

基于多元信息融合的风电功率预测模型研究

白雪 范仲昊 赵彦龙 王历洋 李彦芝
华能(甘肃)新能源有限公司 甘肃 兰州 730000

摘要: 随着全球对清洁能源需求的持续增长,风电在能源结构中的地位愈发重要。然而,风电的间歇性、波动性和随机性给电力系统的稳定运行与调度带来巨大挑战。本文提出一种基于多元信息融合的风电功率预测模型,详细阐述多元信息融合的方法与策略,构建融合多种信息的预测模型框架。通过实验案例分析,表明该模型相较于传统单一信息预测模型,在预测精度和稳定性方面均有显著提升,为风电功率的准确预测提供了有效途径。

关键词: 风电功率预测; 多元信息融合; 气象数据; 地形数据; 风机运行状态

1 引言

在全球能源转型的大背景下,风能作为一种取之不尽、用之不竭的可再生能源,得到了广泛关注和快速发展。近年来,风电装机容量持续攀升,在能源供应中所占比例不断增加。然而,风电功率的不确定性给电力系统的规划、运行和调度带来了诸多难题。因此,准确的风电功率预测对于保障电力系统的安全稳定运行、提高风电消纳能力以及降低电力系统运行成本具有重要意义。

2 多元信息融合的方法和策略

2.1 信息融合策略

多元信息融合是指将来自多个信息源的信息进行综合处理,以获得比单一信息源更准确、更全面的信息。根据信息融合的层次,可分为数据层融合、特征层融合和决策层融合^[1]。数据层融合直接对原始数据进行融合,能够保留最多的信息,但对数据的质量和同步性要求较高;特征层融合先对各信息源的数据进行特征提取,然后将特征进行融合,在降低数据维度的同时保留了重要的信息;决策层融合则是对各信息源的决策结果进行融合,具有较高的灵活性和容错性。

2.2 信息融合方法

2.2.1 加权平均法

加权平均法是根据不同信息源的重要程度,为每个信息源的预测结果分配一个权重,然后将加权后的预测结果进行求和。权重可以通过专家经验、历史数据分析或优化算法来确定。例如,假设基于气象数据、地形数据和风机运行状态数据的预测结果分别为 P_1 、 P_2 和 P_3 ,对应的权重分别为 w_1 、 w_2 和 w_3 ,则最终的预测结果 P 为:

$$P = w_1 P_1 + w_2 P_2 + w_3 P_3$$

其中, $w_1 + w_2 + w_3 = 1$ 。

2.2.2 贝叶斯融合法

贝叶斯融合法是基于贝叶斯定理,将不同信息源的

预测结果作为先验概率,然后根据新的观测数据进行后验概率的计算和更新。该方法能够充分利用先验知识和新的观测信息,但计算过程较为复杂。例如,设基于气象数据、地形数据和风机运行状态数据的预测结果分别服从不同的概率分布,作为先验概率分布。当有新的观测数据(如实际风电功率)时,根据贝叶斯定理计算后验概率分布,从而得到更准确的预测结果。

2.2.3 神经网络融合法

神经网络融合法是利用神经网络的非线性拟合能力,将不同信息源的数据或特征作为神经网络的输入,通过网络的学习和训练,得到融合后的预测结果。该方法具有较强的自适应能力和泛化能力,但需要大量的训练数据和较长的训练时间^[2]。例如,构建一个多层感知器(MLP)神经网络,将气象特征、地形特征和风机运行状态特征作为输入,风电功率作为输出,通过网络训练得到最优的模型参数,从而实现多元信息的融合和风电功率的预测。

3 基于多元信息融合的风电功率预测模型构建

3.1 模型总体框架

本文构建的基于多元信息融合的风电功率预测模型主要包括数据采集与预处理、特征提取、多元信息融合和预测模型四个部分。首先,采集气象数据、地形数据和风机运行状态数据等多元信息,并进行预处理,包括数据清洗、缺失值填充、异常值处理等。然后,对预处理后的数据进行特征提取,得到能够反映风电功率影响因素的特征向量。接着,采用特征层融合的方法,将不同信息源的特征进行融合。最后,将融合后的特征输入到预测模型中进行风电功率预测。

3.2 数据采集与预处理

3.2.1 数据采集

气象数据可以通过气象站、卫星遥感等方式获取,包括风速、风向、温度、气压、湿度等。地形数据可以

通过地理信息系统 (GIS) 获取, 包括地形高程、坡度、坡向等。风机运行状态数据可以通过风机的监控系统获取, 包括风机的转速、功率、叶片桨距角、偏航角度等。

3.2.2 数据预处理

(1) 数据清洗: 去除重复数据和明显错误的数据。例如, 检查气象数据中的风速是否在合理范围内 (0-30m/s), 若出现负值或超过30m/s的异常值, 则进行删除或修正。

(2) 缺失值填充: 采用插值法、均值法等方法对缺失值进行填充。对于气象数据中的缺失值, 若相邻时刻的数据存在, 可采用线性插值法进行填充; 若缺失值较多, 可采用该气象站历史同期的均值进行填充。对于风机运行状态数据中的缺失值, 可采用该风机历史同期的均值进行填充。

(3) 异常值处理: 采用 3σ 准则、箱线图法等方法检测和异常值。例如, 对于风速数据, 计算其均值 μ 和标准差 σ , 若某个数据点与均值的差的绝对值大于 3σ , 则认为该数据点为异常值, 可将其替换为均值或采用其他合理的处理方法。

(4) 数据归一化: 为了消除不同量纲的影响, 提高模型的收敛速度和预测精度, 对数据进行归一化处理。本文采用最小-最大归一化方法, 将数据映射到 $[0,1]$ 区间, 公式为:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

其中, x 为原始数据, $x_{\min}$ 和 $x_{\max}$ 分别为数据的最小值和最大值, x_{norm} 为归一化后的数据。

3.3 特征提取

3.3.1 气象特征提取

(1) 统计特征: 计算风速的均值、方差、最大值、最小值等统计量, 以反映风速的总体水平和波动情况。例如, 计算过去1小时内的风速均值、方差等。

(2) 时序特征: 提取风速的日变化、季节变化等时序特征。例如, 计算每天不同时段的风速平均值, 分析风速的日变化规律; 计算不同季节的风速平均值, 分析风速的季节变化规律。

(3) 组合特征: 考虑气象因素之间的相关性, 提取一些组合特征, 如风速与温度的比值、风速与气压的差值等。这些组合特征可能更能反映气象条件对风电功率的影响。

3.3.2 地形特征提取

(1) 地形高程特征: 计算地形高程的标准差、坡度的平均值、坡向的分布等特征。例如, 计算风电场区域内地形高程的标准差, 以反映地形的起伏程度; 计算坡度的平均值, 以反映地形的陡峭程度。

(2) 空间特征: 考虑地形在空间上的分布情况, 提

取一些空间特征, 如地形高程的梯度、地形粗糙度等。这些空间特征可以通过GIS工具进行计算和分析。

3.3.3 风机运行状态特征提取

(1) 统计特征: 计算风机的转速、功率的均值、方差等统计量, 以反映风机的运行稳定性和发电能力。例如, 计算过去1小时内的风机功率均值、方差等。

(2) 动态特征: 提取叶片桨距角、偏航角度的变化率等动态特征, 以反映风机的调节能力和运行状态。例如, 计算叶片桨距角在单位时间内的变化量, 分析风机的变桨控制情况。

(3) 相关特征: 考虑风机运行状态与气象因素之间的关系, 提取一些相关特征, 如风速与风机功率的比值、风向与偏航角度的偏差等^[3]。这些相关特征可以更全面地反映风电功率的影响因素。

3.4 多元信息融合

本文采用特征层融合的方法, 将气象特征、地形特征和风机运行状态特征进行融合。具体步骤如下:

3.4.1 特征拼接

首先, 将不同信息源的特征向量进行拼接, 得到一个综合的特征向量。例如, 气象特征向量为 X_{weather} , 地形特征向量为 X_{terrain} , 风机运行状态特征向量为 $X_{\text{operation}}$, 则综合特征向量 X 为:

$$X = [X_{\text{weather}}, X_{\text{terrain}}, X_{\text{operation}}]$$

3.4.2 特征降维

采用主成分分析 (PCA) 方法对综合特征向量进行降维处理, 去除冗余信息, 提取主要特征。PCA是一种常用的数据降维方法, 它通过线性变换将原始数据投影到新的坐标系中, 使得投影后的数据方差最大。具体步骤如下:

计算综合特征向量 X 的协方差矩阵 C 。

计算协方差矩阵 C 的特征值和特征向量。

选择前 k 个最大的特征值对应的特征向量, 构成投影矩阵 P 。

将综合特征向量 X 投影到投影矩阵 P 上, 得到降维后的特征向量 X_{reduced} :

$$X_{\text{reduced}} = XP$$

其中, k 的取值可以根据累计方差贡献率来确定, 一般要求累计方差贡献率达到85%以上。

3.5 预测模型

本文选择长短期记忆网络 (LSTM) 作为预测模型。LSTM是一种特殊的循环神经网络 (RNN), 能够有效地解决传统RNN存在的梯度消失和梯度爆炸问题, 适用于处理时间序列数据。LSTM神经元包括输入门、遗忘门和输出门, 通过这些门的控制, 可以实现对历史信息的记忆和

遗忘,从而更好地捕捉时间序列数据中的长期依赖关系。

3.5.1 LSTM模型结构

本文构建的LSTM模型包括输入层、LSTM层、全连接层和输出层。输入层接收降维后的特征向量;LSTM层包含多个LSTM神经元,用于提取时间序列数据的特征;全连接层将LSTM层的输出进行线性变换;输出层输出风电功率的预测值。

3.5.2 模型训练与优化

(1) 损失函数:采用均方误差(MSE)作为损失函数,用于衡量预测值与实际值之间的误差。MSE的计算公式为:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

其中, n 为样本数量, y_i 为实际风电功率, $\hat{y}_i$ 为预测风电功率。

(2) 优化算法:采用Adam优化算法对模型参数进行优化。Adam优化算法结合了AdaGrad和RMSProp两种优化算法的优点,能够自适应地调整学习率,提高模型的收敛速度和稳定性。

(3) 模型训练:将训练集数据输入到LSTM模型中进行训练,通过反向传播算法不断调整模型参数,使得损失函数最小化。训练过程中,设置合适的batchsize和epochs,以避免过拟合和欠拟合问题。

4 实验验证与结果分析

4.1 实验数据

本文选取某实际风电场2022年1月1日至2022年12月31日的历史数据作为实验数据,包括气象数据、地形数据和风机运行状态数据。气象数据的时间分辨率为15分钟,共采集了35040个数据点;地形数据为风电场所在区域的数字高程模型(DEM)数据;风机运行状态数据的时间分辨率也为15分钟,共采集了35040个数据点。将前10个月(2022年1月1日至2022年10月31日)的数据作为训练集,后2个月(2022年11月1日至2022年12月31日)的数据作为测试集。

4.2 评价指标

为了评价预测模型的性能,本文采用均方根误差(RMSE)、平均绝对误差(MAE)和相关系数(R)作为评价指标。RMSE和MAE越小,说明预测值与实际值之间的误差越小,预测精度越高;R越接近1,说明预测值与实际值之间的相关性越强,预测效果越好。

均方根误差(RMSE):

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

平均绝对误差(MAE):

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

相关系数(R):

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

其中, $\bar{y}$ 和 $\bar{\hat{y}}$ 分别为实际风电功率和预测风电功率的平均值。

4.3 实验结果与分析

4.3.1 不同信息源对预测结果的影响

为了分析不同信息源对风电功率预测结果的影响,分别构建了基于单一气象信息、单一地形信息和单一风机运行状态信息的预测模型,并与基于多元信息融合的预测模型进行对比。实验结果如表1所示。

表1 不同信息源对预测结果的影响

模型类型	RMSE (MW)	MAE (MW)	R
基于单一气象信息模型	2.15	1.68	0.82
基于单一地形信息模型	3.20	2.56	0.65
基于单一风机运行状态信息模型	2.58	2.01	0.75
基于多元信息融合模型	1.52	1.18	0.90

从实验结果可以看出,基于单一信息源的预测模型性能较差,RMSE和MAE较大,R较小。而基于多元信息融合的预测模型性能明显优于单一信息源模型,RMSE和MAE分别降低了29.3%和29.8%,R提高了9.8%。这说明多元信息融合能够充分利用不同信息源的优势,提高风电功率预测的精度和可靠性。

结语

本文提出了一种基于多元信息融合的风电功率预测模型,通过融合气象、地形及风机运行状态等多源数据,显著提升了预测的精度与稳定性。实验结果显示,该模型在RMSE、MAE和R等指标上优于单一信息源模型。未来研究可从优化融合策略、引入更多影响因素、结合大数据与云计算技术以及开展实际应用等方面深入,进一步提升模型实用性与效率,为风电发展和电力系统稳定提供有力支撑。

参考文献

- [1]董鹏辉,崔世文,苏正雄,等.基于多元信息融合的风电功率预测模型研究[J].微型电脑应用,2025,41(02):153-156.
- [2]李宇钦.基于多变量信息融合与新颖TCN-former模型的短期负荷与光伏发电功率预测研究[D].南昌大学,2024.(03):145-150.
- [3]鲁泓壮,丁云飞,汪鹏宇.基于信息融合和堆叠模型的超短期风电功率预测[J].可再生能源,2022,40(03):344-349.