本文混合预测模型的有效性。4种模型预测曲线对比如图8所示, 抽取其中连续24 h的预测结果数据汇总表见表8。分析对比结果可知, 本文提出的RF-LSTM混合负荷预测模型对输入变量的重要性进行选择, 大大减少了模型的输入变量个数, 预测误差小, 预测精度高, 能够实现对负荷预测输入参数的优化, 较好地完成电力负荷预测任务。

表 4 单一LSTM模型关键参数

Table 4 Key parameters of single LSTM network

相关参数	数值
网络隐藏层数	3
LSTM各层单元数	45、60、60
最大迭代数 (max epochs)	250

续表

相关参数	数值
初始学习速率 (initial learn rate)	0.025
学习速率下降时迭代数 (learn rate drop period)	125
学习速率下降因子 (learn rate drop factor)	0.7

表 5 单一随机森林模型关键参数

Table 5 Key parameters of single random forest

相关参数	数值
最佳分裂特征数	4
决策树数量	300

表 6 单一BP神经网络关键参数

Table 6 Key parameters of single BP neural network

相关参数	数值
输入层神经元数	15
隐藏层神经元数	8
输出层神经元数	1
隐藏层激活函数	tansig()
输出层激活函数	purelin()
训练次数	500
学习速率	0.05
目标精度	0.000 1
权值及阈值	PSO优化

表 7 不同模型预测结果比较

Table 7 Comparison of prediction results of different models

预测模型	$E_{RMSE}$	$E_{MAPE}$	$A_F/\%$
LSTM模型	0.210 3	0.046 2	95.41
随机森林算法	0.227 4	0.048 1	95.14
BP神经网络模型	0.251 9	0.059 7	94.02
RF-LSTM混合模型	0.183 2	0.040 2	96.24

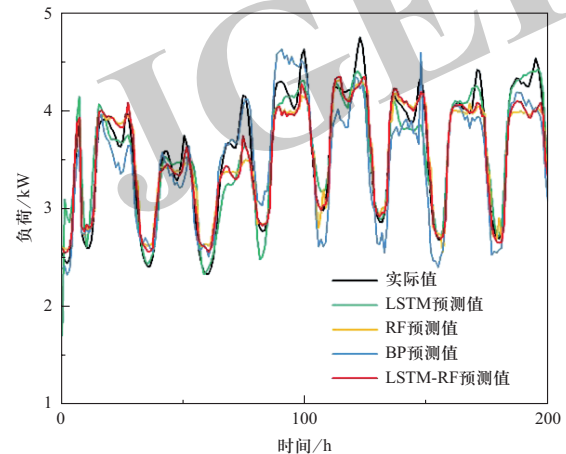


图 8 不同模型预测结果对比图

Fig. 8 Comparison of prediction results of different models



表8 不同模型预测数据汇总  
Table 8 Summary of forecast data for different models

时间/h	负荷真实值/kW	LSTM模型		随机森林算法		BP神经网络模型		RF-LSTM混合预测模型	
		负荷预测值/kW	平均绝对百分比误差/%	负荷预测值/kW	平均绝对百分比误差/%	负荷预测值/kW	平均绝对百分比误差/%	负荷预测值/kW	平均绝对百分比误差/%
1	4.007	3.810	4.91	4.074	1.67	3.764	6.06	4.045	0.95
2	4.266	3.854	9.66	4.119	3.45	3.654	14.35	4.107	3.73
3	4.341	3.844	11.45	4.128	4.91	4.589	5.71	4.182	3.66
4	4.175	3.708	11.19	4.120	1.32	3.587	14.08	4.173	0.05
5	3.865	3.483	9.88	4.050	4.79	3.317	14.18	3.976	2.87
6	3.446	3.284	4.70	3.535	2.58	3.044	11.67	3.504	1.68
7	3.093	3.177	2.72	3.408	10.18	2.558	17.30	3.250	5.08
8	2.847	2.969	4.29	2.803	1.55	2.470	13.24	2.839	0.28
9	2.738	2.792	1.97	2.754	0.58	2.460	10.15	2.746	0.29
10	2.685	2.706	0.78	2.741	2.09	2.400	10.61	2.705	0.74
11	2.677	2.695	0.67	2.733	2.09	2.509	6.28	2.694	0.64
12	2.756	2.800	1.60	2.553	7.37	2.580	6.39	2.716	1.45
13	3.042	3.188	4.80	2.893	4.90	2.976	2.17	2.878	5.39
14	3.585	3.668	2.32	3.641	1.56	3.580	0.14	3.641	1.56
15	3.965	3.956	0.23	3.931	0.86	3.936	0.73	3.944	0.53
16	4.054	4.058	0.10	3.991	1.55	4.021	0.81	4.029	0.62
17	4.059	4.096	0.91	3.998	1.50	4.062	0.07	4.047	0.30
18	4.081	4.075	0.15	3.989	2.25	4.053	0.69	4.048	0.81
19	4.067	4.096	0.71	3.986	1.99	4.022	1.11	4.058	0.22
20	4.021	4.060	0.97	3.988	0.82	3.980	1.02	4.051	0.75
21	4.019	4.076	1.42	3.971	1.19	3.905	2.84	4.021	0.05
22	3.990	4.082	2.31	3.964	0.65	3.942	1.20	3.999	0.23
23	3.969	4.140	4.31	4.070	2.54	3.816	3.85	3.976	0.18
24	4.030	4.221	4.74	3.965	1.61	3.926	2.58	3.975	1.36

#### 4 结语

为了更加精确地预测电力负荷,本文提出一种基于RF-LSTM的混合预测模型,利用随机森林算法对高维特征变量进行重要性排序和选择,将筛选出的重要特征量与历史负荷数据作为LSTM预测模型的输入,构建RF-LSTM混合预测模型。通过对比实验可知RF-LSTM混合模型的主要优势如下。

1) 使用随机森林算法对众多可能影响电力负荷的时间日期及天气因素进行重要性评估,筛选重要变量,从而减少后续预测模型输入变量个数,降低预测

模型的复杂性,有利于提高预测精度。

2) LSTM模型对于具有明显非线性、时序性的电力负荷数据学习能力较强。面对较多超参数寻优问题时使用粒子群算法进行参数优化可以较为快速地获得优化结果,并使预测结果更为准确,避免人为选参对预测结果带来的不利影响。根据实验结果可知,本文提出的方法可有效提高电力负荷的预测精度,RF-LSTM混合预测模型具有充分的可行性和有效性。

本文中所构建的特征数据集未考虑不同时段电价不同这一影响因素,因此后续研究中考虑加入电价等特征,构建更为丰富的特征集并深入探究各因素间的



关系。同时将对模型作进一步改进, 提高预测速度和精度。此外, 还将考虑将电力负荷预测应用于电网数字孪生的建设中。基于负荷预测数据和实时工作环境数据, 研究数据驱动的智能配电网容量挖掘技术, 构建配变及线路重过载多维度评估体系。

## 参考文献

- [1] 姚程文, 杨苹, 刘泽健. 基于CNN-GRU混合神经网络的负荷预测方法[J]. 电网技术, 2020, 44(9): 3416-3424.  
YAO Chengwen, YANG Ping, LIU Zejian. Load forecasting method based on CNN-GRU hybrid neural network[J]. Power System Technology, 2020, 44(9): 3416-3424(in Chinese).
- [2] 张宇航, 邱才明, 贺兴, 等. 一种基于LSTM神经网络的短期用电负荷预测方法[J]. 电力信息与通信技术, 2017, 15(9): 19-25.  
ZHANG Yuhang, QIU Caiming, HE Xing, et al. A short-term load forecasting based on LSTM neural network[J]. Electric Power Information and Communication Technology, 2017, 15(9): 19-25(in Chinese).
- [3] 陆继翔, 张琪培, 杨志宏, 等. 基于CNN-LSTM混合神经网络模型的短期负荷预测方法[J]. 电力系统自动化, 2019, 43(8): 131-137.  
LU Jixiang, ZHANG Qipei, YANG Zhihong, et al. Short-term load forecasting method based on CNN-LSTM hybrid neural network model[J]. Automation of Electric Power Systems, 2019, 43(8): 131-137(in Chinese).
- [4] 李焱, 贾雅君, 李磊, 等. 基于随机森林算法的短期电力负荷预测[J]. 电力系统保护与控制, 2020, 48(21): 117-124.  
LI Yan, JIA Yajun, LI Lei, et al. Short term power load forecasting based on a stochastic forest algorithm[J]. Power System Protection and Control, 2020, 48(21): 117-124(in Chinese).
- [5] 胡海琴, 蒋传文, 蔡宏欣. 中长期负荷预测的傅里叶级数残差修正模型[J]. 安徽电力, 2010(2): 39-41.  
HU Haiqin, JIANG Chuanwen, CAI Hongxin. Mid- and long-term load forecasting using Fourier series residual correction[J]. Anhui Electric Power, 2010(2): 39-41(in Chinese).
- [6] 王宝财. 基于温度近因效应的多元线性回归电力负荷预测[J]. 水电能源科学, 2018, 36(10): 201-205.  
WANG Baocai. Load forecasting of multiple linear regression based on temperature recency effect[J]. Water Resources and Power, 2018, 36(10): 201-205(in Chinese).
- [7] 李琛, 郭文利, 吴进, 等. 基于BP神经网络的北京夏季日最大电力负荷预测方法[J]. 气候与环境研究, 2019, 24(1): 135-142.  
LI Chen, GUO Wenli, WU Jin, et al. A method for prediction of daily maximum electric loads in the summer in Beijing based on the BP neural network[J]. Climatic and Environmental Research, 2019, 24(1): 135-142(in Chinese).
- [8] 王吉权, 王福林, 董志贵, 等. 基于改进BP神经网络的电力负荷预测[J]. 数学的实践与认识, 2017, 47(9): 276-284.  
WANG Jiquan, WANG Fulin, DONG Zhigui, et al. Electrical load forecasting based on improved BP neural network[J]. Mathematics in Practice and Theory, 2017, 47(9): 276-284(in Chinese).
- [9] 鲍伟强, 陈娟, 谢伟, 等. 一种基于自适应神经模糊推理系统的短期负荷预测方法[J]. 电气开关, 2019, 57(1): 8-11.  
BAO Weiqiang, CHEN Juan, XIE Wei, et al. Short-term load forecasting approach using adaptive neuro-fuzzy inference system[J]. Electric Switchgear, 2019, 57(1): 8-11(in Chinese).
- [10] ALI D, YOHANNA M, PUWU M I, et al. Long-term load forecast modelling using a fuzzy logic approach[J]. Pacific Science Review A: Natural Science and Engineering, 2016, 18(2): 123-127.
- [11] 钱志. 基于改进型SVR的电网短期负荷预测[J]. 中国电力, 2016, 49(8): 54-58.  
QIAN Zhi. Short-term power load forecasting based on improved SVR[J]. Electric Power, 2016, 49(8): 54-58(in Chinese).
- [12] CHEN Y B, XU P, CHU Y Y, et al. Short-term electrical load forecasting using the support vector regression (SVR) model to calculate the demand response baseline for office buildings[J]. Applied Energy, 2017, 195: 659-670.
- [13] 戚任远. 上海用电负荷预测中气温累积效应分析[J]. 现代电力, 2018, 35(2): 38-42.  
QI Renyuan. Analysis on temperature accumulation effect in Shanghai power load forecast[J]. Modern Electric Power, 2018, 35(2): 38-42(in Chinese).
- [14] 吴潇雨, 和敬涵, 张沛, 等. 基于灰色投影改进随机森林算法的电力系统短期负荷预测[J]. 电力系统自动化, 2015, 39(12): 50-55.  
WU Xiaoyu, HE Jinghan, ZHANG Pei, et al. Power system short-term load forecasting based on improved random forest with grey relation projection[J]. Automation of Electric Power Systems, 2015, 39(12): 50-55(in Chinese).
- [15] 宋建, 束洪春, 董俊, 等. 基于GM(1, 1)与BP神经网络的综合负荷预测[J]. 电力建设, 2020, 41(5): 75-80.  
SONG Jian, SHU Hongchun, DONG Jun, et al. Comprehensive load forecast based on GM(1, 1) and BP neural network[J]. Electric Power Construction, 2020, 41(5): 75-80(in Chinese).
- [16] 谷云东, 马冬芬, 程红超. 基于相似数据选取和改进梯度提升决策树的电力负荷预测[J]. 电力系统及其自动化学报, 2019, 31(5): 64-69.  
GU Yundong, MA Dongfen, CHENG Hongchao. Power load forecasting based on similar-data selection and improved gradient boosting decision tree[J]. Proceedings of the CSU-EPSA, 2019, 31(5): 64-69(in Chinese).
- [17] 姜建, 刘海琼, 李衡, 等. 基于XGBoost的配电网线路峰值负荷预测方法[J]. 电力系统保护与控制, 2021, 49(16): 119-127.  
JIANG Jian, LIU Haiqiong, LI Heng, et al. Peak load

- forecasting method of distribution network lines based on XGBoost[J]. Power System Protection and Control, 2021, 49(16): 119-127(in Chinese).
- [18] 黄晗, 孙堃, 刘达. 基于随机森林的电力系统小时负荷预测研究[J]. 智慧电力, 2018, 46(5): 8-14.  
HUANG Han, SUN Kun, LIU Da. Hourly load forecasting of power system based on random forest[J]. Smart Power, 2018, 46(5): 8-14(in Chinese).
- [19] 李婉华, 陈宏, 郭昆, 等. 基于随机森林算法的用电负荷预测研究[J]. 计算机工程与应用, 2016, 52(23): 236-243.  
LI Wanhua, CHEN Hong, GUO Kun, et al. Research on electrical load prediction based on random forest algorithm[J]. Computer Engineering and Applications, 2016, 52(23): 236-243(in Chinese).
- [20] YIN L F, SUN Z X, GAO F, et al. Deep forest regression for short-term load forecasting of power systems[J]. IEEE Access, 2020, 8: 49090-49099.
- [21] 陈亮, 王震, 王刚. 深度学习框架下LSTM网络在短期电力负荷预测中的应用[J]. 电力信息与通信技术, 2017, 15(5): 8-11.  
CHEN Liang, WANG Zhen, WANG Gang. Application of LSTM networks in short-term power load forecasting under the deep learning framework[J]. Electric Power Information and Communication Technology, 2017, 15(5): 8-11(in Chinese).
- [22] 李香龙, 马龙飞, 赵向阳, 等. 基于LSTM网络的时间多尺度电采暖负荷预测[J]. 电力系统及其自动化学报, 2021, 33(4): 71-75.  
LI Xianglong, MA Longfei, ZHAO Xiangyang, et al. Multi-time scale electric heating load forecasting based on long short-term memory network[J]. Proceedings of the CSU-EPSA, 2021, 33(4): 71-75(in Chinese).
- [23] 陈纬楠, 胡志坚, 岳菁鹏, 等. 基于长短期记忆网络和LightGBM组合模型的短期负荷预测[J]. 电力系统自动化, 2021, 45(4): 91-97.  
CHEN Weinan, HU Zhijian, YUE Jingpeng, et al. Short-term load prediction based on combined model of long short-term memory network and light gradient boosting machine[J]. Automation of Electric Power Systems, 2021, 45(4): 91-97(in Chinese).
- [24] 庞传军, 张波, 余建明. 基于LSTM循环神经网络的短期电力负荷预测[J]. 电力工程技术, 2021, 40(1): 175-180.  
PANG Chuanjun, ZHANG Bo, YU Jianming. Short-term power load forecasting based on LSTM recurrent neural network[J]. Electric Power Engineering Technology, 2021, 40(1): 175-180(in Chinese).
- [25] YIN R, LI D X, WANG Y F, et al. Forecasting method of monthly wind power generation based on climate model and long short-term memory neural network[J]. Global Energy Interconnection, 2020, 3(6): 571-576.
- [26] 何邦乐, 黄勇, 叶颀, 等. 基于PSO-LSSVM的高压电力电缆接头温度预测[J]. 电力工程技术, 2019, 38(1): 31-35.  
HE Bangle, HUANG Yong, YE Ting, et al. Temperature prediction of power cable joint based on PSO-LSSVM predict model[J]. Electric Power Engineering Technology, 2019, 38(1): 31-35(in Chinese).

收稿日期: 2021-09-26; 修回日期: 2021-12-16。



董彦军

作者简介:

董彦军(1974), 男, 高级工程师, 研究方向为电网监控运行, E-mail: 291026624@qq.com。

岳凡丁(1997), 女, 硕士研究生, 研究方向为电力系统运行规划与可靠性分析, E-mail: yuefanding@stu.xjtu.edu.cn。

袁欢(1992), 男, 助理教授, 研究方向为电力设备寿命评估与预测, 电力系统可靠性分析。通信作者, E-mail: huanyuan@xjtu.edu.cn。

(责任编辑 李锡)