2022-2023学年空间天气预报期末报告

姓名 王彪 学号 BX2215901

**一、数据处理**

1）经对数据文件处理后，我共获得 10079 个数据点

2）对该数据绘图得到结果如图1所示

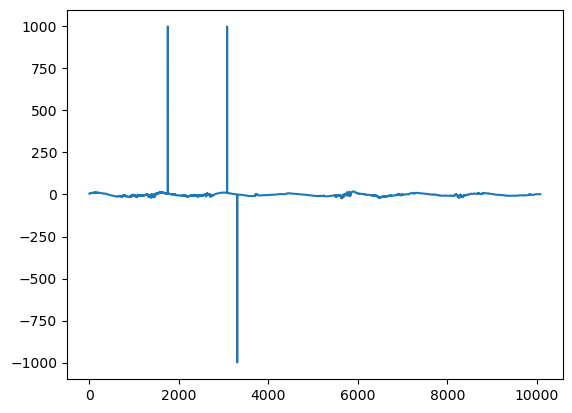


图1 原始数据图示

3）经过观察原始数据，发现 3 个数据异常点。

4）采用 基于标准差的异常值处理，计算该数据点与均值之间的距离除以标准差，将标准差较大的点用前后插值代替 方法消除这些异常数据。代码如下：

def remove\_outliers\_and\_interpolate(*data*, *threshold*=3):

    mean = np.mean(data)

    std = np.std(data)

    z\_scores = [(x - mean) / std *for* x *in* data]

    filtered\_data = [x *if* abs(z) < threshold *else* 0 *for* x, z *in* zip(data, z\_scores)]

*return* np.interp(np.arange(len(filtered\_data)), np.flatnonzero(filtered\_data), np.array(filtered\_data)[np.flatnonzero(filtered\_data)])

图表

描述已自动生成

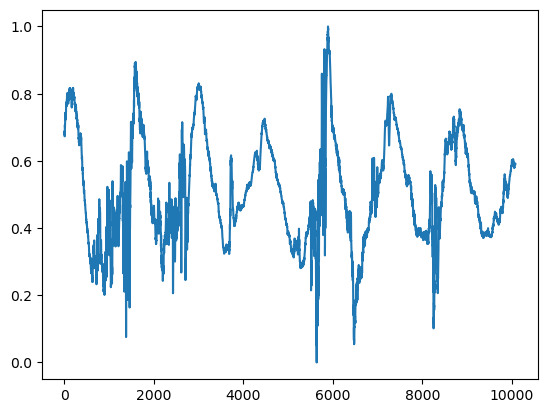
异常处理后数据图示

5）采用 最小-最大归一化 方法将数据归一化到[0,1] 区间，代码如下：

数据归一化代码：

min\_max\_scaler = preprocessing.MinMaxScaler(*feature\_range*=(0,1))

train = min\_max\_scaler.fit\_transform(train)



归一后数据

6）为了建立训练集，采用 前100个数据 作为输入， 后10个数据 作为输出，共建立 9970 组输入输出单元，在它们当中采用 前9000 组作为训练数据，

后970 组作为验证数据。代码如下：

训练数组构建代码：

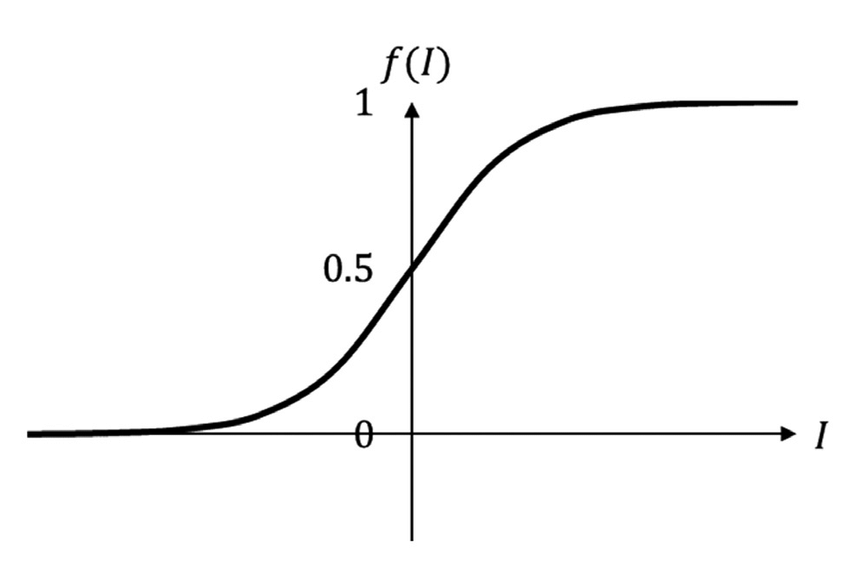
model.fit(new\_X\_train[0:9000], new\_Y\_train[0:9000], *epochs*=1000, *batch\_size*=128)

**二、神经网络构建与训练**

1）建立了 1 层 LSTM 类型神经网络，每层神经元个数分别为 100 个。

2）该神经网络训练函数和激励函数选择如图2所示

训练函数为：均方误差作为损失函数



Sigmoid激励函数

图2神经网络训练函数和激励函数

3）对神经网络训练后，训练界面和误差变化分别如图3和图4所示

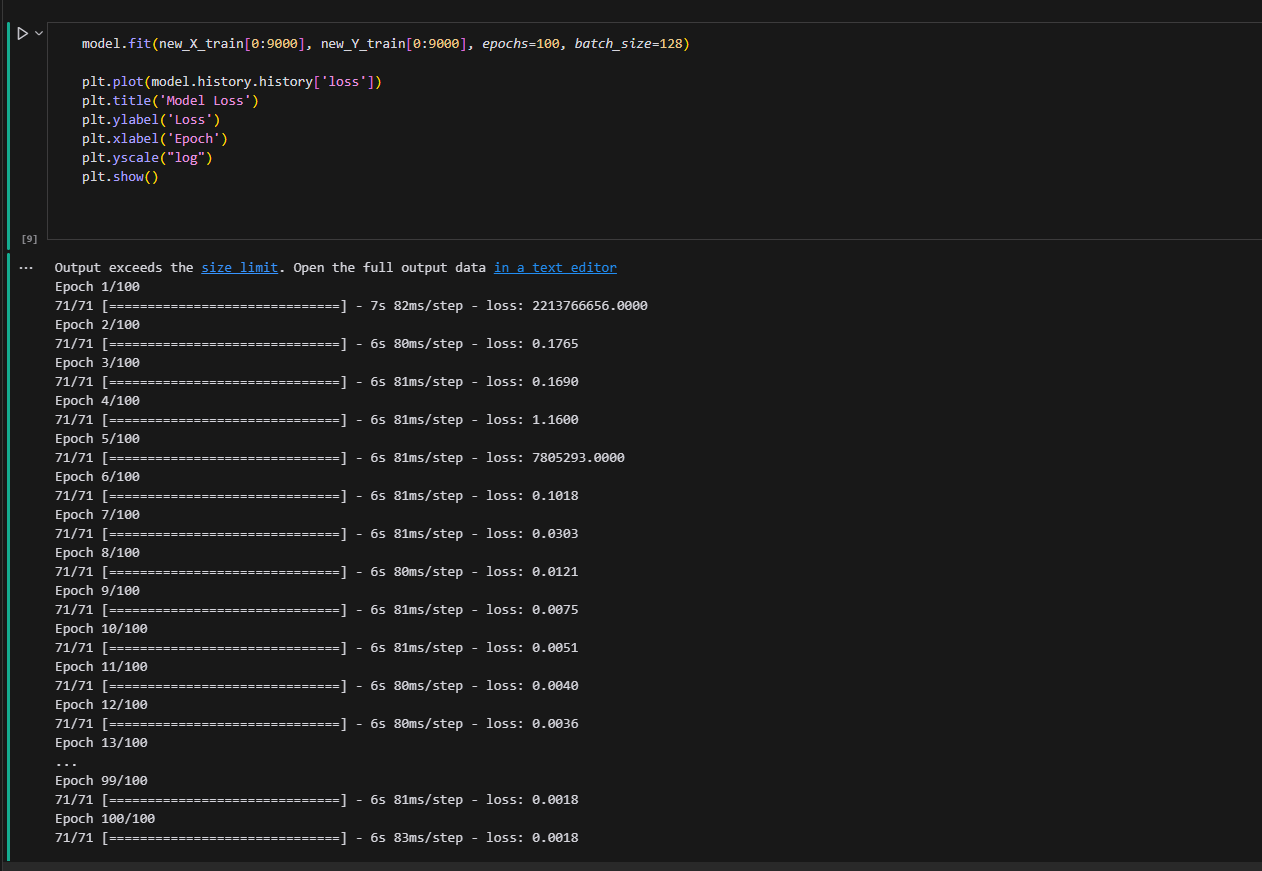


图3 训练界面

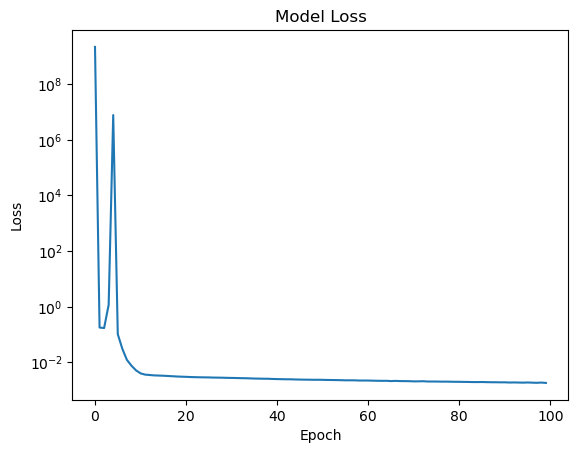


图4 训练误差结果

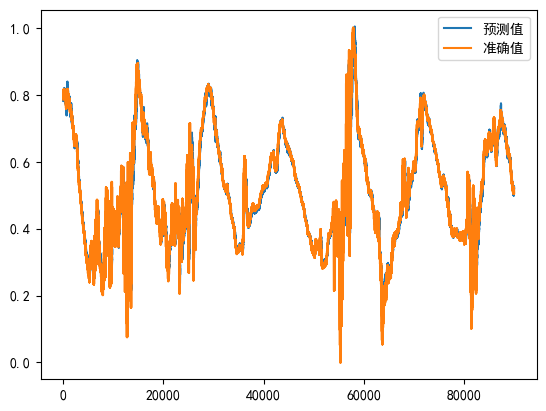


图5 神经网络输出与训练数据目标输出比较

4）采用训练好的神经网络，预测验证集的结果，预测结果与实际结果比较如图6所示。

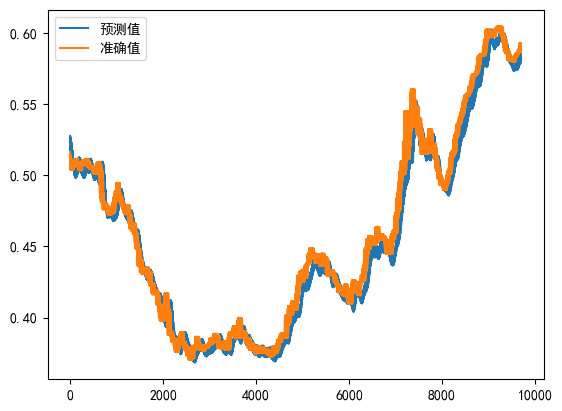


图6 预测结果与验证结果比较

**三、预报结果**

1）根据神经网络训练时的输入和输出关系，一次可以预报 10 个点；

2）为了预报一天的数据，采用了 for循环递归 方法，代码如下：

预报代码：

Y\_=[]

train2=train.reshape(-1)

*for* i *in* np.arange(144):

    X\_=train2[-100:].copy()

    X\_=X\_.reshape(1,-1,1)

    outputs = model(X\_)

    outputs=np.array(outputs).reshape(-1)

    train2=np.concatenate((train2,outputs),*axis*=0)

    print(train2.shape)

    Y\_=np.concatenate((Y\_,outputs),*axis*=0)

3）经过逆归一化后，最终对未来一天的预报结果如图7所示。

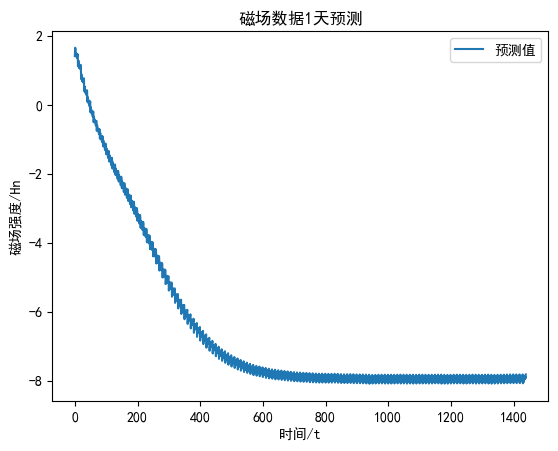


图7 未来一天预报结果

不太清楚是什么卫星数据，找不到预测当天的观测值。且由于误差不断叠加，越往后，预测越不准。

图8 随时间变化的观测数据（蓝色）与预报数据（红色）