基于Wide & Deep-XGB2LSTM模型的超短期光伏功率预测

粟 然[1],丁 星[1],孙 帆[1],韩 怡[1],刘会兰[1],严敬汝[2]
(1. 华北电力大学 电气与电子工程学院,河北 保定 071003;
2. 国网河北省电力有限公司电力科学研究院,河北 石家庄 050021)

摘要:为了充分利用电网自身的海量历史数据进行光伏功率预测,提出一种宽度&深度(Wide & Deep)框架下融合极限梯度提升(XGBoost)算法和长短时记忆网络(LSTM)的Wide & Deep-XGB2LSTM超短期光伏功率预测模型。对历史数据进行特征提取,获得时间、辐照度、温度等原始特征,在此基础上进行特征重构,通过交叉组合和挖掘统计特征构造辐照度×辐照度、均值、标准差等组合特征,并通过Filter法和Embedded法进行特征选择。在TensorFlow框架下通过算例对比验证了所提模型及特征工程工作对光伏功率预测性能的提升效果。

关键词:光伏功率预测;宽度&深度模型;极限梯度提升;长短时记忆网络;特征工程;模型融合
中图分类号:TM 615 **文献标志码**:A DOI:10.16081/j.epae.202103020

0 引言

随着光伏装机容量的逐年上升,其大规模并网给电力系统带来的挑战也日趋严峻。准确的光伏功率预测对电网规划和安全稳定运行意义重大[1-3]。

现有光伏功率预测方法主要分为统计法和物理法两大类[4-6]。统计法主要是通过对影响光伏输出的气象因素进行分析从而建立光伏发电预测模型。文献[7]考虑季节类型、天气类型和气象等因素,以综合相似度为指标进行相似日的选取,并以云自适应粒子群优化CAPSO(Cloud Adaptive Particle Swarm Optimization)算法优化Spiking神经网络SNN(Spiking Neural Network)的多突触连接权值,提高算法的收敛精度。文献[8]通过概率神经网络和主成分分析法得到每种天气类型下影响光伏出力的主成分,并基于分散搜索和支持向量机建立光伏发电功率短期预测模型。文献[9-10]分别引入天气类型指数和气溶胶指数,以此提高各种天气类型和突变天气下光伏出力预测的准确度。文献[11]通过估计气溶胶光学厚度和不同积灰状态下光伏电池板的功率输出衰减率,研究雾霾对光伏发电的影响。

上述方法对光伏预测精度有一定提高,但由于气象、云图信息往往覆盖范围较大,无法准确体现光伏电池板所处位置的准确信息,对超短期预测精度的提高有限。考虑到光伏发电系统自身功率历史数据隐含的信息,文献[12]在分析光伏出力混沌特性的基础上,分析光伏出力相空间轨迹局部变化的规律性,建立基于混沌-径向基函数RBF(Radial Basis Function)神经网络的光伏发电功率超短期预测模型。文献[13]采用双维度顺序填补方法补齐缺失数据,基于完整数据建立改进Kohonen天气聚类模型,并利用S-Kohonen实现预测日天气类型识别,采用多种群果蝇优化广义回归神经网络(MFOA-GRNN)模型对预测日光伏发电功率进行预测。

上述文献大多是围绕气象条件和历史数据展开建模,并应用不同算法进行预测,而较少涉及运用深度学习挖掘数据深层信息及特征工程,选取特征时大多以显性物理含义为导向,未能尝试挖掘特征数据间可能隐含的其他关系。

本文利用深度学习算法在数据挖掘方面的优势,在宽度&深度(Wide & Deep)框架下,融合极限梯度提升XGBoost(eXtreme Gradient Boosting)算法和长短期记忆网络LSTM(Long Short-Term Memory network)构建用于超短期光伏功率预测的Wide & Deep-XGB2LSTM模型。同时,深入挖掘常规预测中应用的辐照度、温度等特征及其之间蕴含的深层信息,构造大量特征并通过Filter法和Embedded法进行特征选择作为模型的输入。所建模型融合了上述3种算法的优点,特征工程能够较好地挖掘用于光伏功率预测的历史数据蕴含的信息。最后,通过算例对比不同算法和融合方式间的性能差异,验证了本文模型和特征工程对预测性能的提升效果。

1 Wide & Deep模型及模型融合

1.1 Wide & Deep模型

Wide & Deep模型是谷歌在2016年提出的用于推荐系统的深度学习方法,其通过Wide和Deep两部分进行联合训练得到最终结果[14]。

Wide部分通常指的是一种广义线性模型,它通过输入数据的原始特征和cross-product transformation生成的组合特征来体现模型的记忆能力。cross-

收稿日期:2020-06-03;**修回日期**:2021-01-13
基金项目:中央高校基本科研业务费专项资金资助项目(2017MS093)
Project supported by the Fundamental Research Funds for the Central Universities(2017MS093)

product transformation的定义为：

$$\phi_k(x_i)=\prod_{i=1}^{d}x_i^{c_{ki}}\quad c_{ki}\in\{0,1\} \tag{1}$$

其中，$\phi_k(x_i)$为第k个组合特征；x_i为输入的第i维特征；d为输入的维度；c_{ki}表示第i维特征是否参与第k个组合特征的构造，其值为0表示不参与构造，为1表示参与构造。

该部分通过线性模型+特征交叉的方式实现了结果的记忆能力和可解释性，但无法通过学习得到训练集中未出现过的特征，即泛化能力较弱。

Deep部分是一个多层感知器，其输入通常为连续特征和Embedding后的离散特征，通过深度神经网络DNN（Deep Neural Network）等方法学习训练集中没有出现过的特征组合，以实现模型的泛化能力，但可能会由于输入过于稀疏而导致结果过度泛化。

Wide & Deep模型平衡了Wide部分的记忆能力和Deep部分的泛化能力，适用于多维特征的回归和分类问题，其结构见附录中图A1。

1.2 XGBoost算法

XGBoost算法是文献[15]在梯度提升决策树GBDT（Gradient Boosting Decision Tree）的基础上提出的一种改进算法，广泛应用于Kaggle等机器学习竞赛中解决分类和预测问题。

传统GBDT基于boosting框架，通过迭代的方式产生若干弱分类器，每个分类器依赖上一个分类器产生的残差（即预测值与真实值之间的误差）来调整权重，最终通过对所有弱分类器的线性组合得到预测结果。由于训练过程是通过不断降低偏差来提高预测精度，因此要求弱分类器足够简单，具有“高偏差、低方差”的特点，以使得整体结果趋于“低偏差、低方差”。

由于GBDT的计算依赖前一次训练的结果，当数据规模变大时效率相应降低。为了改善该情况，XGBoost通过对目标函数进行泰勒展开，使得目标函数只依赖于每个数据点在误差函数上的一阶导数和二阶导数，并且加入正则项来控制模型的复杂程度，防止过拟合并提高泛化能力。XGBoost的目标函数为：

$$f_{\text{obj}}^{(t)}=\sum_{i=1}^{n}l\left(y_i,\hat{y}_i^{t-1}+f_t(x_i)\right)+\omega(f_t)+f_{\text{cons}} \tag{2}$$

其中，n为样本总数；y_i为第i个样本的实际值；$l\left(y_i,\hat{y}_i^{t-1}+f_t(x_i)\right)$为误差函数，$\hat{y}_i^{t-1}+f_t(x_i)$为第$i$个样本第$t$轮预测值$\hat{y}_i^t$，$f_t(x_i)$是为使目标函数值尽可能降低而加入的函数，与误差函数相关；$\omega(f_t)$为正则项；f_{cons}为常数项。

用标准二阶泰勒展开式对上述目标函数进行泰勒展开，标准二阶泰勒展开式表示为：

$$f(x+\Delta x)\approx f(x)+f'(x)\Delta x+\frac{1}{2}f''(x)(\Delta x)^2 \tag{3}$$

其中，f、x、Δx分别为函数、变量、变量的变化量。

同时定义$l\left(y_i,\hat{y}_i^{t-1}\right)$的一阶偏导和二阶偏导分别为$g_i$、$h_i$：

$$\begin{aligned}g_i&=\partial_{\hat{y}_i^{t-1}}l(y_i,\hat{y}_i^{t-1})\\h_i&=\partial^2_{(\hat{y}_i^{t-1})^2}l(y_i,\hat{y}_i^{t-1})\end{aligned} \tag{4}$$

由此得到目标函数近似取值为：

$$f_{\text{obj}}^{(t)}\approx\sum_{i=1}^{n}\left(l\left(y_i,\hat{y}_i^{t-1}\right)+g_if_t(x_i)+\frac{1}{2}h_if_t^2(x_i)\right)+\omega(f_t)+f_{\text{cons}} \tag{5}$$

如前所述，决策树中第t轮回归树由第$t-1$轮回归树的残差得到，这意味着第t轮时第$t-1$棵树的$\hat{y}_i^{t-1}$是已知的，而y_i作为实际值也是已知的，则此时的误差函数$l\left(y_i,\hat{y}_i^{t-1}\right)$对上述目标函数的优化就不再产生影响，可以直接去掉，同时移除常数项，得到一个统一的目标函数表达式为：

$$f_{\text{obj}}^{(t)}=\sum_{i=1}^{n}\left(g_if_t(x_i)+\frac{1}{2}h_if_t^2(x_i)\right)+\omega(f_t) \tag{6}$$

由此可以得出，此时XGBoost的目标函数只依赖于每个数据点在误差函数上的一阶导数g_i和二阶导数h_i，相较于GBDT，整体计算速度得到了提升。详细的公式推导过程见文献[15]。

1.3 LSTM

LSTM是为了解决传统循环神经网络RNN（Recurrent Neural Network）存在的长依赖问题而提出的，其避免了RNN在解决长序列问题时可能出现的梯度爆炸或梯度弥散问题[16]。对于单个tanh层，标准RNN结构见附录中图A2。它通过不断地将信息进行循环操作来保证信息持续存在，从而利用之前的信息辅助当前任务。但当任务复杂，之前的相关信息与当前的预测位置之间间隔很大时，RNN将会丧失学习如此远距离信息的能力。为了弥补该缺陷，LSTM在类似的链式结构上加入细胞状态以及对细胞状态进行控制的遗忘门、输入门、输出门，其结构如附录中图A3所示。

该结构中，对细胞状态$\boldsymbol{C}_{t-1}$的信息保留或丢弃量由遗忘门$\boldsymbol{f}_t$通过查看输入门$\boldsymbol{h}_{t-1}$和$\boldsymbol{x}_t$的信息来决定，而给$\boldsymbol{C}_{t-1}$更新的信息则由输入门$\boldsymbol{i}_t$、$\boldsymbol{h}_{t-1}$、$\boldsymbol{x}_t$通过tanh层得到的候选细胞状态$\bar{\boldsymbol{C}}_t$决定，更新完细胞状态后RNN单元的最终输出由输出门$\boldsymbol{o}_t$根据输入门$\boldsymbol{h}_{t-1}$和$\boldsymbol{x}_t$进行判断。该过程计算公式为：

$$\begin{cases}\boldsymbol{f}_t=\sigma(\boldsymbol{W}_{\text{f}}[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_{\text{f}})\\\boldsymbol{i}_t=\sigma(\boldsymbol{W}_{\text{i}}[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_{\text{i}})\\\bar{\boldsymbol{C}}_t=\tanh(\boldsymbol{W}_{\text{C}}[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_{\text{C}})\\\boldsymbol{C}_t=\boldsymbol{f}_t\boldsymbol{C}_{t-1}+\boldsymbol{i}_t\bar{\boldsymbol{C}}_t\\\boldsymbol{o}_t=\sigma(\boldsymbol{W}_{\text{o}}[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_{\text{o}})\\\boldsymbol{h}_t=\boldsymbol{o}_t\tanh\boldsymbol{C}_t\end{cases} \tag{7}$$

其中，$\boldsymbol{W}_f$、$\boldsymbol{W}_i$、$\boldsymbol{W}_C$、$\boldsymbol{W}_o$ 和 $\boldsymbol{b}_f$、$\boldsymbol{b}_i$、$\boldsymbol{b}_C$、$\boldsymbol{b}_o$ 分别为各状态的权重矩阵和偏置矩阵；σ 为 sigmoid 激活函数。

1.4　BP 神经网络

反向传播 BP 神经网络作为传统的神经网络，其最大的特点是对误差的反向传播。该方法通过将每次训练得到的结果与预期结果进行误差分析，修改权值和阈值，使模型输出逐渐逼近预期结果。该方法目前已经发展十分成熟，本文不再赘述。

1.5　模型融合

Wide & Deep 模型本质上解决的是推荐系统的点击率预测问题，与本文研究的光伏功率预测问题存在一定差异。XGBoost 和 LSTM、BP 神经网络在解决预测问题上均表现优良，特别是 LSTM 在解决时间序列的问题上具有突出优势。

模型融合是在训练多个模型后按照一定策略将这些模型集成为一个模型的方法，该方法能够提高模型的准确率，并且不同模型结果之间的相关性越低，融合模型越能集成不同模型的优势，模型融合的结果也会越好。常见的模型融合方式包括投票法、加权法、学习法等。本文参考文献[17]的命名方法，选取基于树模型的 XGBoost 和基于神经网络的 LSTM，提出一种 Wide & Deep 框架下嵌入 XGBoost 和 LSTM 的模型融合方法，称为 Wide & Deep-XGB2LSTM。该方法先训练 XGBoost 得到初步预测结果，再将初步预测结果和特征一起作为 LSTM 的输入特征进行预测，以发挥 LSTM 在序列问题上的优势，提高预测精度，最后将 Wide 部分与经过全连接层的 LSTM 部分进行拼接，其结构如图 1 所示。

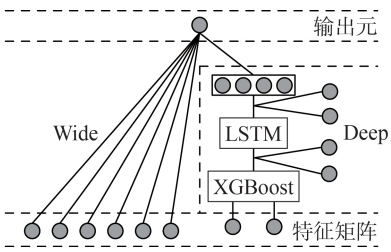

图 1　Wide & Deep-XGB2LSTM 模型结构

Fig.1　Structure of Wide & Deep-XGB2LSTM model

2　光伏功率预测模型建模

2.1　数据分析与预处理

数据作为算法的根本，直接影响模型和算法的性能，因此，在进行训练前对数据进行前期探查和预处理至关重要。

（1）缺失数据和异常数据处理。查看数据基本信息，若存在缺失数据则进行填充，再根据箱型图的离散数据判定原则[18]对异常数据进行筛选和替换。替换可以使用平均值、前值、特殊值等，考虑到一般情况下气象数据具有短时间内不会出现突变的特点，采用前值对异常数据进行替换。

（2）数据无量纲化。机器学习算法中，将不同规格的数据无量纲化可以加快算法求解速度。本文对 Wide & Deep 部分数据进行标准化处理，对 LSTM 部分数据进行归一化处理，XGBoost 属于决策树，不需要进行无量纲化处理。

（3）类别型数据编码。对于反映样本特征的类别型数据，需要将其通过独热编码（one-hot encoding）变成计算机能够识别的形式，编码后的结果通常为一组二元稀疏矩阵。本文中涉及的需要进行编码的特征主要为时间特征和风向特征，其中时间特征需要对标准格式的时间信息进行切片，提取样本记录的年、月、日、时、分属性。

2.2　特征工程

特征工程是从原始数据中提取对模型和算法有用特征的过程，对提升模型性能、充分发挥算法的作用至关重要。该过程主要包括特征提取、特征重构和特征选择 3 个过程，针对本文研究的光伏功率预测问题，具体过程如下。

（1）特征提取。考虑到光伏发电功率主要受自然因素影响，将温度、湿度、辐照度、风向、风速、压强作为原始特征，并将编码得到的时间特征也加入其中。

（2）特征重构。考虑到上述特征与光伏发电功率之间可能存在更深层的联系，对这些特征进行计算或组合得到新特征，常见的方法有基于多项式的变换、基于指数函数的变换等。本文对温度、湿度、辐照度、风向、风速、压强 6 个特征进行交叉组合，得到辐照度×辐照度、湿度×辐照度等 51 个新的组合特征。同时，提取上述 6 个特征每日的最大值、最小值、平均值和标准差作为 24 个统计特征，全部特征组合见附录中表 A1。

（3）特征选择。当特征维度较高时，需要通过特征选择选取对模型有帮助的特征，减轻计算压力。本文首先采用 Filter 法中的方差过滤法筛选特征，删除方差为 0 的特征。由于所选数据均为 2019 年的历史数据，上述所选特征中年份特征方差为 0，因此予以剔除。然后采用 Embedded 法对特征进行选择，该方法先通过常规机器学习算法进行训练，得到各个特征的权重系数，再根据系数大小对特征进行选择。根据 XGBoost 可以输出特征对预测结果重要性的特点，本文利用 XGBoost 进行训练，得到上述 85 个特征的权重排序，将权重排序过低的最后 15 个特征予以剔除。特征工程流程如图 2 所示。

2.3　Wide & Deep-XGB2LSTM 光伏功率预测模型

2.3.1　模型评估指标

预测问题中常用平均绝对误差（MAE）、均方误

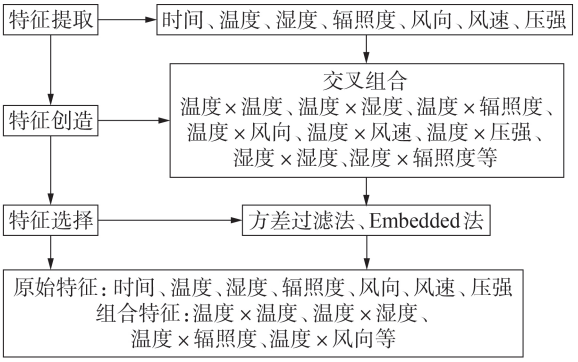

图2 特征工程流程图

Fig.2 Flowchart of feature engineering

差(MSE)、均方根误差(RMSE)、平均绝对百分误差(MAPE)等指标来评价预测结果。为了直观地展示和比较预测结果，本文选取RMSE、MAPE和MAE作为评估指标，具体计算方式见附录中式(A1)—(A3)。

2.3.2 模型参数设置

本文所提模型涉及Wide & Deep模型、XGBoost模型、LSTM模型和传统机器学习算法BP神经网络。其中，Wide & Deep模型设置3层隐藏层，各层神经元数分别为1 024、512、256，激活函数为ReLU，Embedding层更换为全连接层，目标函数为随机梯度下降SGD(Stochastic Gradient Descent)，损失函数为MSE，迭代次数为100。

同时，为了对比不同算法的预测性能，XGBoost和LSTM也均设置迭代次数为100。由于本文研究的是预测问题，XGBoost中booster选择gbtree，目标函数为reg:liner。参数调优利用XGBoost提供的cv函数和sklearn的GridSearchCV进行自动调参，学习率设置为0.2，采用5折交叉验证，早期停止次数设置为100，分别对最大树深度等参数进行测试，得到弱学习器数量为100，最大树深度为6，最小叶子节点样本权重为5，节点分裂所需的最小损失函数下降值为0.3，每棵树随机采样比例为0.8，每棵树随机采样列数占比为0.7，其他参数采用算法默认值。LSTM设置为3层(类似Wide & Deep)，神经元数目分别设置为256、128、64，优化器基于Adam算法，为防止过拟合，Dropout设置为0.2。类似地，BP神经网络采用3层隐藏层，迭代次数为100，神经元数分别为256、128、64，学习率为0.2，损失函数为MSE，Adam梯度下降方式优化损失函数，激活函数为ReLU。批大小均设置为256。

2.3.3 Wide & Deep-XGB2LSTM模型

将经过预处理和特征工程的数据集划分为训练集、验证集和测试集，由训练集和验证集进行模型训练和参数优化，将得到的最优模型对测试集进行测试并得到预测结果，模型具体搭建流程如图3所示。

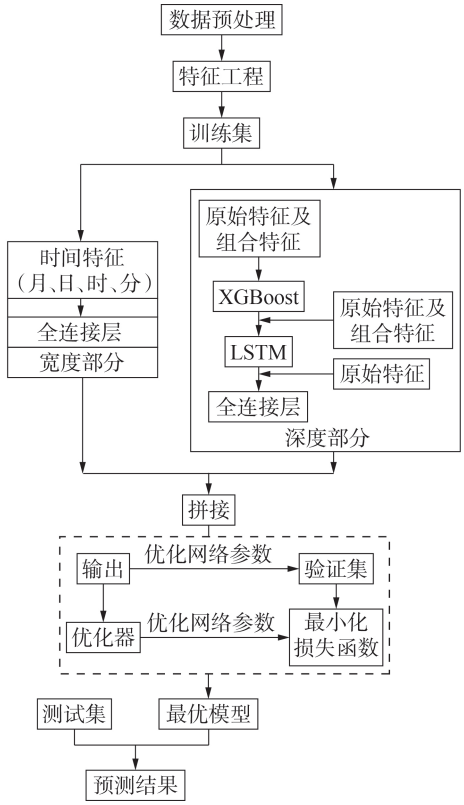

图3 Wide & Deep-XGB2LSTM模型搭建流程

Fig.3 Flowchart of building Wide & Deep-XGB2LSTM model

3 算例分析

选取某地光伏电站2019年4月1日至9月1日的历史数据进行分析，采样时间为每天05:00—20:00，采样分辨率为15 min，每天采集60个样本点。

为了直观地体现本文模型和特征工程对预测结果的提升效果，分别使用单独的Wide & Deep模型、XGBoost模型、LSTM模型、BP神经网络以及并行融合Wide & Deep、XGBoost和LSTM的模型(称为Wide & Deep+XGBoost+LSTM，结构见附录中图A4)对光伏发电功率进行预测。

将4月1日至8月30日的数据按80%和20%的比例划分为训练集和验证集，8月31日至9月2日连续三天作为待预测日，采用单步预测的方式，XGBoost以预测时间点前6 h的数据作参考得到初步预测结果，并将结果作为输入传递到LSTM。LSTM以该结果和特征选择得到的特征作为输入，以预测时间点前4 h时的数据对待预测时间点进行预测。

按照2.3.2节设置的参数，所提模型迭代过程中最优MSE和平均MSE变化趋势如图4所示。由图中可见，模型可以较好地收敛。

在此基础上，上述几种算法的预测结果对比曲线见附录中图A5。误差对比见表1。

其中，Wide & Deep模型将时间特征作为Wide

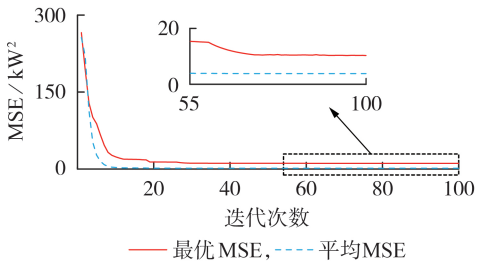

图 4　Wide & Deep-XGB2LSTM 迭代过程 MSE 变化

Fig.4　Change of MSE during iteration process of Wide & Deep-XGB2LSTM

表 1　不同算法误差对比

Table 1　Comparison of error among different algorithms

时间	算法	RMSE / kW	MAPE / %	MAE / kW
8 月 31 日 05:00—20:00	1	7.01	2.18	5.05
	2	28.84	10.57	22.49
	3	36.09	11.27	25.25
	4	18.25	14.65	15.30
	5	17.63	7.98	13.89
	6	32.87	8.81	23.01
9 月 1 日 05:00—20:00	1	6.06	5.35	5.04
	2	21.04	18.31	17.00
	3	29.31	17.04	23.02
	4	17.16	12.43	13.73
	5	9.34	12.39	7.93
	6	25.38	11.64	17.73
9 月 2 日 05:00—20:00	1	8.60	9.50	7.55
	2	23.28	13.29	18.18
	3	28.14	9.98	18.13
	4	16.92	19.92	13.86
	5	13.42	6.14	9.91
	6	25.73	27.54	20.89
8 月 31 日—9 月 2 日 平均误差	1	7.22	5.68	5.88
	2	24.39	14.06	19.22
	3	31.18	12.76	22.13
	4	17.44	15.67	14.30
	5	13.46	8.84	10.58
	6	27.99	16.00	20.54

注：算法 1—6 分别表示 Wide & Deep-XGB2LSTM、XGBoost、Wide & Deep、LSTM、Wide & Deep+XGBoost+LSTM、BP，后同。

部分的输入，将由特征重构得到的特征和 6 个原始特征（除时间特征）作为 Deep 部分输入。XGBoost、LSTM、BP 神经网络将特征选择得到的全部特征作为输入。Wide & Deep+XGBoost+LSTM 中的 3 种算法预测结果并行拼接时占比分别为 0.2、0.4、0.4。

需要指出的是，尽管本文采用固定的特征组合对非线性的光伏出力进行预测，但由于特征组合数量较多，所包含的数据信息量巨大，因此本文认为经过这样大量数据训练的模型可以弥补采用固定特征组合预测非线性光伏出力带来的不足。同时，大量的数据训练有助于挖掘与光伏出力显性相关性不明显的因素和光伏出力间的深层关系。

由上述结果可以发现，单独使用 XGBoost、LSTM、BP 算法时，平均 RMSE 均低于主要用于推荐系统的 Wide & Deep 模型，四者平均 MAPE 指标相差不大。当采用模型融合的方式将 XGBoost 和 LSTM 的结果按一定方式嵌入 Wide & Deep 模型再进行预测时，无论是串行融合的 Wide & Deep-XGB2LSTM 还是并行融合的 Wide & Deep+XGBoost+LSTM，预测结果的评价指标相比上述单独算法均有不同程度改善。同时，对本文所研究的光伏功率预测问题而言，XGBoost 和 LSTM 串行连接后嵌入 Wide & Deep 模型的预测效果要优于 XGBoost、LSTM、Wide & Deep 三者并行拼接。

从运算时间上看，XGBoost、Wide & Deep、LSTM 和 BP 神经网络单独训练的平均时间分别为 133.34、211.76、226.44、238.12 s，Wide & Deep-XGB2LSTM 和 Wide & Deep+XGBoost+LSTM 分别为 432.21、325.67 s。单步预测时间相差不大，前 4 种单独算法的预测时间均在 4 s 左右，后 2 种融合算法预测时间约为 6 s（一般而言，计算机性能越好则所需时间越短）。可见，尽管训练网络时模型融合方式会导致所需时间长于单一算法，但在预测阶段时间差异不大，且模型融合对预测精度的提高显著，本文认为牺牲部分时间来提高预测精度是可行的。

利用 XGBoost 可以输出特征对预测结果重要程度影响的性质，可以得到 Wide & Deep-XGB2LSTM 在训练集上重要程度排序前十的特征如图 5 所示。

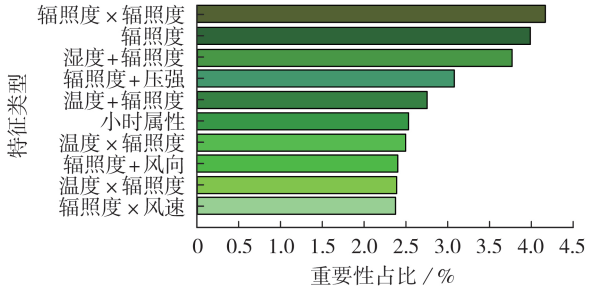

图 5　重要程度排序前十的特征及其占比

Fig.5　Top ten important features and their proportions

图 5 表明，在原始特征的基础上进行合适的特征重构有利于挖掘数据深层信息，为网络训练过程提供更多信息参考。对于本文所研究的光伏功率预测问题，除小时属性特征外，辐照度及由其交叉组合得到的湿度+辐照度等相关特征对发电功率影响较大。

另外，为了更清晰地对比不同特征工程对预测结果的影响，本文仅使用原始特征和统计特征作为上述 6 种模型的输入对待预测日进行预测，预测结果见附录中图 A6，得到的平均评价指标对比结果如表 2 所示。

可以看出，上述所有算法仅采用原始特征和统计特征进行预测的误差均高于进行交叉组合后再预测的误差，说明交叉组合有利于挖掘原始特征与光伏功率间的深层隐含关系，合理的特征工程有利于

表2 不考虑交叉组合特征时的平均评价指标对比

Table 2 Comparison of average evaluation indexes without considering cross-combination features

算法	8月31日至9月2日平均误差		
	RMSE / kW	MAPE / %	MAE / kW
1	17.56	9.14	12.75
2	42.34	15.74	28.67
3	49.65	30.72	31.03
4	33.85	15.07	22.04
5	28.03	14.04	33.68
6	42.91	14.62	29.55

提高预测精度。另外,与采用交叉组合后预测的情况类似,仅用于推荐系统的Wide&Deep模型单独使用效果较差。

4 结论

为了充分发挥电网历史数据在光伏功率预测问题中的作用,本文提出一种基于Wide & Deep-XGB2LSTM模型的超短期光伏功率预测方法,并通过大量特征工程深入挖掘历史数据蕴含的信息。本文主要得到如下结论:

(1)所提模型通过训练大量历史数据可得到较准确的预测结果,充分利用了电网资源,降低了光伏功率预测工作对外部其他信息采集工作的依赖性;

(2)特征工程作为前期数据处理工作至关重要,充分、合理的特征工程能够深入挖掘数据的深层信息,较好地提升模型性能、发挥算法优势;

(3)多种算法的融合有利于提高预测精度,发挥算法各自优势,实际预测工作中可以多加尝试,寻找解决问题的最优融合方式。

需要指出的是,本文研究内容侧重于利用电网自身的历史数据进行预测,未考虑气溶胶、雾霾等需要借助外部监测手段的影响因素,后续研究将结合电网自身数据与其他气象监测数据,利用深度学习进行数据挖掘,提高预测精度。

附录见本刊网络版(http://www.epae.cn)。

参考文献:

[1] 舒印彪,张智刚,郭剑波,等. 新能源消纳关键因素分析及解决措施研究[J]. 中国电机工程学报,2017,37(1):1-9.
SHU Yinbiao,ZHANG Zhigang,GUO Jianbo,et al. Study on key factors and solution of renewable energy accommodation [J]. Proceedings of the CSEE,2017,37(1):1-9.

[2] 赵书强,胡利宁,田捷夫,等. 基于中长期风电光伏预测的多能源电力系统合约电量分解模型[J]. 电力自动化设备,2019,39(11):13-19.
ZHAO Shuqiang,HU Lining,TIAN Jiefu,et al. Contract power decomposition model of multi-energy power system based on mid-long term wind power and photovoltaic electricity forecasting[J]. Electric Power Automation Equipment,2019,39(11):13-19.

[3] 鲁宗相,黄瀚,单葆国,等. 高比例可再生能源电力系统结构形态演化及电力预测展望[J]. 电力系统自动化,2017,41(9):12-18.
LU Zongxiang,HUANG Han,SHAN Baoguo,et al. Morphological evolution model and power forecasting prospect of future electric power systems with high proportion of renewable energy[J]. Automation of Electric Power Systems,2017,41(9):12-18.

[4] KOSTER D,MINETTE F,BRAUN C,et al. Short-term and regionalized photovoltaic power forecasting,enhanced by reference systems,on the example of Luxembourg[J]. Renewable Energy,2019,132:455-470.

[5] 谭津,邓长虹,杨威,等. 微电网光伏发电的Adaboost天气聚类超短期预测方法[J]. 电力系统自动化,2017,41(21):33-39.
TAN Jin,DENG Changhong,YANG Wei,et al. Ultra-short-term photovoltaic power forecasting in microgrid based on Adaboost clustering[J]. Automation of Electric Power Systems,2017,41(21):33-39.

[6] 单英浩,付青,耿炫,等. 基于改进BP-SVM-ELM与粒子化SOM-LSF的微电网光伏发电组合预测方法[J]. 中国电机工程学报,2016,36(12):3334-3343.
SHAN Yinghao,FU Qing,GENG Xuan,et al. Combined forecasting of photovoltaic power generation in microgrid based on the improved BP-SVM-ELM and SOM-LSF with particlization[J]. Proceedings of the CSEE,2016,36(12):3334-3343.

[7] 陈通,孙国强,卫志农,等. 基于相似日和CAPSO-SNN的光伏发电功率预测[J]. 电力自动化设备,2017,37(3):66-71.
CHEN Tong,SUN Guoqiang,WEI Zhinong,et al. Photovoltaic power generation forecasting based on similar day and CAPSO-SNN[J]. Electric Power Automation Equipment,2017,37(3):66-71.

[8] 王昕,黄柯,郑益慧,等. 基于PNN/PCA/SS-SVR的光伏发电功率短期预测方法[J]. 电力系统自动化,2016,40(17):156-162.
WANG Xin,HUANG Ke,ZHENG Yihui,et al. Short-term forecasting method of photovoltaic output power based on PNN/PCA/SS-SVR[J]. Automation of Electric Power Systems,2016,40(17):156-162.

[9] 袁晓玲,施俊华,徐杰彦. 计及天气类型指数的光伏发电短期出力预测[J]. 中国电机工程学报,2013,33(34):57-64,12.
YUAN Xiaoling,SHI Junhua,XU Jieyan. Short-term power forecasting for photovoltaic generation considering weather type index[J]. Proceedings of the CSEE,2013,33(34):57-64,12.

[10] LIU J,FANG W L,ZHANG X D,et al. An improved photovoltaic power forecasting model with the assistance of aerosol index data[J]. IEEE Transactions on Sustainable Energy,2015,6(2):434-442.

[11] 刘卫亮,刘长良,林永君,等. 计及雾霾影响因素的光伏发电超短期功率预测[J]. 中国电机工程学报,2018,38(14):4086-4095,4315.
LIU Weiliang,LIU Changliang,LIN Yongjun,et al. Super short-term photovoltaic power forecasting considering influence factor of smog[J]. Proceedings of the CSEE,2018,38(14):4086-4095,4315.

[12] 王育飞,付玉超,孙路,等. 基于混沌-RBF神经网络的光伏发电功率超短期预测模型[J]. 电网技术,2018,42(4):1110-1116.
WANG Yufei,FU Yuchao,SUN Lu,et al. Ultra-short term prediction model of photovoltaic output power based on chaos-RBF neural network[J]. Power System Technology,2018,42(4):1110-1116.

[13] 李燕青,杜莹莹. 基于双维度顺序填补框架与改进Kohonen天气聚类的光伏发电短期预测[J]. 电力自动化设备,2019,39(1):60-65.

LI Yanqing,DU Yingying. Short-term photovoltaic power forecasting based on double-dimensional sequential imputation framework and improved Kohonen clustering[J]. Electric Power Automation Equipment,2019,39(1):60-65.
[14] CHENG H T,LEVENT K,JEREMIAH H. Wide & deep learning for recommender systems[EB/OL]. [2020-03-27]. https://arxiv.org/abs/1606.07792.
[15] CHEN T Q,GUESTRIN C. XGBoost:a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York,USA:ACM,2016:785-794.
[16] 栗然,孙帆,丁星,等. 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/OL]. 电网技术. [2020-05-12]. https://doi.org/10.13335/j.1000-3673.pst.2020.0006a.
[17] LING X L,DENG W W,GU C,et al. Model ensemble for click prediction in Bing search ads[C]//Proceedings of the 26th International Conference on World Wide Web Companion-WWW 17 Companion. New York,USA:ACM,2017:689-698.
[18] 吕海灿,王伟峰,赵兵,等. 基于Wide & Deep-LSTM模型的短期台区负荷预测[J]. 电网技术,2020,44(2):428-436.
LÜ Haican,WANG Weifeng,ZHAO Bing,et al. Short-term substation load forecast based on Wide & Deep-LSTM model[J]. Power System Technology,2020,44(2):428-436.

作者简介:


栗　然

栗　然(1965—),女,河北保定人,教授,博士,主要研究方向为电力系统分析、运行与控制以及新能源发电与并网(**E-mail**:liranlelele@163.com);

丁　星(1995—),女,河北保定人,硕士研究生,主要研究方向为分布式能源规划与调度(**E-mail**:dingxing2014@126.com);

孙　帆(1995—),男,河北保定人,硕士研究生,主要研究方向为综合能源系统负荷预测与运行优化(**E-mail**:15932266113 @163.com)。

(编辑　王锦秀)

Ultra-short-term photovoltaic power prediction based on Wide & Deep-XGB2LSTM model

LI Ran[1],DING Xing[1],SUN Fan[1],HAN Yi[1],LIU Huilan[1],YAN Jingru[2]

(1. School of Electrical and Electronic Engineering,North China Electric Power University,Baoding 071003,China;
2. Electric Power Research Institute of State Grid Hebei Electric Power Company,Shijiazhuang 050021,China)

Abstract:In order to fully use massive historical data of power grid for photovoltaic power prediction,XGBoost (eXtreme Gradient Boosting) algorithm and LSTM (Long Short-Term Memory network) are fused under Wide & Deep framework,and a ultra-short-term photovoltaic power prediction model based on Wide & Deep-XGB2LSTM is proposed. The feature extraction is performed on historical data to obtain primitive features of time,irradiance,temperature and so on,on this basis,feature reconstruction is carried out and combination features such as irradiance×irradiance,mean value,standard deviation are constructed by cross combination and statistical feature mining,and Filter method and Embedded method are used for feature selection. Case comparison under TensorFlow framework verifies the promotion effect of the proposed model and feature engineering work on prediction performance of photovoltaic power.

Key words: photovoltaic power prediction; Wide & Deep model; XGBoost; LSTM; feature engineering; model fusion

附录：

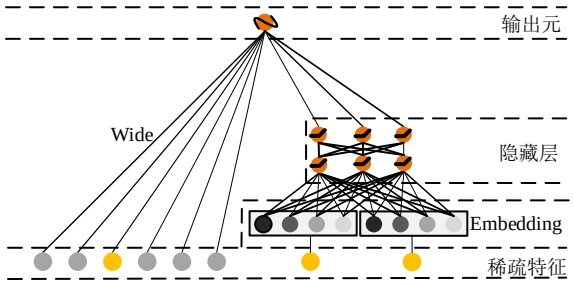

图 A1 Wide & Deep 模型结构

Fig.A1 Structure diagram of Wide & Deep model

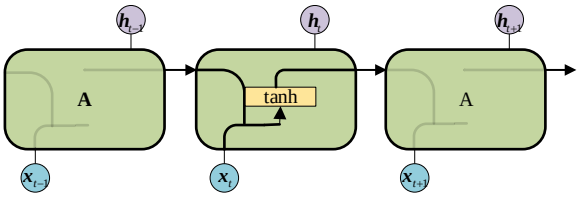

图 A2 标准 RNN 结构

Fig.A2 Structure of standard RNN

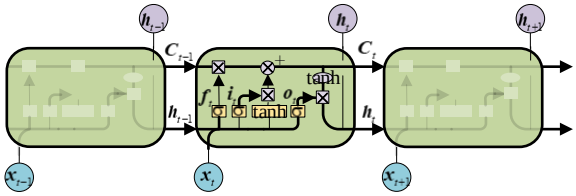

图 A3 LSTM 结构

Fig.A3 Structure of LSTM

表 A1 全部特征及剔除特征

Table A1 All features and removed features

原始特征(11)	统计特征(24)	交叉组合特征(51)	剔除特征(16)
时间(年、月、日、时、分属性)，温度，湿度，辐照度，风向，风速，压强	温度、湿度、辐照度、风向、风速、压强每日的最大最小值、均值和标准差	温度×温度，温度×湿度，温度×辐照度，温度×风向，温度×风速，温度×压强，湿度×湿度，湿度×辐照度，湿度×风向，湿度×风速，湿度×压强，辐照度×辐照度，辐照度×风向，辐照度×风速，辐照度×压强，风向×风向，风向×风速，风向×压强，风速×风速，风速×压强，压强×压强，温度+湿度，温度+辐照度，温度+风向，温度+风速，温度+压强，湿度+辐照度，湿度+风向，湿度+风速，湿度+压强，辐照度+风向，辐照度+风速，辐照度+压强，风向+风速，风向+压强，温度/湿度，温度/辐照度，温度/风向，温度/风速，温度/压强，湿度/辐照度，湿度/风向，湿度/风速，湿度/压强，辐照度/风向，辐照度/风速，辐照度/压强，风向/风速，风向/压强，风速/压强	年属性，风速，风速标准差，天属性，湿度均值，湿度标准差，压强标准差，湿度×湿度，风速×压强，风速均值，风向，湿度×压强，风向/压强，风速×风速，温度×湿度，风向+压强

RMSE、MAPE 和 MAE 分别定义为：

$$\delta_{\mathrm{RMSE}}=\sqrt{\frac{1}{n}\sum_{i=1}^{m}(Y_i-\hat{Y}_i)^2} \tag{A1}$$

$$\delta_{\mathrm{MAPE}}=\frac{1}{n}\sum_{i=1}^{m}\frac{|Y_i-\hat{Y}_i|}{Y_i} \tag{A2}$$

$$\delta_{\mathrm{MAE}}=\frac{1}{n}\sum_{i=1}^{m}|Y_i-\hat{Y}_i| \tag{A3}$$

其中，m 为测试集样本数；Y_i、$\hat{Y}_i$ 分别为第 i 个采样点的发电功率实际观测值和预测值。

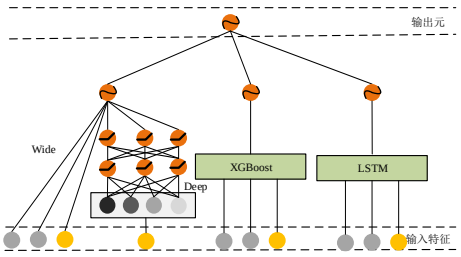

图 A4 Wide & Deep+XGB+LSTM 模型结构

Fig.A4 Structure of Wide & Deep+XGB+LSTM model

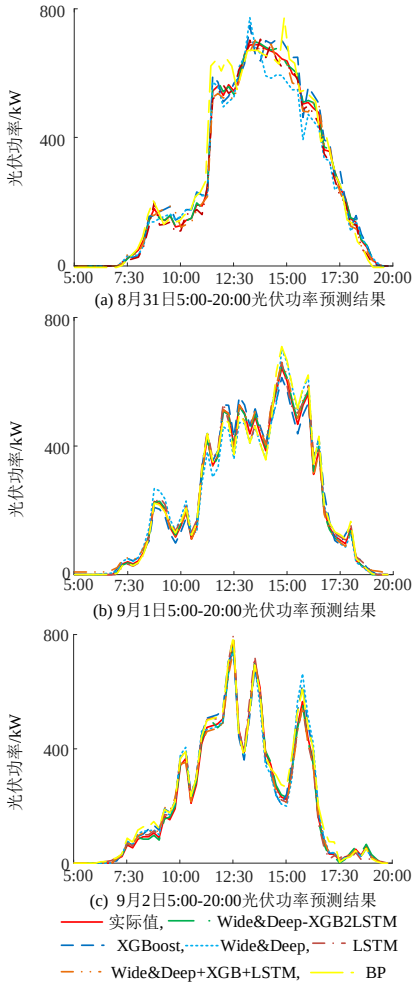

图 A5 不同算法预测结果对比

Fig.A5 Comparison of prediction results among different algorithms

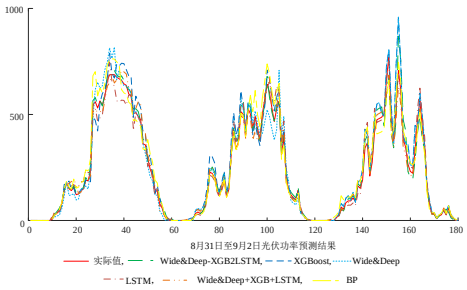

图 A6 不考虑交叉组合特征时不同算法预测结果对比

Fig.A6 Comparison of prediction results among different algorithms without considering cross-combination features