

# 光伏电站发电功率预测建模研究

## 问题一：太阳几何-气候耦合光伏发电偏差分析模型

### 问题一分析

光伏电站的发电功率是一个受多重因子共同驱动的非线性系统，其主要影响来源包括太阳辐射强度、地理位置、季节时间、天气变化（如云量、湿度等）、设备效率衰减等。为了系统地揭示其发电特性，我们建立理论可发电功率模型作为“理想参考基准”，并与实际历史功率数据对比，进而构建“功率偏差函数”以识别潜在影响机制。

地理信息决定了单位时间内太阳辐照在单位面积上投射的角度与强度，因此，我们首要任务是通过地理位置  $(\phi, \lambda)$  结合日地关系推导理论辐照模型，并计算出对应的理论发电功率函数  $P_{th}(t)$ 。

此模型将长周期与短周期的发电变化模式通过时域与频域耦合的方式综合建模，并利用傅里叶变换、经验模态分解（EMD）、小波变换等方法剖析波动的周期结构与成因，最终揭示实际发电系统中偏差的结构性与非结构性来源。

### 问题一建模

#### 1、理论可发电功率模型

首先建立地球绕太阳运动中太阳赤纬角  $\delta$  与时角  $\omega$  的计算公式：

$$\delta(n) = 23.45 \cdot \sin\left(\frac{2\pi}{365} \cdot (284 + n)\right)$$

$$\omega(t_{solar}) = 15 \cdot (t_{solar} - 12)$$

其中：

- $n \in [1, 365]$  表示当年的第  $n$  天，

- $t_{\text{solar}}$  表示太阳时（单位：小时）。

太阳高度角  $h$  与太阳方位角  $\alpha$  的计算表达为：

$$h(t) = \arcsin(\sin \phi \cdot \sin \delta + \cos \phi \cdot \cos \delta \cdot \cos \omega)$$

$$\alpha(t) = \arccos\left(\frac{\sin \delta \cdot \cos \phi - \cos \delta \cdot \sin \phi \cdot \cos \omega}{\cos h}\right)$$

基于太阳高度角可估计单位面积接受到的太阳直接辐射强度  $G(t)$ （忽略大气散射时近似）：

$$G(t) = G_{\text{sc}} \cdot \cos \theta(t)$$

其中  $G_{\text{sc}}$  为太阳常数， $\theta(t)$  表示太阳光与面板法线的夹角，设面板倾角为  $\beta$ ，方位角为  $\gamma$ ，则：

$$\cos \theta(t) = \sin h(t) \cdot \cos \beta + \cos h(t) \cdot \sin \beta \cdot \cos(\alpha(t) - \gamma)$$

理论发电功率为：

$$P_{\text{th}}(t) = \eta \cdot A \cdot G(t)$$

其中：

- $\eta$  为光伏组件效率 ( $0 < \eta < 1$ )，
- $A$  为组件总面积（单位： $\text{m}^2$ ）。

## 2、功率偏差比模型

定义实际功率  $P_{\text{act}}(t)$  与理论功率之间的偏差比：

$$\Delta(t) = \frac{P_{\text{act}}(t) - P_{\text{th}}(t)}{P_{\text{th}}(t)}$$

函数  $\Delta(t)$  表征系统实际运行状态与理论预期的相对偏离，用以探测系统性误差、运行效率衰退以及外部气候干扰等。

## 3、周期性与波动结构提取

我们对偏差函数  $\Delta(t)$  进行频域与小波域分析：

1. **傅里叶变换：** 提取周期分量，刻画长期季节性变化。

定义  $\Delta(t)$  的傅里叶变换为：

$$\mathcal{F}(\omega) = \int_{-\infty}^{+\infty} \Delta(t) \cdot e^{-i\omega t} dt$$

得到功率谱密度函数  $S(\omega) = |\mathcal{F}(\omega)|^2$ ，以识别年周期、半年周期、日周期等主导频率成分。

2. **小波分析：** 结合时频分析优势，提取短期异常波动，分析云层遮挡、天气骤变等因素的瞬时影响：

$$W(a,b) = \int_{-\infty}^{+\infty} \Delta(t) \cdot \psi^* \left( \frac{t-b}{a} \right) dt$$

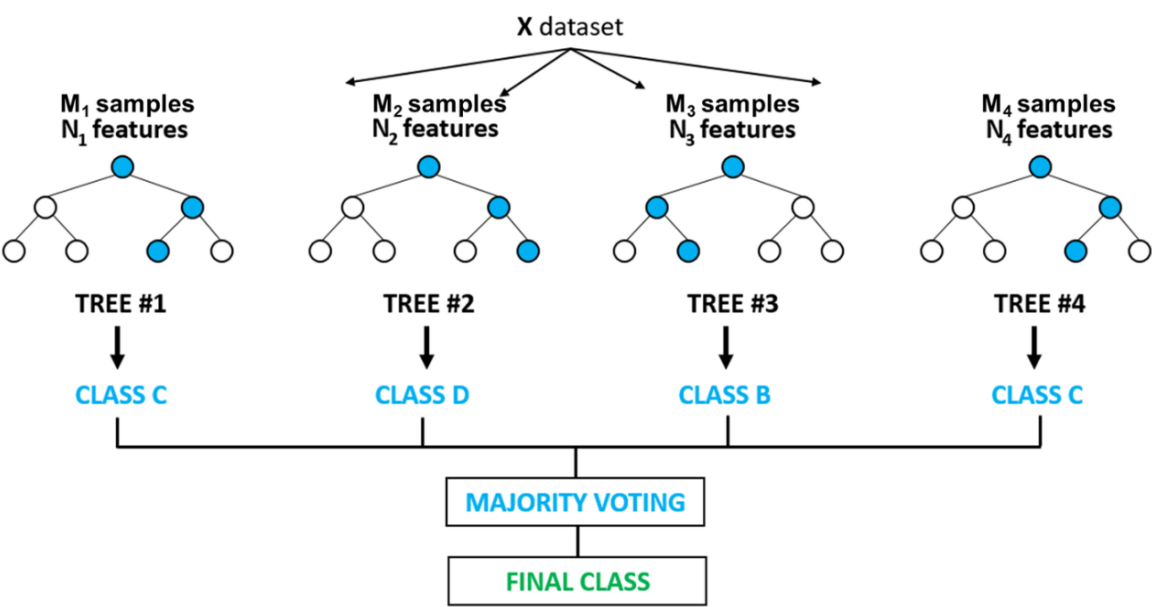
其中  $a$  为尺度参数,  $b$  为时间平移参数,  $\psi$  为母小波函数。

#### 4、偏差成因诊断与模式归纳

最后，将时间序列  $\Delta(t)$  与气象数据（如云量  $C(t)$ 、湿度  $H(t)$ 、风速  $v(t)$  等）联合建模，构建偏差函数的多元回归或非线性回归模型：

$$\Delta(t) = f(C(t), H(t), v(t), \dots) + \varepsilon(t)$$

其中  $f(\cdot)$  可选用随机森林方法,  $\varepsilon(t)$  为不可解释的噪声分量。



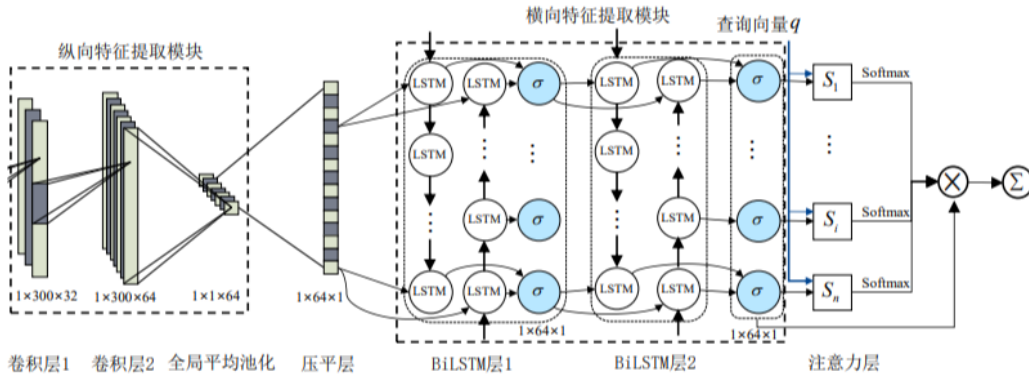
通过该模型可进一步实现：发电效率异常报警；天气因子对发电影响的敏感性评估；光伏电站性能退化诊断；区域能效差异分析。该模型为光伏系统运营维护、发电预测优化提供理论支持。

# 问题二：基于CNN-BiLSTM的光伏电站功率预测模型

## 问题二分析

本问题要求在仅依赖历史发电功率数据的前提下，构建一种准确预测光伏电站未来24小时发电功率的模型。由于不包含气象等外生变量，该任务属于典型的时间序列建模问题。光伏功率序列通常具有明显的日周期性、趋势性以及短期扰动特征，这对模型的时序依赖捕捉与局部特征提取能力提出较高要求。基于此，我们采用卷积神经网络（CNN）与双向长短期记忆网络（BiLSTM）融合的方法。CNN具备良好的局部模式识别能力，有利于提取局部波动特征；BiLSTM则能从前后两个方向捕捉时序依赖关系，更适合建模长期依赖和动态非线性特征。

为提高模型的泛化能力，我们采用滑动窗口的方式构造训练数据，同时结合多层感知机（MLP）实现最终输出。整个模型以均方误差作为目标函数，利用验证集调参，采用早停机制以防止过拟合。



## 问题二建模

设光伏电站历史功率序列为一维离散时间序列：

$$x = \{x_1, x_2, \dots, x_T\}, \quad x_t \in \mathbb{R}, \quad t \in \{1, 2, \dots, T\}$$

采用长度为  $l$  的滑动窗口构造输入序列，每个输入样本为：

$$X^{(t)} = \{x_{t-l+1}, x_{t-l+2}, \dots, x_t\}, \quad t \in \{l, l+1, \dots, T-N\}$$

其中  $N = 96$  表示未来24小时内每15分钟一次的功率预测点数，预测目标为：

$$Y^{(t)} = \{x_{t+1}, x_{t+2}, \dots, x_{t+N}\}$$

## 1. 卷积特征提取层

设输入样本  $X^{(t)}$  为一维向量，记作列向量形式：

$$\mathbf{x}^{(t)} \in \mathbb{R}^{l \times 1}$$

卷积层采用一维卷积操作，设卷积核大小为  $k$ ，卷积核权重为：

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_k \end{bmatrix} \in \mathbb{R}^{k \times 1}, \quad b \in \mathbb{R}$$

卷积层输出为：

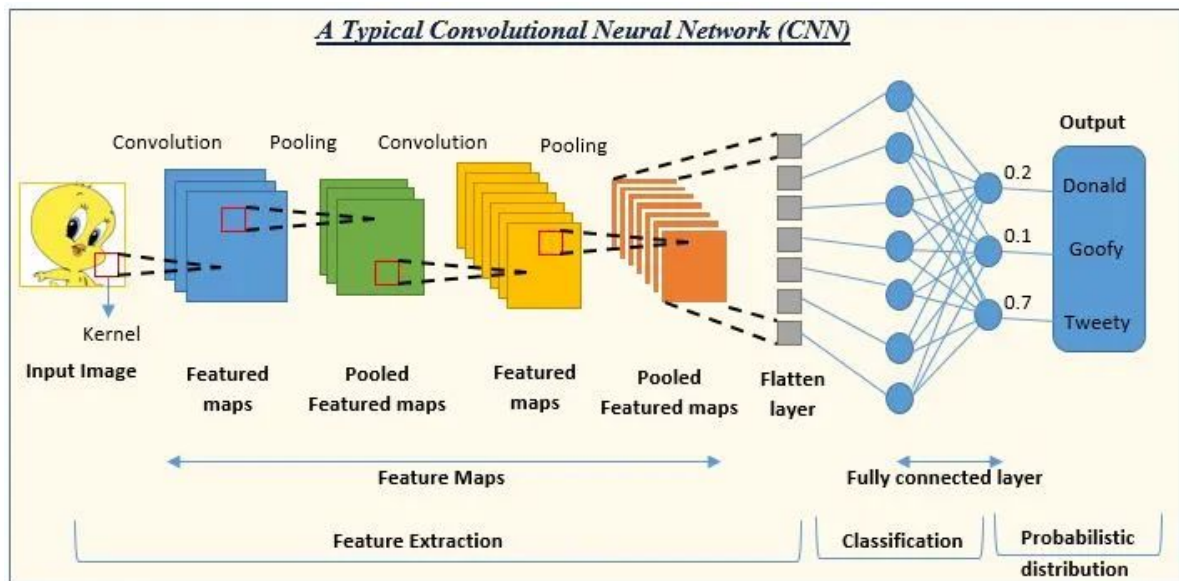
$$z_i = \sigma \left( \sum_{j=1}^k w_j \cdot x_{i+j-1} + b \right), \quad i = 1, 2, \dots, l - k + 1$$

其中  $\sigma(\cdot)$  表示激活函数，采用修正线性单元 (ReLU)：

$$\sigma(z) = \max(0, z)$$

输出序列记作

$$\mathbf{z} = \{z_1, z_2, \dots, z_{l-k+1}\} \in \mathbb{R}^{(l-k+1) \times 1}$$



## 2. 双向LSTM层

将卷积层输出作为BiLSTM的输入，记为  $\mathbf{z}$ 。BiLSTM分别从前向与后向两端处理该序列，设隐藏单元维度为  $d$ ，则前向LSTM输出为：

$$\mathbf{h}_f = \{h_f^{(1)}, h_f^{(2)}, \dots, h_f^{(l-k+1)}\}, \quad h_f^{(i)} \in \mathbb{R}^d$$

后向LSTM输出为：

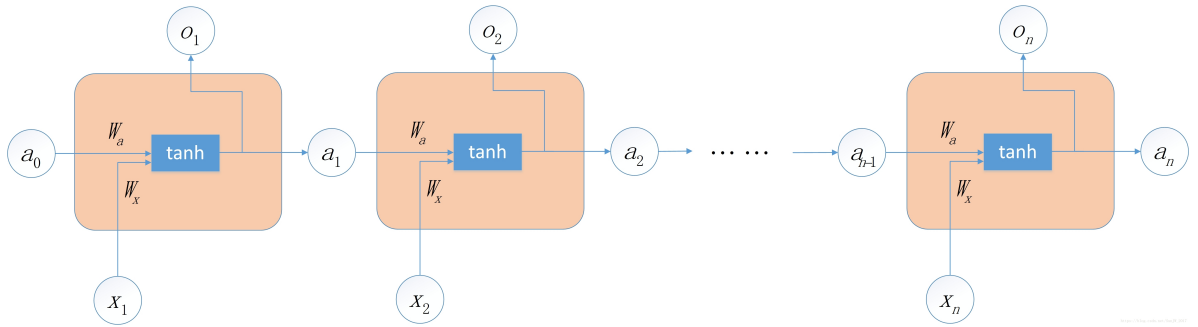
$$\mathbf{h}_b = \{h_b^{(1)}, h_b^{(2)}, \dots, h_b^{(l-k+1)}\}, \quad h_b^{(i)} \in \mathbb{R}^d$$

拼接前向与后向隐藏状态得到最终输出：

$$\mathbf{h}^{(i)} = \begin{bmatrix} h_f^{(i)}; h_b^{(i)} \end{bmatrix} \in \mathbb{R}^{2d}$$

全序列表示为：

$$\mathbf{H} = \{\mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \dots, \mathbf{h}^{(l-k+1)}\}$$



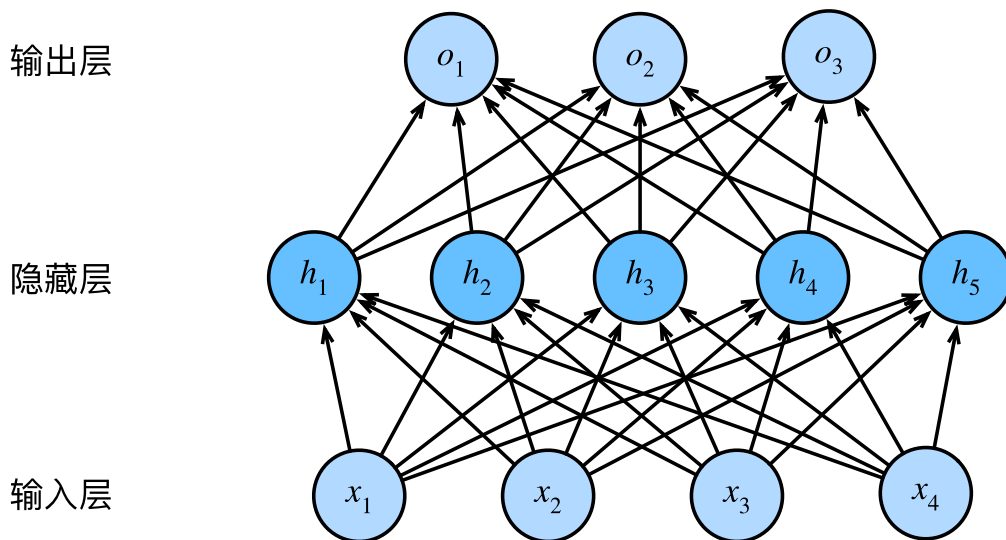
## 3. MLP输出层

我们将BiLSTM输出  $\mathbf{H}$  通过平均池化简化为定长向量：

$$\bar{\mathbf{h}} = \frac{1}{l-k+1} \sum_{i=1}^{l-k+1} \mathbf{h}^{(i)} \in \mathbb{R}^{2d}$$

再通过多层感知机 (MLP) 映射至预测值：

$$\hat{\mathbf{Y}}^{(t)} = \mathbf{W}_2 \cdot \sigma(\mathbf{W}_1 \cdot \bar{\mathbf{h}} + \mathbf{b}_1) + \mathbf{b}_2, \quad \hat{\mathbf{Y}}^{(t)} \in \mathbb{R}^N$$



## 4. 损失函数

采用均方误差（MSE）作为损失函数，形式为：

$$\mathcal{L} = \frac{1}{M} \sum_{t=1}^M \frac{1}{N} \sum_{i=1}^N \left( \hat{x}_{t+i}^{(t)} - x_{t+i}^{(t)} \right)^2$$

其中  $M$  为样本数量， $N = 96$  为预测步数。

模型训练采用Adam优化器进行参数更新，使用滑动窗口生成训练数据，划分验证集调节超参数。最终在测试集上评估预测性能，采用考核要求中指定的误差度量标准（如均方误差、平均绝对误差等）进行模型效果评估与对比分析。

# 问题三：融合NWP信息的光伏电站日前发电功率预测模型

## 问题三分析

为提升光伏电站的日前发电功率预测精度，我们在前期模型基础上引入数值天气预报（Numerical Weather Prediction，记为 NWP）信息。NWP 模拟结果包含多个关键气象变量，如温度、湿度、风速、云量与太阳辐射等，它们与光伏组件的输出存在显著的物理相关性。

为了有效利用这一类异构数据，我们设计一种融合历史功率序列与未来气象预测数据的多模态深度学习模型。该模型以功率

## 问题三建模

设：

- 光伏电站在时刻  $t$  的发电功率为  $x_t \in \mathbb{R}$ ;
- 历史发电功率序列为  $\{x_{t-l+1}, x_{t-l+2}, \dots, x_t\}$ ;
- 在时间  $t$ , NWP 提供的  $d$  维气象预测变量为  $M_t = (m_t^{(1)}, m_t^{(2)}, \dots, m_t^{(d)})^\top \in \mathbb{R}^d$ ;
- 历史 NWP 序列为  $\{M_{t-l+1}, M_{t-l+2}, \dots, M_t\}$ ;
- 最终的多模态融合输入向量为：

$$\tilde{X}_t = \left[ x_{t-l+1}, \dots, x_t, m_{t-l+1}^{(1)}, \dots, m_t^{(1)}, \dots, m_{t-l+1}^{(d)}, \dots, m_t^{(d)} \right]^\top \in \mathbb{R}^{l(1+d)}$$

我们采用**卷积神经网络 (CNN)** 与**双向长短期记忆网络 (BiLSTM)** 进行多层次特征提取：

- CNN 层**：用于捕捉输入序列中局部时间片段的局部模式与趋势；
- BiLSTM 层**：用于建模远程时间依赖与趋势变化；
- 全连接层**：将提取的深度特征映射至预测目标空间。

记最终模型为函数  $f(\cdot; \theta)$ ，其中  $\theta$  表示模型参数，则发电功率的预测结果为：

$$\hat{x}_{t+1} = f(\tilde{X}_t; \theta)$$

为了量化预测性能，我们使用以下两种误差评价指标：

- 均方根误差 (Root Mean Squared Error)：

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_{t+i} - \hat{x}_{t+i})^2}$$

- 平均绝对百分比误差 (Mean Absolute Percentage Error)：

$$\text{MAPE} = \frac{100}{N} \sum_{i=1}^N \left| \frac{x_{t+i} - \hat{x}_{t+i}}{x_{t+i}} \right|$$

其中  $x_{t+i}$  表示真实值， $\hat{x}_{t+i}$  表示预测值， $N$  为测试样本总数。



在完成模型训练与测试后，我们基于不同气象情景评估 NWP 信息的引入是否对预测精度有显著贡献。我们定义如下四类典型天气场景：

1. **晴天场景**：气象变量稳定，太阳辐射高且波动小；
2. **多云场景**：云层变化频繁，短时辐射波动大；
3. **阴雨场景**：日照弱，天气系统稳定但输出低；
4. **雾霾场景**：辐射受污染影响，表现为低频非线性扰动。

对于每类天气场景，分别计算有无 NWP 信息条件下的预测误差指标：

- 若满足：

$$\text{RMSE}_{\text{NWP}} < \text{RMSE}_{\text{baseline}} \quad \text{或} \quad \text{MAPE}_{\text{NWP}} < \text{MAPE}_{\text{baseline}}$$

则判定该场景下 NWP 信息对提升模型预测精度有效。

我们进而构建如下场景划分图谱：

- 晴天：误差下降较小，NWP 效果不显著；
- 多云：误差显著下降，NWP 预测提升效果最好；
- 阴雨：模型有一定提升，但效果有限；
- 雾霾：模型表现不稳定，需进一步精细建模与变量筛选。

最终通过在多个时间段与地点上的实验验证以上划分策略的有效性。

## 问题四：空间降尺度光伏功率预测模型

### 问题四分析

传统数值天气预报（NWP, Numerical Weather Prediction）系统的空间分辨率通常为 1km 至 10km，其空间粒度远大于MW级光伏电站的物理覆盖范围（通常为  $10^2 \sim 10^3 \text{m}^2$ ）。这导致NWP在对局地光照、温度、风速等关键气象变量的估计上存在一定误差，进而影响基于气象驱动变量的光伏功率预测模型精度。

空间降尺度旨在从粗分辨率的NWP数据中，利用地理特征、历史观测数据及空间统计关系，构建 finer-resolution 的气象估计，从而提升下游光伏功率预测模型的性能。

因此，我们尝试构建融合空间插值、机器学习回归及混合策略的空间降尺度方法，将降尺度结果引入光伏发电功率预测模型，检验其在不同误差指标下的改进幅度，并分析其适用性与有效性。

## 问题四建模

我们将建模过程分为三个核心阶段：

### 1. 空间降尺度气象建模

我们考虑以下三类方法对粗分辨率气象变量进行空间降尺度：

- (a) 基于空间距离的权重插值模型 (IDW)：

定义目标站点位置为  $\mathbf{x}_0$ ，其周围  $K$  个NWP格点的气象观测为  $M_1, M_2, \dots, M_K$ ，格点位置为  $\mathbf{x}_j$ ，则目标点气象变量的估计值为：

$$\hat{M}_0 = \frac{\sum_{j=1}^K w_j M_j}{\sum_{j=1}^K w_j}, \quad w_j = \frac{1}{\|\mathbf{x}_j - \mathbf{x}_0\|^p}$$

其中  $p > 0$  为经验调节参数，通常取  $p = 2$ ， $\|\cdot\|$  表示欧氏距离。

- (b) 克里金插值 (Kriging)：

引入空间协方差结构，设  $\text{Cov}(M_i, M_j) = C(h)$  为距离  $h = \|\mathbf{x}_i - \mathbf{x}_j\|$  的协方差函数，则估计值为：

$$\hat{M}_0 = \sum_{j=1}^K \lambda_j M_j$$

使得方差  $\text{Var}[\hat{M}_0 - M_0]$  最小，约束条件为  $\sum \lambda_j = 1$ 。求解过程涉及半方差函数估计及协方差矩阵求逆，体现对空间结构的显式建模。

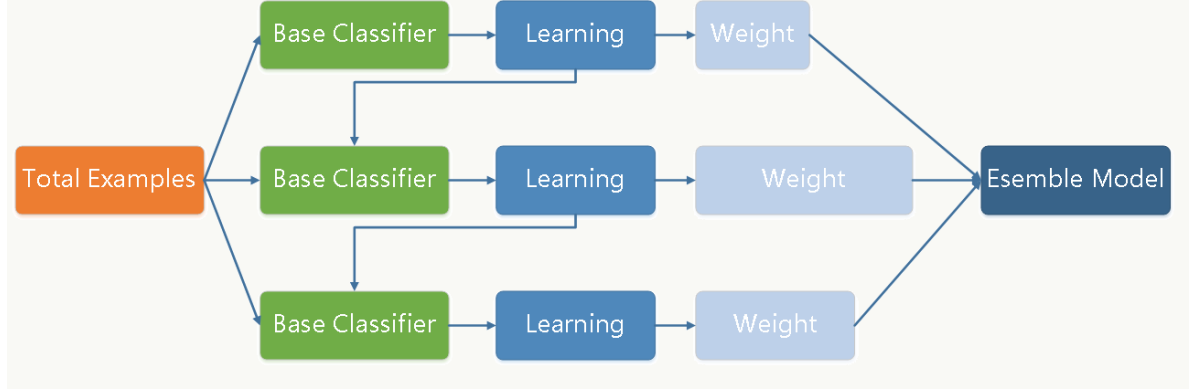
- (c) 基于机器学习的非线性降尺度回归模型：

构建映射  $f: \mathbb{R}^d \rightarrow \mathbb{R}$ ，以周边多个格点的气象变量  $\{M_j\}_{j=1}^K$  及地理特征（经度、纬度、海拔等）作为输入  $\mathbf{x} \in \mathbb{R}^d$ ，目标气象变量为输出  $y$ ，训练非线性回归模型——梯度提升树以逼近：

$$\hat{y} = f(\mathbf{x}; \theta)$$

其中  $\theta$  为模型参数，通过最小化预测误差目标函数如均方误差（MSE）训练得到：

$$\mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^N (y_i - f(\mathbf{x}_i; \theta))^2$$



## 2. 光伏功率预测模型构建

光伏功率预测模型以气象驱动变量为输入，输出未来时刻的光伏功率。我们采用标准的回归框架，记输入特征为  $\mathbf{z}_t$ （包括辐照度、温度、风速等气象变量），输出为功率  $P_t$ ，模型结构为：

$$\hat{P}_t = g(\mathbf{z}_t; \phi)$$

其中  $g(\cdot)$  为回归模型（CNN-BILSTM模型）， $\phi$  为模型参数。我们分别以原始NWP数据与降尺度NWP数据作为输入特征  $\mathbf{z}_t$ ，构建两个版本的预测模型，以比较其性能。

## 3. 模型性能对比与评估

采用以下误差指标评估降尺度方法的有效性：

- 均方根误差（RMSE）：

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T (P_t - \hat{P}_t)^2}$$

- 平均绝对百分比误差（MAPE）：

$$\text{MAPE} = \frac{100\%}{T} \sum_{t=1}^T \left| \frac{P_t - \hat{P}_t}{P_t} \right|$$

- 判定系数 ( $R^2$ ) :

$$R^2 = 1 - \frac{\sum_{t=1}^T (P_t - \hat{P}_t)^2}{\sum_{t=1}^T (P_t - \bar{P})^2}$$

其中  $\bar{P}$  为实测功率的均值。

我们对比原始NWP与降尺度NWP在上述指标上的表现，若后者显著优于前者，说明降尺度过程有效增强了局地气象描述能力，进而提高了光伏功率预测精度。

进一步的敏感性分析可以探讨以下问题：

- 不同地形、季节、天气类型下降尺度的稳定性。
- 插值/学习模型误差对最终功率预测误差的传递规律。
- 空间降尺度模型与时间序列建模结合的耦合效应。

综上所述，构建精细化的空间降尺度模型，显著提升NWP数据在小尺度光伏应用中的预测能力，是提升功率预测精度的有效途径，尤其适用于气象条件复杂、地形异质性强的区域。

通过上述四个问题的系统建模，结合 PV0DDataset 提供的高质量数据，以及深度融合 CNN 与 BiLSTM 的多模态建模策略，能够在理论与工程应用中显著提升光伏功率预测精度，为电网调度提供坚实的数据支持。