WRF数据：

2013-2017年数据为6h输出一次，文件名和实际输出的预报时间为：

例如：

文件名为2013040900-2013041612，输出次数7\*4+3 =31

实际输出2013040912-2013041412，输出次数2+4\*4+3=21

2018-2019年数据为1h输出一次，文件名和实际输出的预报时间为：

例如：

文件名为2018010100-2018010812，输出次数7\*24+13 =181

实际输出2018010100-2018010612，输出次数5\*24+13=133

WRF数据最终将一年的数据压缩到一个pkl文件中，其中每一天的数据维度如下：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SpotTime | PredTime  截止时间 | XLONG | XLAT | U10 | V10 |
| 2013040900 | 2013041400 | 21\*299\*549 | 21\*299\*549 | 21\*299\*549 | 21\*299\*549 |

训练集：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Station\_Id | XLONG | XLAT | SpotTime | PredTime | Direction\_x | Speed\_x |
| 台站号 | 台站经度 | 纬度 | 起报时间 | 预报时间 | 预报风向 | 预报风速 |
| SeaPressure | StaPressure | P3 | Temp | DPT | Direction\_y | Speed\_y |
| 海平面气压 | 台站气压 | 三小时变压 | 气温 | 露点温度 | 观测风向 | 观测风速 |

海平面气压订正=本站气压+测站海拔高度的假想气柱的重量

SeaPressure = StaPressure×10^[h/18400(1+ Temp /273)]，h为高度。

风向、风速与U10、V10之间的转化（观测数据为风向风速）：

风向是按正北方向起算的，0度表示北风。90度表示东风。

u为正，表示西风，从西边吹来的风。v为正，表示南风，从南边从来的风。假如u为1，v为1，则表示西南风。

def compute(u, v):  
 if u > 0 and v > 0:  
 fx = 270 - math.atan(v / u) \* 180 / math.pi  
 elif u < 0 and v > 0:  
 fx = 90 - math.atan(v / u) \* 180 / math.pi  
 elif u < 0 and v < 0:  
 fx = 90 - math.atan(v / u) \* 180 / math.pi  
 elif u > 0 and v < 0:  
 fx = 270 - math.atan(v / u) \* 180 / math.pi  
 elif u == 0 and v > 0:  
 fx = 180  
 elif u == 0 and v < 0:  
 fx = 0  
 elif u > 0 and v == 0:  
 fx = 270  
 elif u < 0 and v == 0:  
 fx = 90  
 elif u == 0 and v == 0:  
 fx = 999.9  
 elif pd.isnull(u) and pd.isnull(v):  
 fx = np.nan  
 fs = math.sqrt(math.pow(u, 2) + math.pow(v, 2))  
 return fx, fs

台站预报数据插值（距离权重插值）：

选取距离站点最近的四个网格点作为数据源，根据距离计算权重，最终得到插值结果。

格点数量548\*251=137548，（经度范围28.775745-131.22426，纬度范围-21.052437-32.30636）

台站数量67个

GTS观测数据，数据中存在一些无法识别的编码，最终使用windows-1252编码识别。

存在同一时间、统一观测站存在两个数据，目前没有进行处理。