新能源出力分析与预测

——《能源互联网导论》课程期末大作业

苏博文	电 05	2020010599
卢梁宇宸	电 05	2020010547
杨旻	电 05	2020010618

2022 年春

目　　录

1 必做任务一：基于历史数据的日前新能源出力预测

题目回顾：利用所给数据中的天气历史数据和新能源出力历史数据，对欧洲各个国家（至少选择 3 个国家）的光伏出力、风电出力（总出力，onshore 出力，offshore 出力）进行日前点预测。

1.1 数据预处理

1.1.1 对气象数据的处理

weather 数据集中对于每个国家有三个特征'temperature'、'(name_of_country)_radiation_direct_horizontal'、'(name_of_country)_radiation_diffuse_horizontal'。缺失值比较少，直接用时序前后两个数的平均值替代。对于连续缺失的特殊情况，采用线性平均插值方法处理。

1.1.2 对负荷数据的处理

time_series 数据集中对于每个国家的光伏和风电有许多标签，例如电价，装机容量，光伏总出力，风电的岸上出力、海上出力和总出力。我们提取其中本次预测的重点标签'(name_of_country)_solar(wind)_generation_actual'，由于气象数据非常有限，而且直射辐射与漫射辐射主要用于预测光伏出力，对风电出力的预测作用不大。故对于风电的数据集，我们加入'power_1h_ahead'、'power_24h_ahead'、'power_48h_ahead'、'power_72h_ahead'、'power_168h_ahead'（分别是 1 小时、1 天、2 天、3 天和一周前的数据）辅助预测。

1.1.3 对时间数据的处理

所给气象和负荷数据的时间戳粒度都是 1h，我们分别提取星期、月份和年份的信息。通过光伏的出力数据可以看出，各国的数据都是基于各国时区统计，因此不再进行时差的处理。为了平滑处理小时、星期、月份的周期性，引入虚拟日（Virtual_Day）概念，对小时、星期、月份的信息进行三角函数数值抽样，得到特征'Virtual_Day_sin'、'Virtual_Day_cos'、'hour_sin'、'hour_cos'、'month_sin'、'month_cos'。

1.1.4 数据归一化

得到最终 dataset 后，对所有数据进行[0,1]区间归一化处理，以提高模型预测精度。

1.2 模型的简单介绍

1.2.1 LSTM（长短时记忆网络）[1]

在使用深度学习处理时序问题时，LSTM 是最常使用的模型之一。LSTM 之所以在时序数据上有着优异的表现是因为 LSTM 在 t 时间片时会将 t 时间片之前的隐节点作为当前时间片的输入，也就是 LSTM 具有图 1.2.1 的结构。这样有效的原因是之前时间片的信息也用于计算当前时间片的内容，而传统模型的隐节点的输出只取决于当前时间片的输入特征。[2]

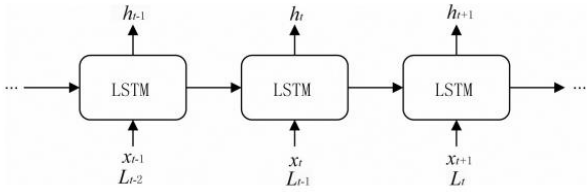

图 1.2.1：LSTM

1.2.2 SVM（支持向量机）

支持向量机（SVM）是一种分类算法，sklearn 中的 model_SVR 模块可以用来做回归分析。回归的目的是得到一个能够尽量拟合训练集样本的模型$f(x)$，通常用的方法是构建一个样本标签与模型预测值的损失函数，使损失函数最小化从而确定模型$f(x)$。SVR 试图通过找到一个回归超平面，让一个集合的所有数据到该平面的距离最近。

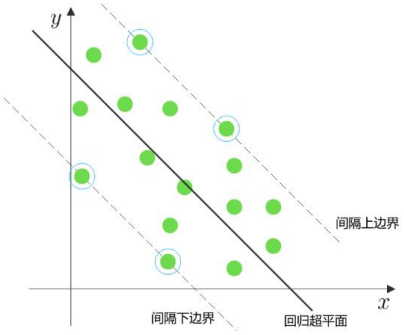

图 1.2.2：SVR 使靠超平面最远的样本点之间的间隔最小

1.2.3 GBDT（梯度提升决策树）

GBDT 属于集成算法的一种，基分类器是回归树。分类问题也是回归树，最后再用 sigmoid 或者 softmax 函数计算类别，是一种 boosting 算法，即逐步拟合逼近真实值，是一个串行的算法，可以减少 bias（误差）却不能减少 variance（偏差），因为每次基本都是全样本参与训练，不能消除偶然性的影响，但每次都逐步逼近真实值，可以减少误差。模型构建步骤如下：初始化，就是所有样本值都初始化为均值，算出来第一轮残差值（残差减去当前值的平方可以作为损失函数）；误差和特征值带进去构建树，划分标准为平方误差减少最多的那个特征值，然后划分到叶子节点，用均值作为这一轮的预测值，然后更新目标值和残差值，加上学习率；重复第以上步骤，直至满足终止条件。终止条件不是达到最大树的个数，就是最终的目标值变化幅度小于特定值；更新累积目标值作为最终的预测结果。

1.2.4 RF（随机森林）

随机森林指的是利用多棵决策树对样本进行训练并预测的一种分类器。它包含多个决策树的分类器，并且其输出的类别是由个别树输出的类别的众数而定。随机森林是一种灵活且易于使用的机器学习算法，即便没有超参数调优，也可以在大多数情况下得到很好的结果。随机森林也是最常用的算法之一，因为它很简易，既可用于分类也能用于回归。

随机森林集成了所有的分类投票结果，将投票次数最多的类别指定为最终的输出，这就

是一种最简单的 Bagging 思想。

1.2.5 MLP（多层感知器）

多层感知器（Multi-Layer Perceptron），即 MLP 算法，也被称为前馈神经网络，或者被称为人工神经网络（Artificial Neural Network, ANN）。

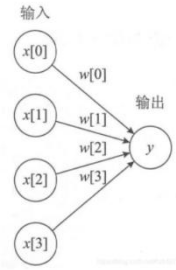

图 1.2.3(a)：线性网络

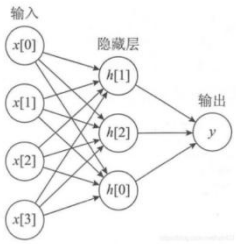

图 1.2.3(b)：加入隐藏层的神经网络

图 1.2.3(b)中，输入的特征和预测的结果用节点进行表示，系数 W 用来连接这些节点。而在 MLP 模型中，算法在过程里添加了隐藏层（Hidden Layers），然后在隐藏层重复进行上述加权求和计算，最后再把隐藏层所计算的结果用来生成最终结果。这样一来，模型要学习的特征系数（权重）就会多很多了。每一个输入的特征和隐藏单元（hidden unit）之间，都有一个系数，这一步也是为了生成这些隐藏单元。而每个隐藏单元到最终结果之间，也都有一个系数。在生成隐藏层之后，会使用激活函数对激活单元进行非线性化，因为非线性处理是为了将样本特征进行简化，从而使神经网络可以对复杂的非线性数据集进行学习。

1.3 实验结果

1.3.1 奥地利光伏出力预测结果

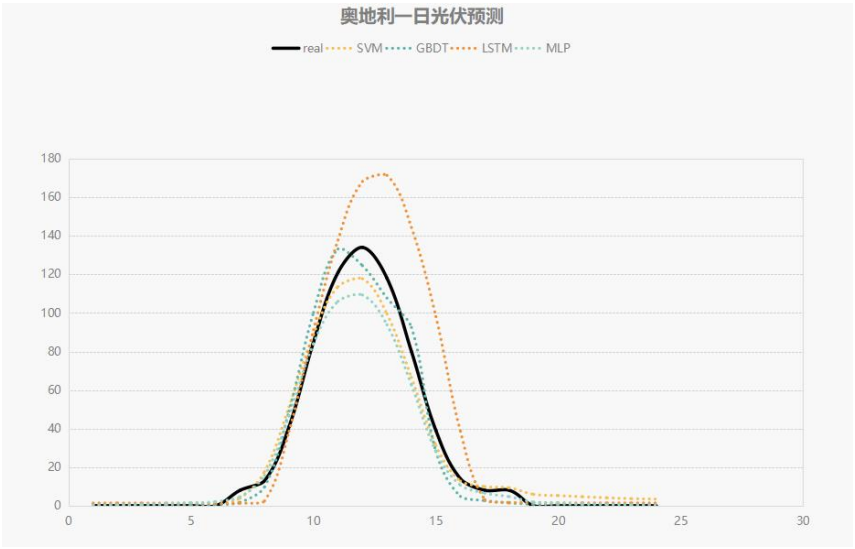

图 1.3.1：奥地利 2017.1.7 光伏出力预测结果

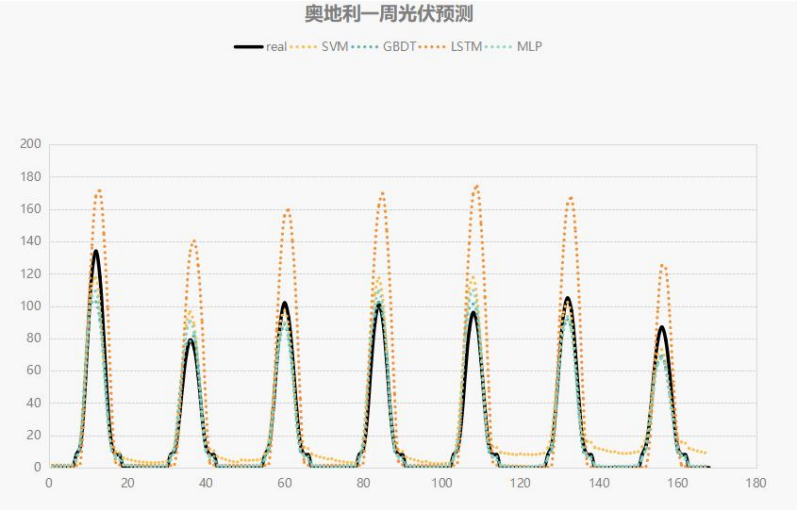

图 1.3.2：奥地利 2017.1.7 至 2017.1.13 光伏出力预测结果

1.3.2 奥地利风电出力预测结果

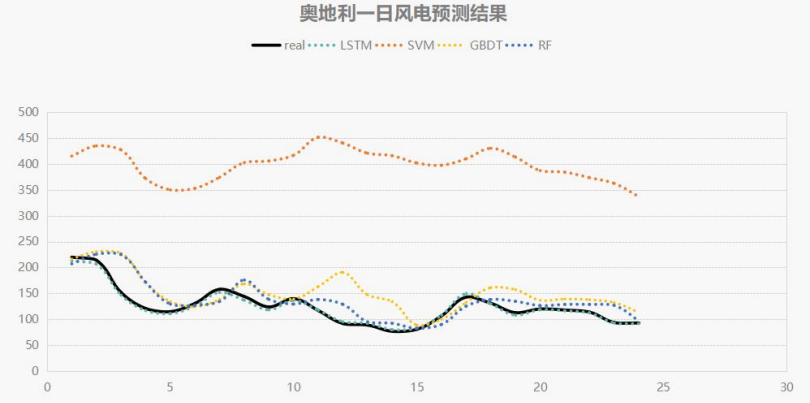

图 1.3.3：奥地利 2017.1.1 风电出力预测结果

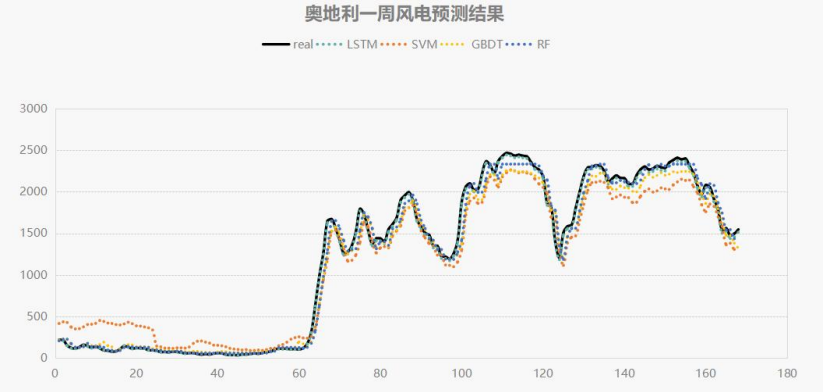

图 1.3.4：奥地利 2017.1.1 至 2017.1.7 风电出力预测结果

1.3.3 比利时光伏出力预测结果

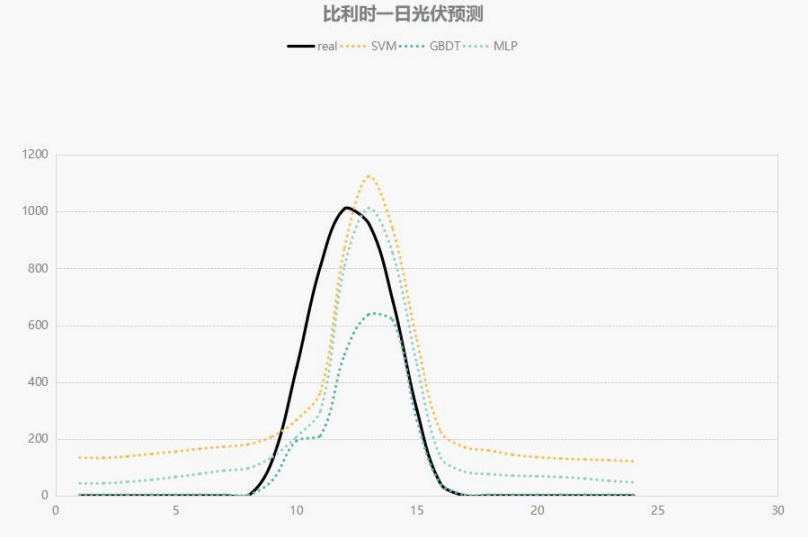

图 1.3.5：比利时 2017.1.7 光伏出力预测结果

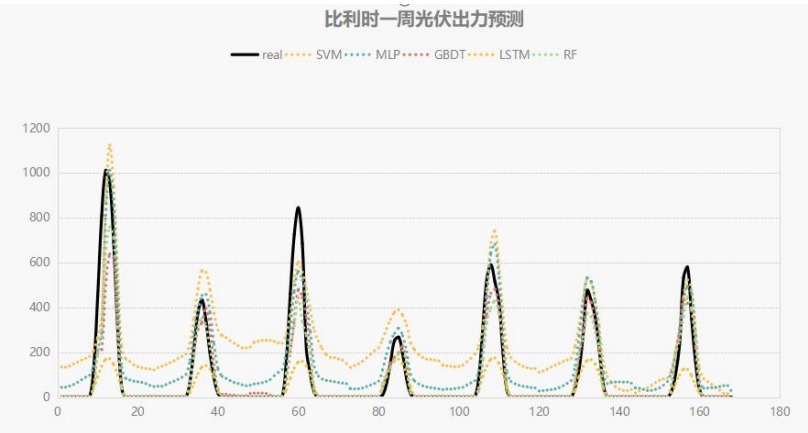

图 1.3.6：比利时 2017.1.7 至 2017.1.13 光伏出力预测结果

1.3.4 比利时风电出力预测结果

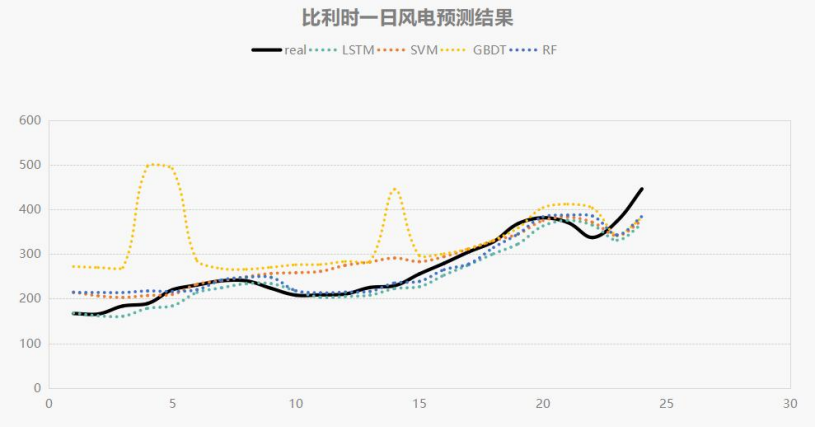

图 1.3.7：比利时 2017.1.1 风电出力预测结果

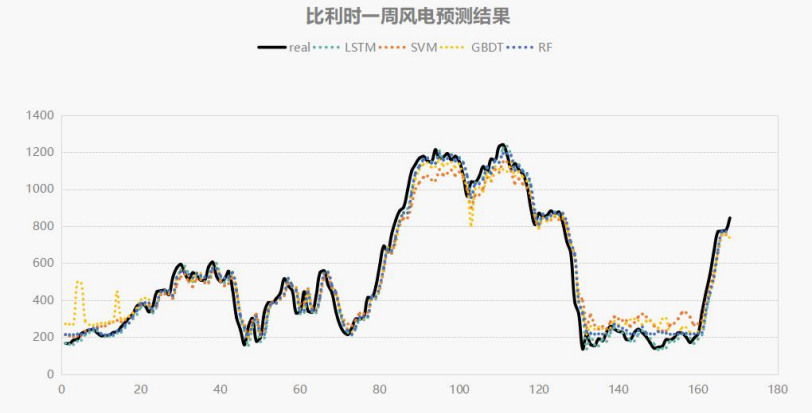

图 1.3.8：奥地利 2017.1.1 至 2017.1.7 风电出力预测结果

1.3.5 丹麦光伏出力预测结果

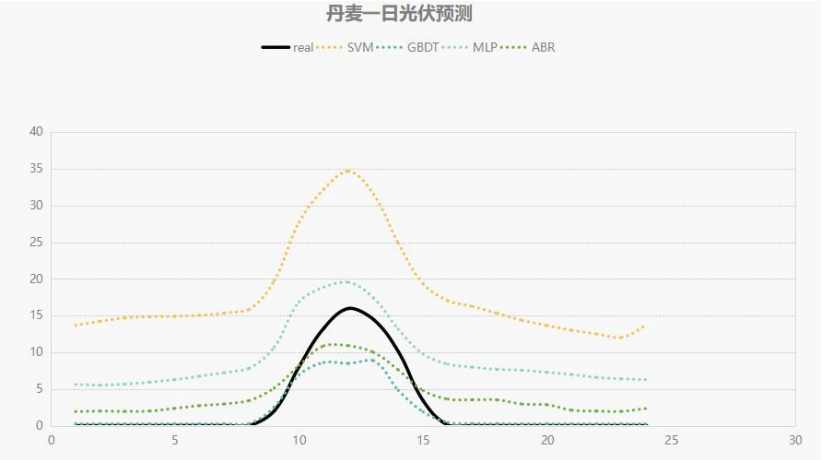

图 1.3.9：丹麦 2017.1.7 光伏出力预测结果

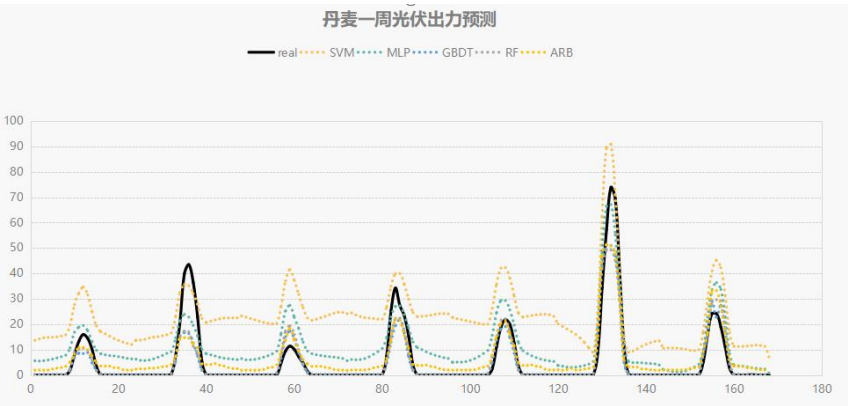

图 1.3.10：丹麦 2017.1.7 至 2017.1.13 光伏出力预测结果

1.3.6 丹麦风电出力预测结果

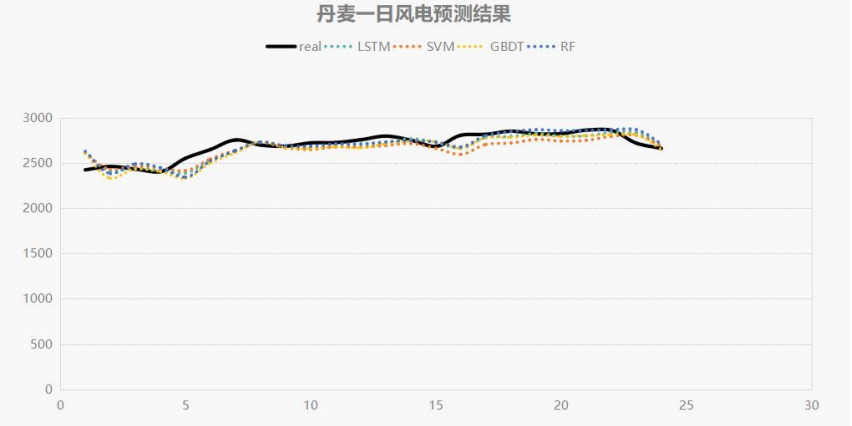

图 1.3.11：丹麦 2017.1.1 风电出力预测结果

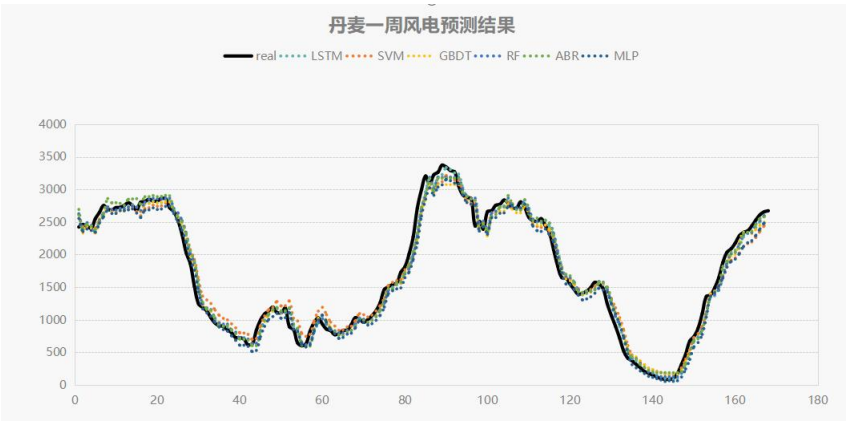

图 1.3.12：丹麦 2017.1.1 至 2017.1.7 风电出力预测结果

下面对预测结果进行分析。

首先分别给出光伏预测和风电预测的预测精度，考虑到周预测结果的数据量更大，得到的数据随机性更小，预测精度更可靠，故以下预测精度均是周预测的精度。

以 RMSE 为衡量指标（真实值中含有 0，故 MAPE 在光伏预测中不适用）的光伏预测精度如下（单位：MW）：

表 1.3.1：光伏周预测精度

	LSTM	SVM	GBDT	MLP	RF	ABR
AT_solar	28.08	8.48	6.37	6.54	7.02	
BE_solar		168.48	92.07	94.33	87.88	
DK_solar		18.46	5.14	7.51	5.17	5.71

结合实际数据和上表，实际出力越多，以 RMSE 为指标衡量的误差会增大。结合光伏出力图，可以看出各个模型基本都能契合光伏的分布特点，也就是夜晚为 0，而午后达到峰值；但有些模型对于峰值的预测偏高或者偏低，不是特别准确。

以 MAPE 为衡量指标的风电出力预测精度如下：

表 1.3.2：风电周预测精度

	LSTM	SVM	GBDT	MLP	RF	ABR
AT_wind	2.15%	70.04%	17.43%		11.04%	
BE_wind	11.64%	18.76%	19.43%		12.54%	35.33%
DK_wind	8.16%	12.63%	10.72%	10.52%	8.74%	10.79%

结合实际数据和上表，丹麦的风机容量大于奥地利和比利时的风力容量，而通过表中数据明显可以看出丹麦风电出力 MAPE 都比较小，预测精度较高。其中 LSTM 在预测中的表现很好，是因为 LSTM 的模型特征决定了其在时序预测场景中的强大能力，详细分析见“必做任务三”部分。

2 必做任务二：不同国家的新能源出力分析

题目回顾：利用所给数据（数据集 1，2）中的天气历史数据和新能源出力历史数据，对欧洲各个国家的光伏出力、风电出力（总出力，onshore 出力，offshore 出力）进行统计分析，例如可以分析对比不同国家在光伏出力和风电出力占比上的区别，不同国家在不同季节的光伏和风电出力规律差异，并结合具体国家的地理位置等信息进行分析。

2.1 数据预处理

2.1.1 发电数据处理

选取奥地利，比利时，丹麦三个国家为主要分析对象，其所给发电数据自 2015-2020 年，将所给发电数据分为光伏发电，陆上风力发电以及海上风力发电并单独创建文件以便于数据读取处理。

2.2.2 天气数据处理

同样选取奥地利，比利时，丹麦三国为主要分析国家，其所给天气数据自 1992-2019 年，为了与发电数据对应，因此只选取 2015-2019 年天气数据。并将天气数据分为直射光强与漫射光强，分别创建文件。

2.2 数据分析

2.2.1 数据可视化

应用 MATLAB 软件，读取预处理后的各国各类数据，由于数据较多，直接对数据进行分析工作量较大且分析效果并不理想，所以对其以周为单位进行求天平均值，即先求周和再除以一周的天数，但由于周和求出后与天平均值只差 7 这一常数，因此直接求出周和后，应用 plot 函数创建图表。

2.2.2 数据对比

首先将奥地利、比利时、丹麦三国风力发电量用 plot 函数绘制于一张图中以进行对比，结果如下：

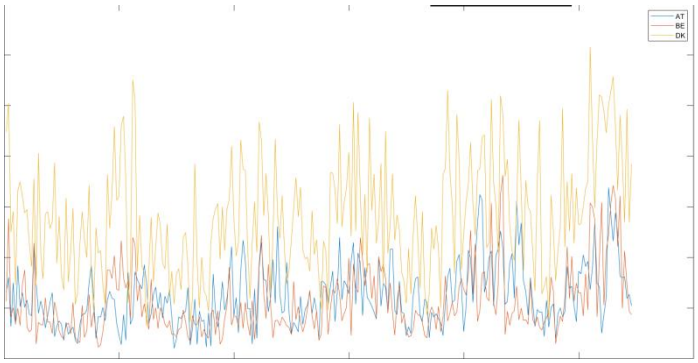

图 2.2.1：奥地利、比利时、丹麦等三国风力发电量对比

可见丹麦风力发电量远高于其他两国。根据其地理位置以及气候条件进行分析。第一，

丹麦本身自然资源相对匮乏，石油，天然气等化石燃料极少且大量依赖进口，因此导致他们应用石油等燃料发电成本较高，动力不足。

第二，丹麦三面环海，地势低平，且为受季风影响较强的温带海洋性气候，经常受到大西洋吹来的西南风影响，并且他广阔的丘陵几乎纵贯了整个半岛，而东部沿岸地区有许多的夹弯和沟谷，因此东海岸地区没有直接受到强风浪冲击但又拥有较多的风能资源，因此十分有利于风力发电。

第三，丹麦开展风力发电历史悠久，技术成熟。他们早在 1891 年就开始风电研究，第一次世界大战和第二次世界大战导致的石油短缺刺激了丹麦的风电发展；而之后的 1973，1979 年石油禁运、能源危机以及环保运动的展开更一步推动了丹麦风电产业发展。

之后对于三个国家光伏发电量进行对比，有：

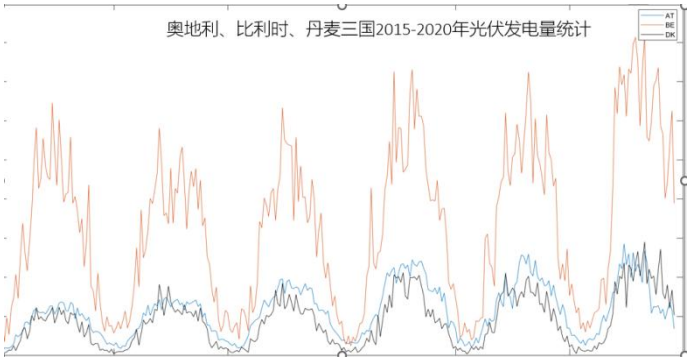

图 2.2.2：奥地利、比利时、丹麦等三国光伏发电量对比

可见比利时光伏发电量高于其他两国家，对其进行进一步分析，将其光伏发电量与光照强度绘制为图像，可得：

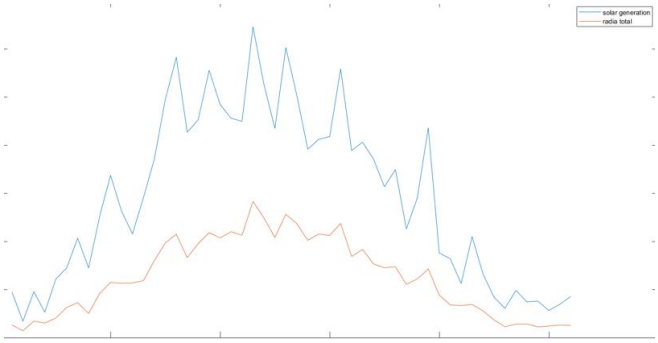

图 2.2.3：比利时光伏发电量与光照强度

体现出来明显的光伏发电的基本特点，冬季光强弱光伏发电量较小，春季光强增长速度较大，光伏发电量快速增长。夏季光强最大，光伏发电量最大。秋季光强下降较快，光伏发电量逐渐减小，总体上可以看到光伏发电量与光照强度正相关，也与直观感受符合。

最后，用比利时光伏发电量除以其直射光强以表征光伏发电效率，再与光照强度总量做对比，有：

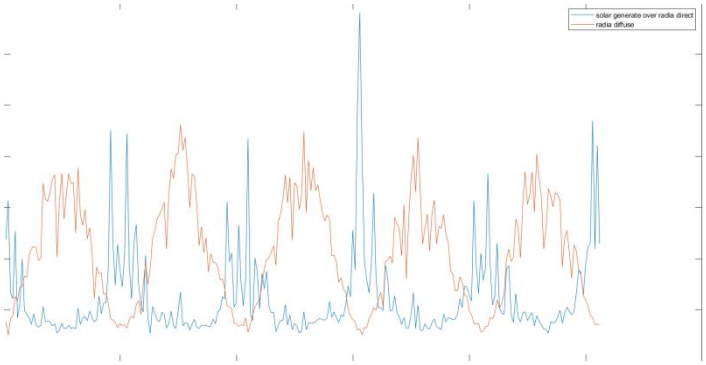

图 2.2.4：比利时光伏发电效率与光照强度

可见当光照强度较强时，其效率相对较低，可以认为比利时仍有较多的光照强度尚未被利用，仍有较大光伏发电发展空间。

3 必做任务三：日前新能源出力点预测模型分析

题目回顾：对训练得到的日前新能源出力点预测模型进行分析，具体分析影响模型预测结果的重要变量，影响程度及其合理性，进而解释模型的决策过程。

3.1 光伏预测模型 model_GBDT 分析（以奥地利为例）

3.1.1 model_GBDT 各特征重要性分析

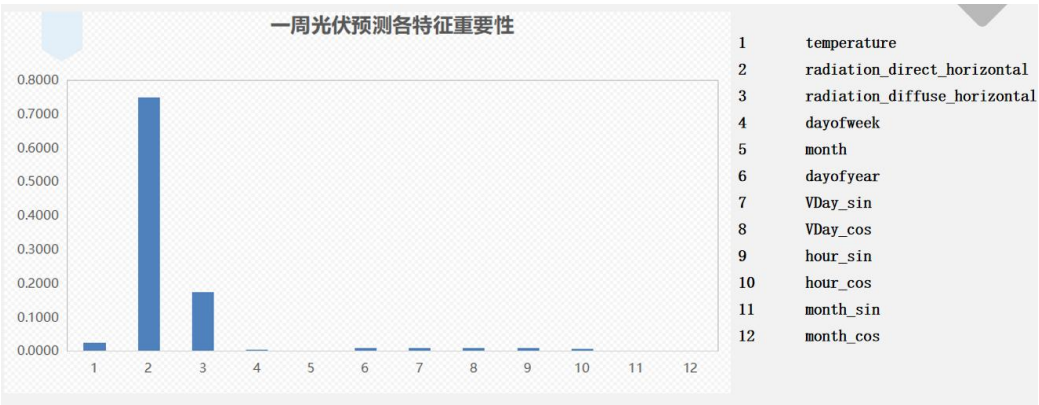

图 3.1.1：光伏出力周预测各特征重要性柱状图

由柱状图可知，在光伏预测模型 model_GBDT 中，对预测起决定性作用的特征为直射辐射强度和漫射辐射强度，符合直观感受。这也是光伏预测的结果比较精确的重要原因。

3.1.2 光伏出力理论模型对比分析

以下是光伏出力的理论模型。

$$P_S(I_T,T) = I_T\eta_S(I_T,T)A_S$$

其中I_T表示辐照度，T表示电池温度，$\eta_S(I_T,T)$表示光伏阵列的效率，与温度和太阳辐射相关，A_S表示光伏阵列面积。相关参数以及计算方式可以在文献[3]中找到。

从计算光伏出力的理论模型来看，在短期（一天、一周）内，光伏阵列的面积是不变的，所以起作用的就是辐照度和效率，而效率也与太阳辐射和温度有关，所以在任务一的光伏预测中用温度、直射辐射和漫射辐射作为特征是比较合理的，这也是光伏出力预测比较精准的原因所在。

3.2 风电预测模型 model_GBDT 分析（以奥地利为例）

3.2.1 model_GBDT 各特征重要性分析

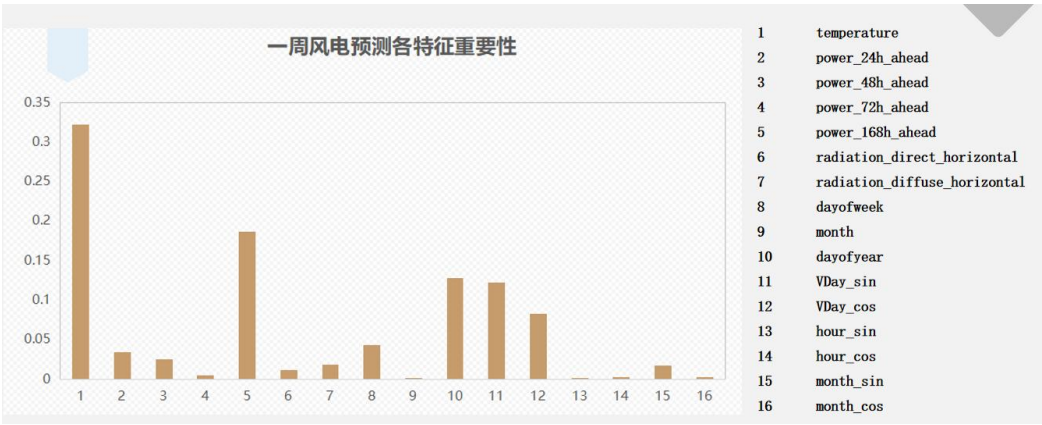

图 3.2.1：风电出力周预测各特征重要性柱状图（优化前）

由上柱状图可知，在风电预测模型 model_GBDT 中，对预测起决定性作用的特征为温度和一周前的出力值，不太符合直观感受。这也是优化前风电预测非常不准确的原因。

与其他模型不同的时，model_LSTM 引入了前一时刻的出力值，因此预测的效果非常好。受此启发，我们把’power_1h_ahead’特征加入 model_SVM 等模型中，对模型的训练过程进行优化。

以下为优化后的特征重要性。

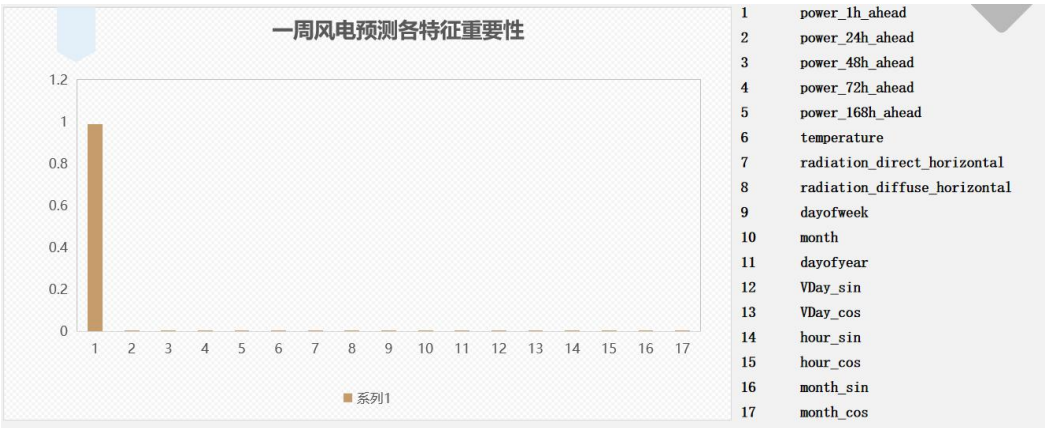

图 3.2.2：风电出力周预测各特征重要性柱状图（优化后）

由上柱状图可知，在风电预测模型 model_GBDT 中，对预测起决定性作用的特征为上一小时的风电出力值。而经过这次优化后，风电预测结果明显优于优化前。说明在缺乏直接相关因素（例如风速、海拔、气压）的情况下，模型将上一时刻的出力值作为最重要的特征。

3.2.2 风电出力理论模型对比分析

以下是风电出力的理论模型[4]。

单个风机出力模型为：

$$P_w\left(v\right)=\begin{cases}0,\ v<v_{ci}\\ P_{\mathrm{wr}}\dfrac{v^3-v_{ci}^3}{v_r^3-v_{ci}^3},\ v_{ci}\le v<v_r\\ P_{\mathrm{wr}},\ v_r\le v<v_{co}\\ 0,\ v\ge v_{co}\end{cases}$$

其中，v_{ci}、v_{co}和v_r分别为风机的切入、切出和额定转速。

风场出力模型为：

$$P_{\mathrm{WF}}\left(v\right)=\eta_w N_w P_w\left(v\right)$$

其中，η_w是考虑如尾流效应、风机损耗等的效率，N_w是风机数量。

从计算风电出力的理论模型来看，在短期（一天、一周）内，风机数量和风机效率是不变的，所以起作用的就是风速。因此对于优化前的模型来讲，以温度作为最重要的特征当然无法得到准确的预测结果。所以在没有外部数据的支撑下，只能将其处理成一个纯时序预测，因此 model_LSTM 的出色表现也就不足为奇。

4 选做任务：基于历史数据进行日前新能源出力概率预测

题目回顾：利用所给数据（数据集 1，2）中的天气历史数据和新能源出力历史数据，对欧洲各个国家的光伏出力、风电出力（总出力，onshore 出力，offshore 出力）进行日前概率预测（即预测未来一段时间内，每个时间点可能的出力概率分布）。

4.1 概率预测思路简介[5]

与传统确定性预测相比，概率预测的预测形式及数学意义具有显著差异。确定性预测通常以数学期望、中位数等单点值作为输出，提供的预测信息较为有限；概率预测则以分位数估计、预测区间估计、概率密度估计为输出，提供的是待预测对象较为完整的概率统计信息。确定性预测与概率预测之间有密切的联系，我们可以通过对确定性误差进行统计分析，对其进行概率分布拟合，得到最佳的概率分布模型。得到概率分布模型后，再由置信度选取置信区间，得到分位数，即可得到对应置信度的百分比误差值。由此可以求出对应的误差上界和误差下界，得到区间分布。

4.2 概率预测结果

4.2.1 误差上界分布的确定

首先计算一年训练集的百分比误差（Mean Percentage Error），利用只对其中正误差进行拟合。以下是其最佳拟合结果，最佳分布为[t]分布。

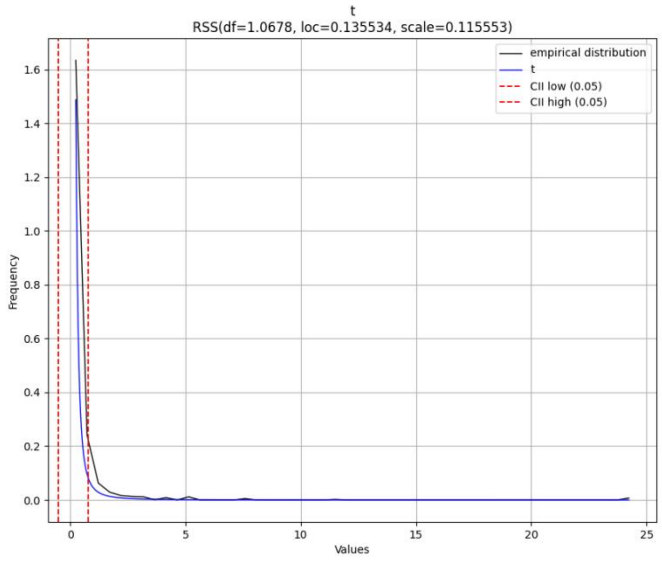

图 4.2.1：误差上界的最佳拟合分布（[t]分布）

得到最佳分布后，很容易根据置信度选取置信区间。由于上界分布只选取了正误差，因此置信区间左端点就是纵轴。根据置信区间求出分位点，就是所对应的误差值。

4.2.2 误差下界分布的确定

首先计算一年训练集的百分比误差，利用只对其中负误差进行拟合（但对其进行了取绝对值处理，所以实际显示正值）。以下是其最佳拟合结果，最佳分布为[gamma]分布。

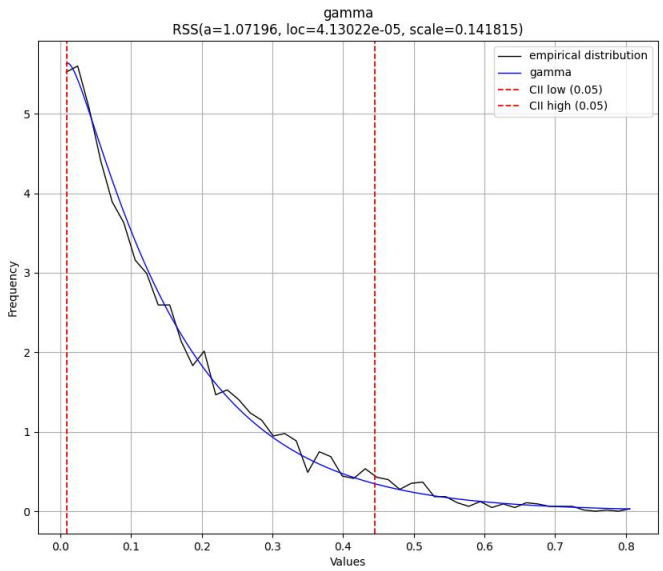

图 4.2.2：误差下界的最佳拟合分布（[gamma]分布）

得到最佳分布后，很容易根据置信度选取置信区间。由于下界分布只选取了负误差，因此置信区间左端点就是纵轴。根据置信区间求出分位点，就是所对应的误差值。

4.2.3 奥地利一周风电的区间概率预测结果（利用 model_RF 的确定性预测结果）

此处只对奥地利一周风电的区间预测结果进行分析，其他预测结果同理。

通过 4.2.1 和 4.2.2 中的 0.9、0.8 等置信度得到的误差上界和误差下界，很容易绘制如下区间预测图。

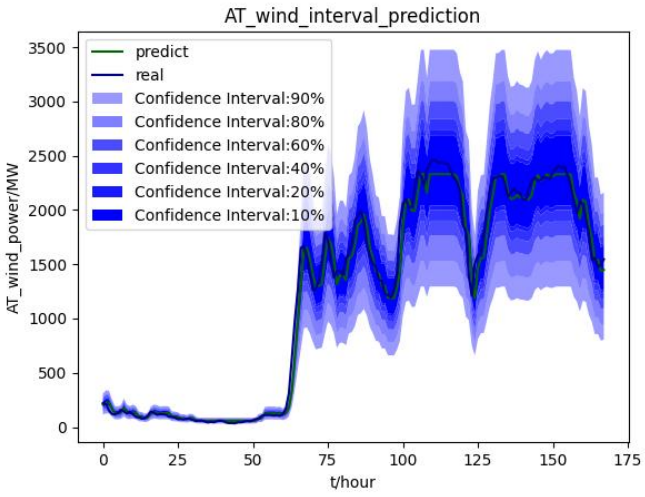

图 4.3.1：奥地利风电区间概率预测结果

从图中可以看出，预测基值越大，其区间预测的范围越宽。预测曲线基本上落在 6 个预测区间的中心，与直观相符。

[参考文献]

[1] 罗澍忻,陆秋瑜,靳冰洁,麻敏华.考虑相关因素的长短时记忆网络短期负荷预测方法[J].机电工程技术,2019,48(12):126-129.
[2] 陆继翔,张琪培,杨志宏,涂孟夫,陆进军,彭晖.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(08):131-137.
[3] Deshmukh M K, Deshmukh S S. Modeling of hybrid renewable energy systems[J]. Renewable and sustainable energy reviews, 2008, 12(1): 235-249.
[4] Deshmukh M K, Deshmukh S S. Modeling of hybrid renewable energy systems[J]. Renewable and sustainable energy reviews, 2008, 12(1): 235-249.
[5] 万灿,宋永华. 新能源电力系统概率预测理论与方法及其应用[J]. 电力系统自动化,2021,45(1):2-16. DOI:10.7500/AEPS20200811008.

附录 任务分工记录表格

组员	主要工作内容
苏博文	Task 2
卢梁宇宸	Task Add
杨旻	Task 1&3