



答辩人：蒋嵩林

指导老师：任超副教授

本科毕业论文答辩

基于 ICEEMDAN 多特征分解和 Prophet-GRU-NN 组合模型多步预测短期风速

本科毕业论文答辩



1 研究背景

2 研究过程

3 总结展望

4 参考文献

5 致谢

6 答疑

多特征组合模型进行风速预测的意义



- (为什么风力发电) 能源变革的需要：
 - ① 传统化石燃料有限、燃烧产生环境问题。
 - ② 风能触手可得、清洁环保，越来越广泛地被应用。
- (为什么风速预测) 风力发电的需求：
 - ① 优化电力分配，提高发电量。
 - ② 预防风力发电机的过载损毁。
 - ③ 安排好风力发电的电网并网问题。
- (为什么组合模型) 单一模型的缺陷：
 - ① 风速数据是非平稳且非线性的。
 - ② 风速数据中往往包含着较大的噪声问题。
 - ③ 数值预报方法求解大气方程，所需计算资源太多。
 - ④ 经典的统计方法如 SARIMA 等，适用于线性关系。
 - ⑤ 传统机器学习方法如 SVR 等，对复杂非线性关系效果不好。
 - ⑥ 直接使用深度学习模型，数据集要求高，收敛性差。
- (为什么多特征) 借鉴数值预报大气动力方程思想，同时预测其他具有强周期性、易预测的气象要素辅助风速的预测。

本科毕业论文答辩



1 研究背景

2 研究过程

3 总结展望

4 参考文献

5 致谢

6 答疑

数据集来源



Nsti 国家科技资源共享服务平台
National Science & Technology Infrastructure



中国区域地面气象要素驱动数据集（1979-2018）

China meteorological forcing dataset (1979-2018)

中国区域地面气象要素驱动数据集，包括近地面气温、近地面气压、近地面空气比湿、近地面全风速、地面向下短波辐射、地面向下长波辐射、地面降水量共7个要素。数据为NETCDF格式，时间分辨率为3小时，水平空间分辨率为0.1°。可为中国区陆面过程模拟提供驱动数据。

该数据集是以国际上现有的Princeton再分析资料、GLDAS资料、GEWEX-SRB辐射资料，以及TRMM降水资料为背景场，融合了中国气象局

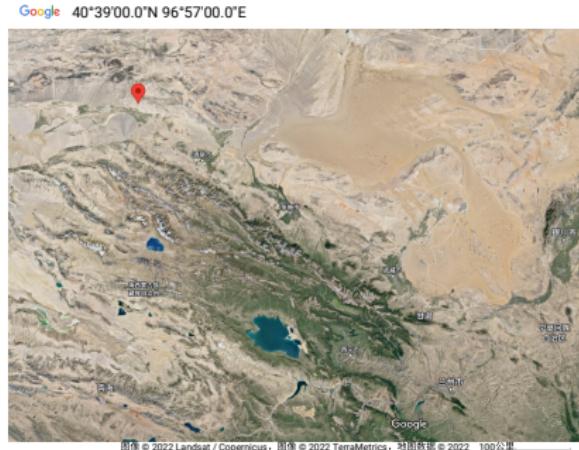
常规气象观测数据制作而成。详细过程请参阅参考文献。原始资料来自于气象局观测数据、再分析资料和卫星遥感数据。已去除非物理范围的值，采用ANU-Spline统计插值。精度介于气象局观测数据和卫星遥感数据之间，好于国际上已有再分析数据的精度。



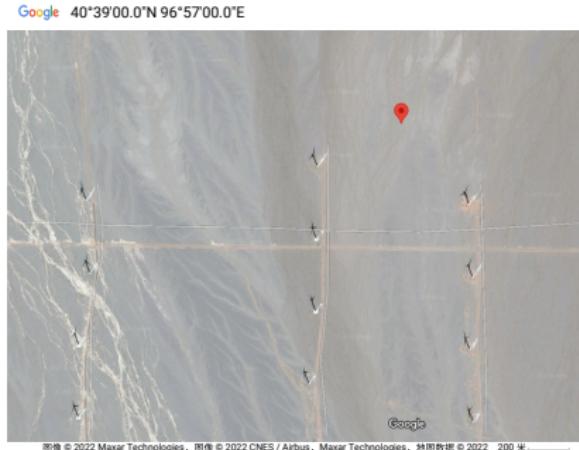
地点选取



甘肃中电酒泉第四风力发电有限公司附近 (40.65° N, 96.95° E)



(a) 甘肃省范围内



(b) 局部范围内

图 2: 选址地坐标使用谷歌地图绘制出的卫星地图

特征选取

表 1: 数据集中所包括特征值的单位以及含义说明^[3]

名称	单位	含义
日期	UTC	YYYY-mm-dd HH:MM:SS
地面向下长波辐射	W/m^2	1.5 小时前开始的 3 小时平均值
地面降水率	mm/h	3 小时前开始的 3 小时平均值
近地面气压	Pa	近地面(距地面 2 米处)瞬时值
近地面空气相对湿度	比值	近地面(距地面 2 米处)瞬时值
地面向下短波辐射	W/m^2	1.5 小时前开始的 3 小时平均值
近地面气温	K	近地面(距地面 2 米处)瞬时值
近地面全风速	m/s	近地面(距地面 2 米处)瞬时值

UTC 2017 年 1 月 1 日 0 时-2018 年 12 月 30 日 21 时 (5832 条)

特征选取



表 2: 风速与其他六种气象要素的斯皮尔曼相关性分析

近地面全风速	相关系数	显著水平
地面向下长波辐射	0.0323507770056**	0.013486168347
地面降水量	0.007876001331	0.5476061054
近地面气压	0.074690860912**	$1.1250462841 \times 10^{-8}$
近地面空气相对湿度	-0.085562076038**	$5.958774799 \times 10^{-11}$
地面向下短波辐射	0.2462649956**	$2.650405346 \times 10^{-81}$
近地面气温	0.056715012962**	$1.4659365394 \times 10^{-5}$

** 表示在 0.01 级别 (双尾), 相关性显著

最终去除掉地面降水量这一特征, 保留剩余 5 种特征。

ICEEMDAN 分解

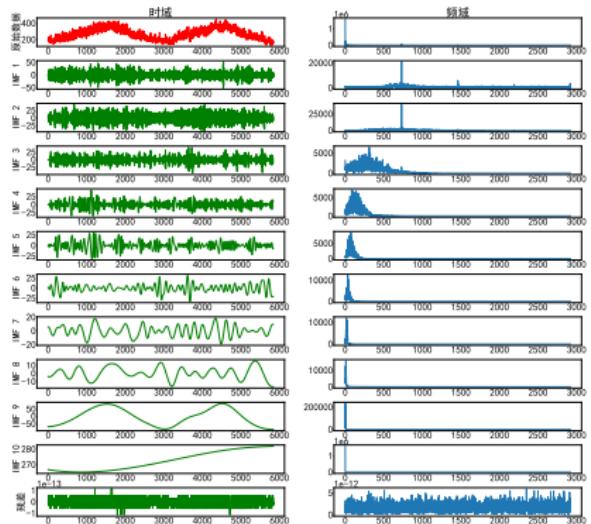


图 3: 地面向下长波辐射

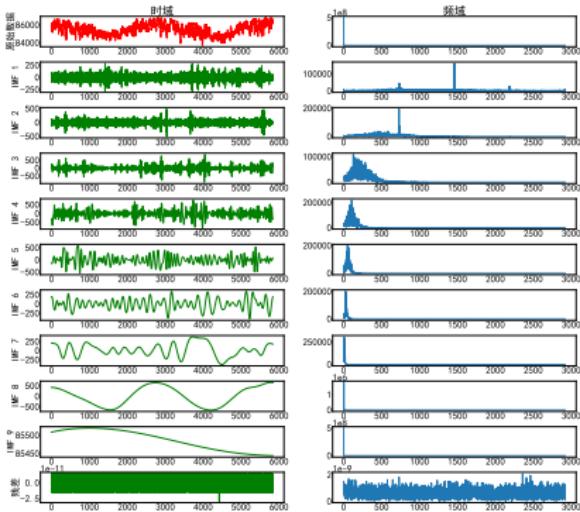


图 4: 近地面气压

ICEEMDAN 分解

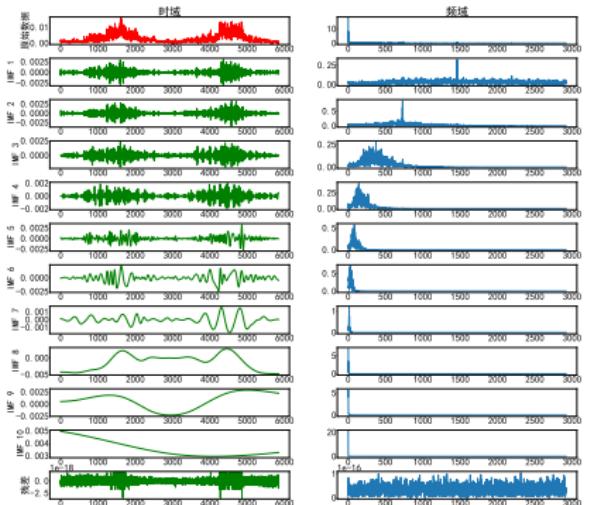


图 5: 近地面空气相对湿度

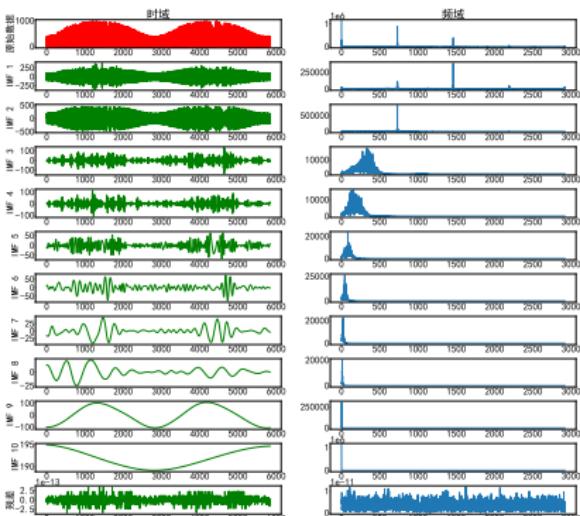


图 6: 地面向下短波辐射

ICEEMDAN 分解

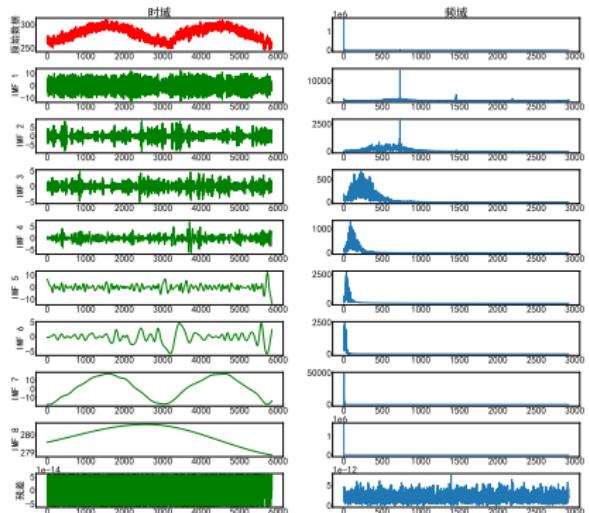


图 7: 近地面气温

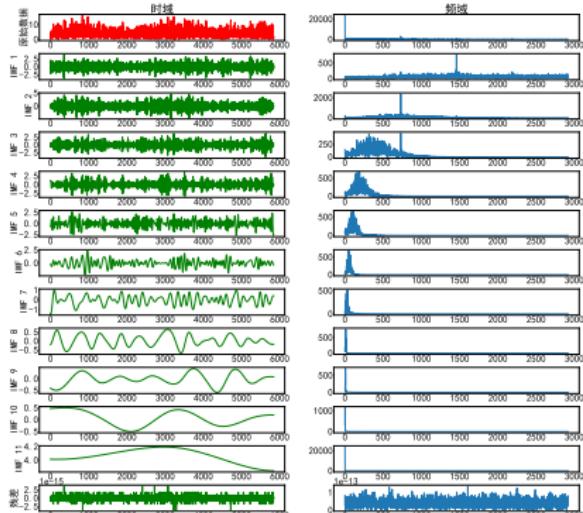


图 8: 近地面全风速

搜寻 Prophet 模型最优超参数



选择前 438 天（3504 条数据）作为初始训练数据。对于后 291 天（2324 条数据），每隔 1 天对未来 12 小时的情况进行一次预测，共计验证 291 次，交叉验证搜寻 Prophet 的最优超参数，从而使得 MAPE 值最小化，得到如表 3 所示最优值。

表 3：风速历史时间序列数据 Prophet 模型的最优超参数

超参数名	值
changepoint_prior_scale	1.0
seasonality_prior_scale	0.1
seasonality_mode	additive
changepoint_range	1

神经网络部分模型训练及验证方法



- 训练时根据 5828 条数据点，以 8 个连续时间序列数据点（1 天）为一个时间窗口制作一个训练样本。每个时间窗口中的数据点整体为一个输入，每个时间窗口中的输出为该输入窗口紧挨着的下一个时间序列数据点（3 小时）的预测值。因而总共得到 5820 个训练样本（时间窗口），每个时间窗口样本都与上一个时间窗口差值为 3 小时。
- 神经网络部分的实现使用 Tensorflow Keras。对于所有 GRU 模型，为统一标准，选择训练 10 个 epoch，NN 模型选择训练 5000 个 epoch。批大小（Batch Size）为默认 32，训练过程中设置回调函数，当损失 5 步内不再下降时自动停止训练过程，并自动保存训练过程中得到的损失最小的模型。损失函数为 Huber，模型优化器为 Nadam，设定学习率为 0.001。
- 进行预测时使用多步预测的方法。每个时间窗口的输入为现在已知的数据点值，混合上一个时间窗口的预测输出值，从而进行四步预测（12 小时）。

风速 ICEEDMAN-GRU 模型

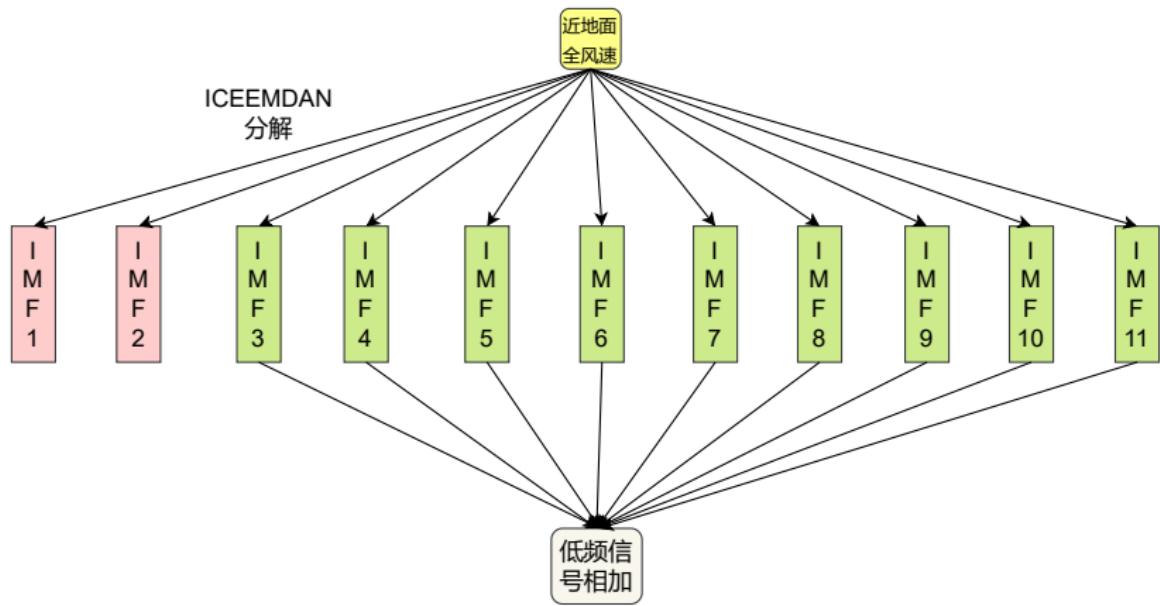


图 9: 风速 ICEEDMAN 分解示意图

风速 ICEEDMAN-GRU 模型

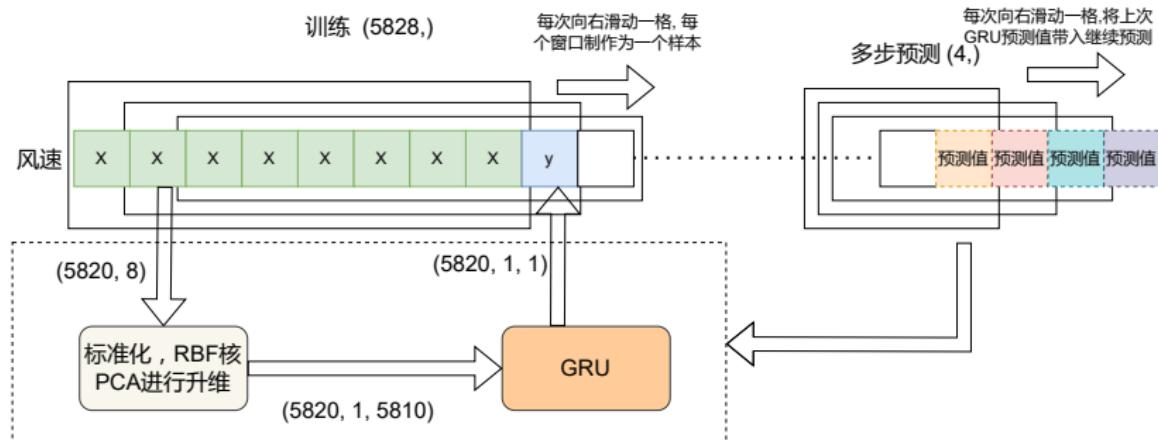


图 10: 风速 GRU 模型结构图

风速 ICEEDMAN-GRU 模型

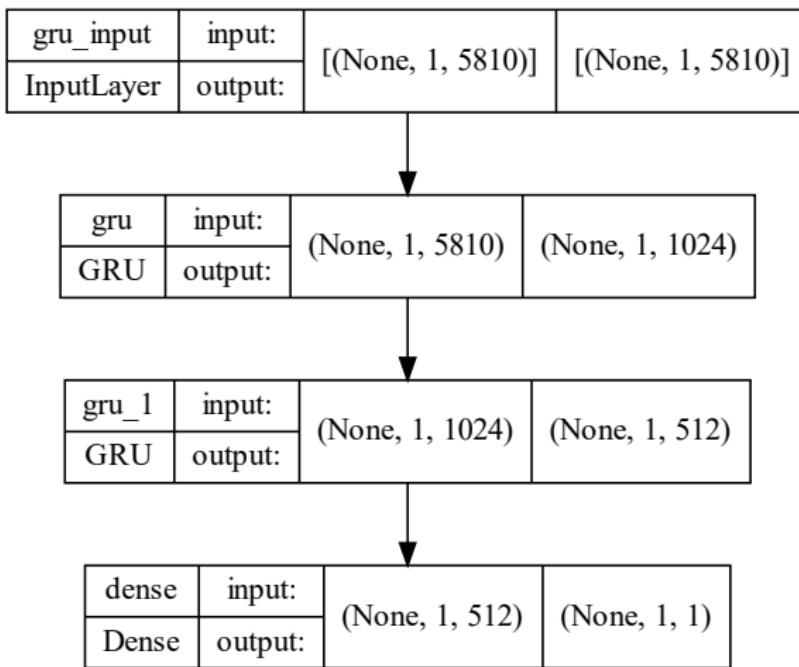


图 11: 本文使用的 GRU 模型网络结构图

风速 ICEEDMAN-Prophet-GRU 模型

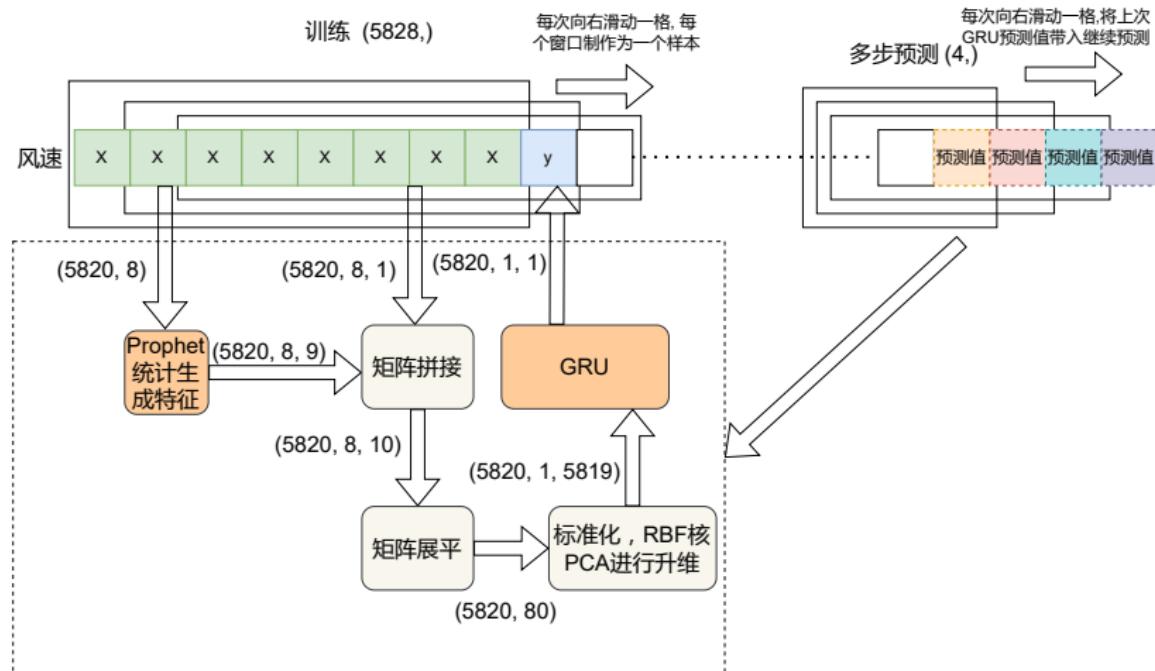


图 12: 风速 Prophet-GRU 模型结构图

多特征 ICEEDMAN-Prophet-GRU-NN 模型

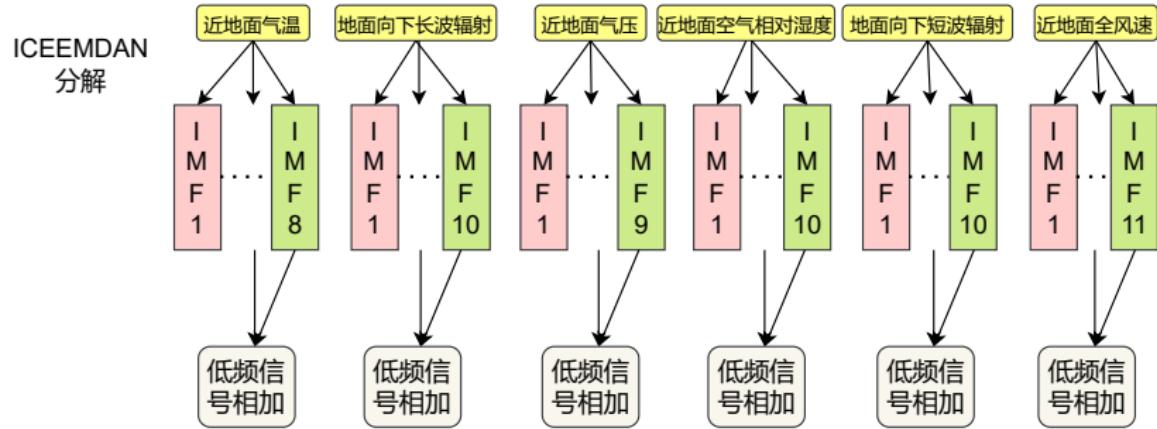


图 13: 多特征 ICEEDMAN 分解示意图

多特征 ICEEDMAN-Prophet-GRU-NN 模型

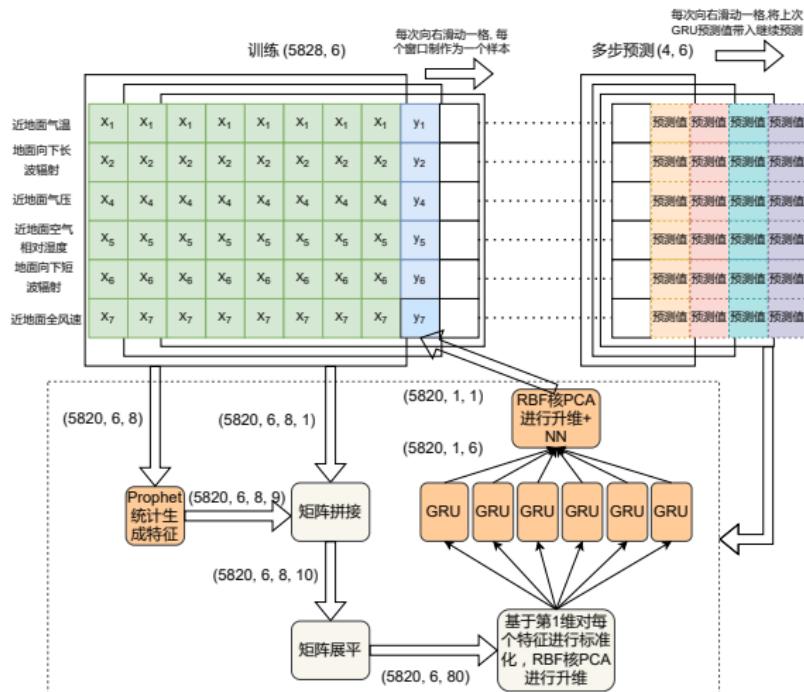


图 14: 多特征 Prophet-GRU-NN 模型结构图

多特征 ICEEDMAN-Prophet-GRU-NN 模型

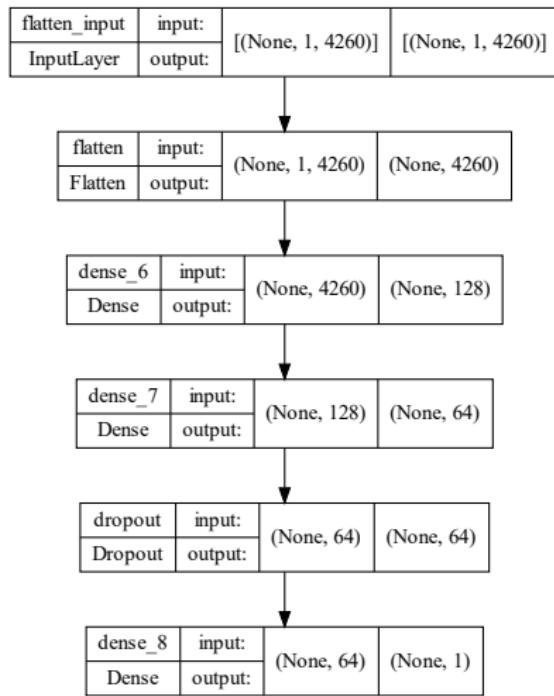


图 15: 多特征 ICEEDMAN-Prophet-GRU-NN 模型中的 NN 网络结构图

模型评价指标



- $MSE = \frac{1}{N} \sum_{t=1}^N (\hat{y}(t) - y(t))^2$
- $MAE = \frac{1}{N} \sum_{t=1}^N |\hat{y}(t) - y(t)|$
- $MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{\hat{y}(t) - y(t)}{y(t)} \right|$
- $RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{y}(t) - y(t))^2}$

上述四种误差指标均为，指标值越小时，模型的精度越高。

- $R^2 = 1 - \frac{\sum_{t=1}^N (\hat{y}(t) - y(t))^2}{\sum_{t=1}^N (\bar{y}(t) - y(t))^2}$

决定系数 (R^2) 取值范围为 $(-\infty, 1]$ ，越接近 1 代表模型的准确度越好。

- 15% 准确率 = 预测值和真实值实际偏差在 15% 以内的样本数/总样本数

模型对比评估



1号：基于风速的 Prophet 模型，2号：基于风速的 ICEEMDAN-Prophet 模型，3号：基于风速的 ICEEMDAN-GRU 模型，4号：风速 ICEEMDAN-Prophet-GRU 模型，5号：多特征 ICEEMDAN-Prophet-GRU-NN 模型。

表 4：五种模型在测试集中四步预测对比评估

编号	MAPE	MAE	MSE	RMSE	R^2	15% 准确度
1	0.7599	1.5475	2.4874	1.5771	-14.4053	0
2	0.7396	1.5054	2.3588	1.5359	-13.6092	0
3	0.1671	0.3	0.2055	0.4533	-0.2728	0.75
4	0.1356	0.286	0.1121	0.3349	0.3051	0.5
5	0.0825	0.1734	0.0307	0.1751	0.81	1

模型对比评估



采用搜寻得到的 Prophet 模型最优超参数，分析历史 5828 条数据
从而直接预测出未来 12 小时，即 4 项风速值。

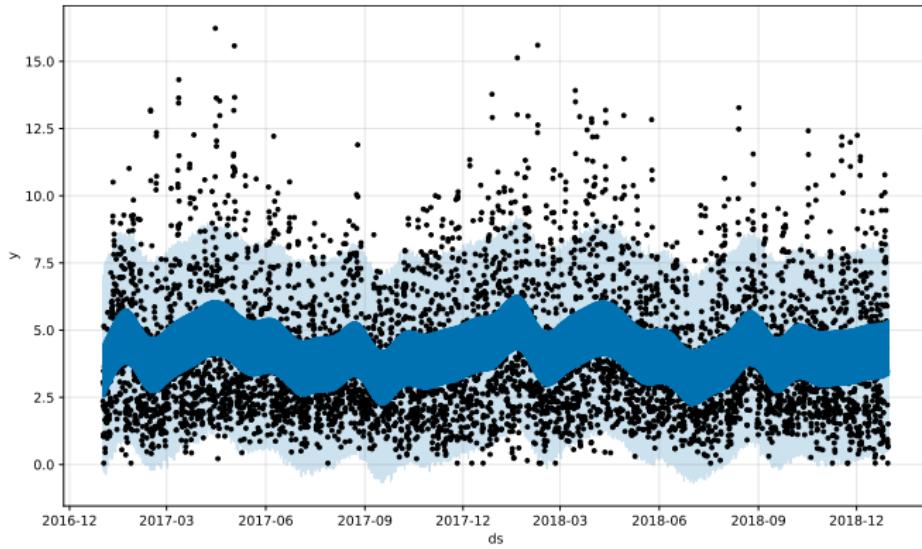


图 16: 基于风速历史的 Prophet 模型可视化图

模型对比评估

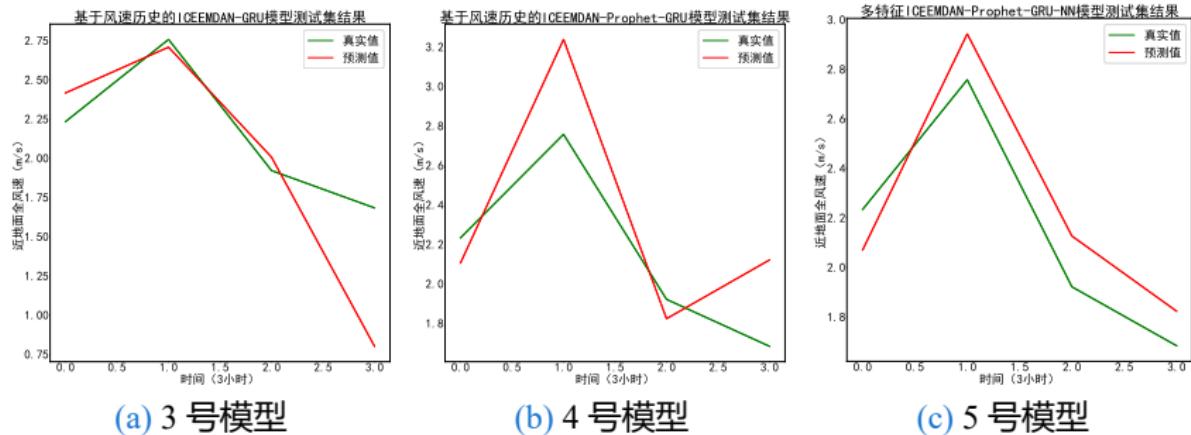


图 17: 三种基于本文提出的神经网络的模型四步预测风速的可视化图

本科毕业论文答辩



1 研究背景

2 研究过程

3 总结展望

4 参考文献

5 致谢

6 答疑

创新点



- ① 从数值预报、大气动力学获得灵感，多气象要素联合预测。
- ② 使用最新信号分解技术 ICEEMDAN (2014 年) 来特征分解，采用 Facebook 推出的新型时间序列框架 Prophet (2018 年) 与最新循环神经网络 GRU 模型 (2014 年) 组合预测。
- ③ 从支持向量回归机获得灵感，使用了 RBF 核 PCA 升维。
- ④ 将 GELU 激活函数 (2016 年) 以及 Nadam 优化器 (2016 年) 和 Huber 损失函数应用到短期风速预测领域。
- ⑤ 选择甘肃酒泉一个真实风力发电厂附近气象数据验证模型。

展望未来



因为论文时间紧迫，展望未来，本次论文写作过程也存在着许多遗憾的点，如果有机会的话可以在后续进一步研究过程中完成：

- 本次建模过程只选择了两年跨度的时间序列数据。
- 只对风速原始时间序列数据这一个模型进行了搜索最优超参数的过程。
- 本文只选择了甘肃酒泉的一个真实风力发电厂附近的地点进行短期风速的预测。
- 没有探究模型中所用神经网络层数进一步加深的情况。
- 可尝试更多各种不同新型模型，用各种不同方法组合起来。

本科毕业论文答辩



1 研究背景

2 研究过程

3 总结展望

4 参考文献

5 致谢

6 答疑

参考文献



-  阳坤, 何杰. 中国区域地面气象要素驱动数据集 (1979-2018) [J]. 国家青藏高原科学数据中心, 2019.
-  Yang J K. He, Tang W, Qin J, et al. On downward shortwave and longwave radiations over high altitude regions: Observation and modeling in the Tibetan Plateau[J]. 2010, 150(1):38–46.
-  He J, Yang K, Tang W, et al. The first high-resolution meteorological forcing dataset for land process studies over China[J]. *Scientific Data*, 7:25.
-  Bengio Y, Boulanger-Lewandowski N, Pascanu R. Advances in optimizing recurrent networks[C]. 2013. 8624–8628.

本科毕业论文答辩



1 研究背景

2 研究过程

3 总结展望

4 参考文献

5 致谢

6 答疑

致谢



- 首先感谢各位在场老师对我的毕业论文答辩的聆听！
- 其次感谢我的毕业论文导师任超副教授的辛勤指导！
- 感谢兰州大学信息科学与工程学院老师们提供的优质课程！
- 另外我还要感谢 2018 级计算机科学与技术基础理论班的各位同学，以及大气科学专业的舍友，和其他同学四年间朝夕相处的鼓励支持！
- 最后，感谢大学四年父母对我的默默关爱和支持！

Q&A

敬请各位在场评委老师们批评指正！

本科毕业论文答辩



1 研究背景

2 研究过程

3 总结展望

4 参考文献

5 致谢

6 答疑

ICEEMDAN 分解



设 x 为待分解信号, $E()$ 表示由 EMD 分解产生的本征模函数 (IMF) 分量, $N()$ 表示产生信号进行 EMD 分解出 IMF 分量后系统平均, $w^{(i)}$ 代表添加的 i 组随机高斯白噪声。则 ICEEMDAN 分解步骤如下:

- ① 向 x 中分别添加 i 组白噪声, 即 $x^{(i)} = x + \beta_0 E(w^{(i)})$ 。
- ② 对 $x^{(i)}$ 进行计算, 得到第一个 IMF 分量 $IMF_1 = N(x^{(i)})$ 。
- ③ 计算残差 $R_1 = x - IMF_1$ 作为新待分解序列, 重复上述步骤直到不能进行 EMD 分解。

ICEEMDAN 和传统基于 EMD 算法的优势:

- 相对于 EMD, 添加一定的白噪声可以平滑极值点的分布, 从而让分解的结果更加健壮。
- 相较于 EEMD, 无需计算过多均值, 因而分解的性能更高。
- 相较于 CEEMDAN, 降低了出现多个无意义的低幅低频伪 IMF 分量的概率。

Prophet 时间序列预测框架



Prophet 类似于 SARIMA，基于公式： $y(t) = g(t) + s(t) + h(t) + \epsilon_t$

- $g(t)$ 表示的是趋势分量，代表时间序列数据在非周期性规律中的变化趋势。
- $s(t)$ 表示周和日、年的季节性分量。
- $h(t)$ 表示节假日项，此处和风速预测无关。
- ϵ_t 为残差，表示噪声分量，其服从高斯白噪声分布。

Prophet 的预测速度快，提供了完全自动化的时间序列预测功能，对异常值和时间序列中的剧烈变化也都具有十分强大的健壮性。

核主成分分析



假设待降维的原始数据有 m 条，每条数据为 n 维，组成了 n 行 m 列的矩阵 X ，则主成分分析 (PCA)：

- ① 将 X 的每一行进行零均值化（减去这一行的均值）。
- ② 求解协方差矩阵 $C = \frac{1}{m}XX^T$ ，以及特征值及对应的特征向量。
- ③ 将协方差矩阵 C 的特征向量按对应特征值大小从前到后按行排列成矩阵，取前 k 行组成新矩阵 P 。
- ④ $Y = PX$ 即为降维到 k 维后的数据。

主成分分析 (PCA) 只能用于线性情况，不适用于非线性的风速。

核主成分分析 (KPCA) 通过核函数把原始非线性的数据映射到高维空间，变成线性的后，用 PCA 来处理映射后的高维数据。

基于高斯径向基函数的核主成分分析



高斯径向基函数 (RBF) 表达式: $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}$

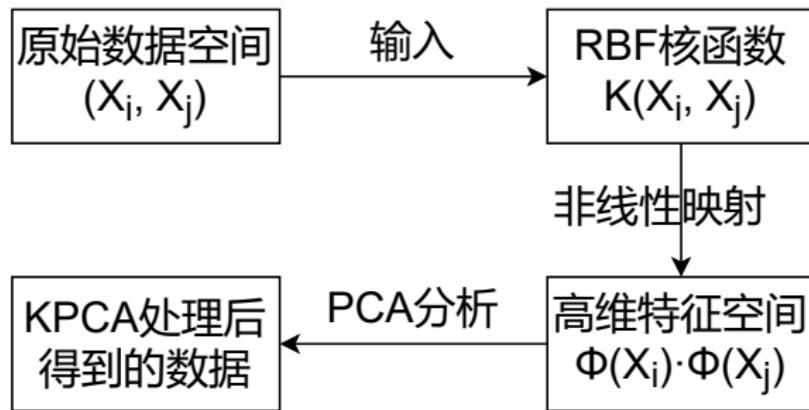


图 18: 基于 RBF 的核主成分分析算法工作流程示意图

经典循环神经网络

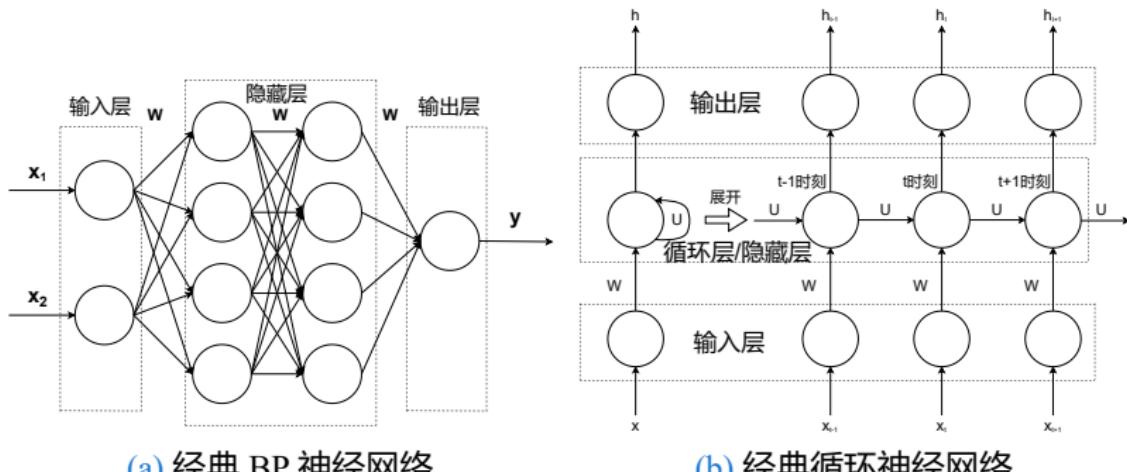
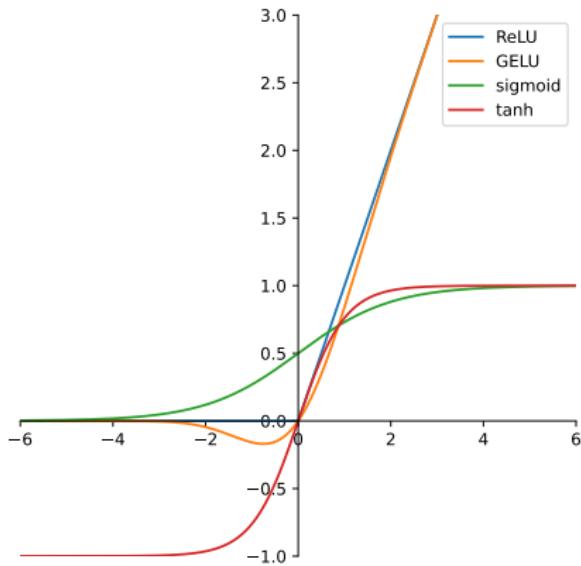


图 19: 两种经典神经网络结构示意图

经典 RNN 中每个神经元: $h_t = \sigma(Wx_t + Uh_{t-1} + b)$

GELU 激活函数



- $ReLU(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases}$
- $GELU(x) = x\Phi(x)$
- $sigmoid(x) = \frac{1}{1+e^{-x}}$
- $tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

图 20: GELU 和其他激活函数的对比

Nadam 优化器



- NAG 算法能够预知未来的更新方向，从而使得更新更加稳定而不至于一直遵循梯度更新的惯性。这种预期性的更新可以防止梯度更新得太快，从而提高响应能力。根据已有的研究成果，NAG 显著提高了 RNN 在许多训练任务中的性能^[4]。
- Adam 是目前最为广泛被使用的一种优化器，能够自动化调整学习率，对与频繁出现的特征相关的参数执行较小的更新（降低学习率）。对与不常见特征相关的参数执行较大的更新（提高学习率），同时还能适应稀疏数据，克服学习率急剧下降的问题。
- Nadam 向 Adam 优化器中融合了 NAG 的思想，添加了 Nesterov 动量，从而使其获得了 NAG 的能够预知未来的更新方向的优点，提高其在 RNN 训练任务中的性能，最终使用 Nadam 能够取得比 Adam 更好的效果。

Huber 损失函数

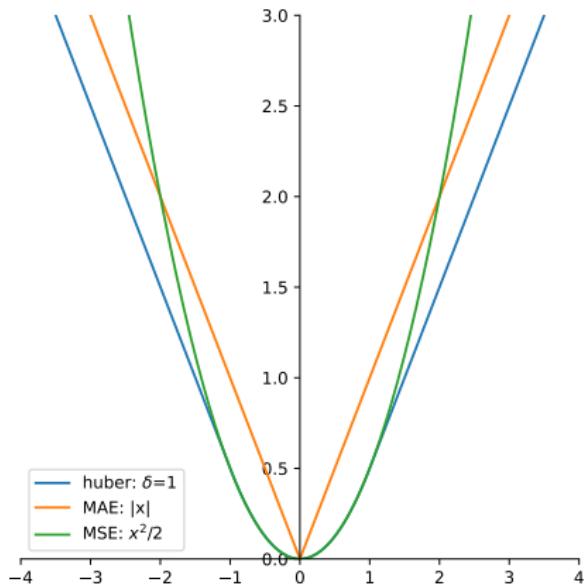


图 21: Huber 和其他损失函数的对比

- *Huber*:
$$\begin{cases} \frac{1}{2}x^2, & |x| < \delta \\ (x - \frac{1}{2}\delta)\delta, & |x| \geq \delta \end{cases}$$
- 当损失值较大时, 由于其是类 MAE 线性损失函数, 梯度始终保持不变, 因而将异常点和正常点同等看待, 克服了 MSE 的缺点并获得了 MAE 的优点。
- 当损失值较小时, 由于其是类 MSE 平方损失函数, 梯度也减小, 从而利于函数的收敛, 克服了 MAE 的缺点并获得了 MSE 的优点。



经典 RNN 在训练时由 sigmoid 激活函数的性质可知，当值较大时，函数趋于 1，并变得平缓，此时进行反向传播会造成整体值过小，造成梯度消失的问题。另外，经典 RNN 的训练过程中每个神经元权重都是直接记忆学习临近时刻的，因而对存在的季节性效应的长期关系学习不佳。

GRU 计算公式：

- 更新门： $z_t = \sigma(W^{(z)}x_t + U^{(z)}h_{t-1})$
- 重置门： $r_t = \sigma(W^{(r)}x_t + U^{(r)}h_{t-1})$
- 确定遗忘掉多少过去的信息： $h'_t = \tanh(Wx_t + r_t \odot Uh_{t-1})$
- 确定当前步记忆多少信息： $h_t = z_t \odot r_t + (1 - z_t) \odot h'_t$