

1.问题重述

1.1 问题的背景

近年来，随着工业化与城市化进程的不断加快，空气污染问题也越发严重。空气污染会对公众的日常生活造成负面影响，甚至会引发一系列健康问题[1-3]。开展环境空气质量预报工作是保障及时妥善应对重污染天气的重要技术手段，对区域大气污染联合减排也具有指导意义。

现有的空气质量预报方法主要有数值分析法 与统计分析法。然而，数值预报法通常需要准确的 输入数据和昂贵的计算资源来进行空气质量预报；而统计预报法对于非线性变化的污染物浓度预报 准确性较低[4]。目前通过人工智能、机器学习等方法 实现环境空气质量预报已成为各国环保领域的研 究热点和发展趋势[5-6]。

1.2 问题的重述

问题一：使用监测点 A 空气质量预报基础数据，按照附录中的方法计算监测点 A 从 2020 年 8 月 25 日到 8 月 28 日每天实测的 AQI 和首要污染物。

问题二：在污染物排放情况不变的条件下，某一地区的气象条件有利于污染物扩散或沉降时，该地区的 AQI 会下降，反之会上升。根据对污染物浓度的影响程度，对气象条件进行合理分类，并阐述各类气象条件的特征。

问题三：使用监测点 A 空气质量预报基础数据和监测点 B、C 空气质量预报基础数据，建立一个同时适用于 A、B、C 三个监测点（监测点两两间直线距离 $>100\text{km}$ ，忽略相互影响）的二次预报数学模型，用来预测未来三天 6 种常规污染物单日浓度值，要求二次预报模型预测结果中 AQI 预报值的最大相对误差应尽量小，且首要污染物预测准确度尽量高。并使用该模型预测监测点 A、B、C 在 2021 年 7 月 13 日至 7 月 15 日 6 种常规污染物的单日浓度值，计算相应的 AQI 和首要污染物。

问题四：相邻区域的污染物浓度往往具有一定的相关性，区域协同预报可能会提升空气质量预报的准确度。监测点 A 的临近区域内存在监测点 A1、A2、A3，使用附件 1、3 中的数据，建立包含 A、A1、A2、A3 四个监测点的协同预报模型，要求二次模型预测结果中 AQI 预报值的最大相对误差应尽量小，且首要污染物预

测准确度尽量高。使用该模型预测监测点 A、A1、A2、A3 在 2021 年 7 月 13 日至 7 月 15 日 6 种常规污染物的单日浓度值，计算相应的 AQI 和首要污染物。并讨论：与问题 3 的模型相比，协同预报模型能否提升针对监测点 A 的污染物浓度预报准确度。

2. 问题分析

问题一：按照附录的计算规则，带入数据进行计算，选择部分代表性数据放在正文中即可，其它的部分已经放在附录里。

问题二：根据问题一计算得到的 AQI 数据以及相关的污染物数据进行无监督聚类，无监督聚类模型有很多，如层次聚类、高斯混合聚类等，采用 SOM 自组织神经网络聚类算法，将原始数据输入网络，后能够自动根据各类数据的特点在不同的步数下生成不同的结果。

问题三：首先分析题目已知数据包括了各监测点逐小时污染物浓度和气象一次预报数据 以及实测的污染物浓度和气象数据等；利用实测数据对预报数据进行误差修正，通过前期的预测数据和实测数据的差，找到相关的误差修正规律；采用神经网络模型，设置一个三层的网络机构，输入层数据是一次预报的气象条件，而标准输出数据为真实污染物浓度与预测污染物浓度的差值，这样就建立了预测气象条件与实际污染物浓度误差之间的关系；使用 基于遗传算法优化的神经网络模型，相对于传统的 BP 神经网络而言，其精度将会更高。得到上述网络关系后，若新得到一组一次预报气象数据结合相关的误差变量进行二次修正。

问题四：区域协同预报主要是为了防止某个点的预测数据存在误差的情况下，可通过其它点的预测数据进行修正(修正的前提是利用其它点得到的相关数据与真实数据存在较大的误差，则此时会进行修正)因此可首先基于真实数据建立 A、A1、A2、A3 四个站点间与位置、污染物浓度数据有关的拟合模型，然后在将该拟合模型之间的关系带入到一次预报数据中，若一次预报数据与拟合得到的结果差距较大则将其替代（可设置一定的误差界限）若一致则保留，此时在利用问题三构建的模型进行计算，看与真实值的误差是否减小，若减小则表示区域协同预报可以提高准确度，若并未减小则表示效果不好。

3.参考文献

- [1] CHEN F L, CHEN Z F. Cost of economic growth: air pollution and health expenditure[J]. Science of the Total Environment, 2021, 755: 142543.
- [2] LIU W L, XU Z P, YANG T. Health effects of air pollution in China[J]. International Journal of Environmental Research and Public Health, 2018, 15 (7) : 1471.
- [3] 张雨梦, 钱鹏, 查书平. 南京市一次大气污染事件时空演化特征及影响因素[J]. 南通大学学报(自然科学版), 2018, 17 (4) : 48-55.
- [4] 邱敬怡, 赵璇. 基于SVR-BP算法的江苏省空气质量指数预测[J]. 南通大学学报(自然科学版), 2020, 19 (1) : 42-47.
- [5] 许治国. 利用Keras构建神经网络在空气质量预测中的应用[J]. 环境监控与预警, 2018, 10 (5) : 18-21.
- [6] RYBACZYK Y, ZALAKEVICIUTE R. Machine learning approaches for outdoor air quality modelling: a systematic review[J]. Applied Sciences, 2018, 8 (12) : 2570.

4.附录

```
%二维自组织特征映射网络设计  
%输入数据为各类实测污染物数据  
clc  
clear  
close all %-----  
%随机生成 100 个二维向量，作为样本，并绘制出其分布  
P=[此处填写污染物数据]  
  
%建立网络，得到初始权值  
net=newsom([0 1;0 1],[5 6]);  
w1_init=net.iw{1,1};  
%-----  
%绘制出初始权值分布图  
figure(2);  
plotsom(w1_init,net.layers{1}.distances)  
%-----  
%分别对不同的步长，训练网络，绘制出相应的权值分布图  
for i=10:30:100  
net.trainParam.epochs=i;  
net=train(net,P);  
figure(3);  
plotsom(net.iw{1,1},net.layers{1}.distances)  
end  
  
%程序一：GA 训练 BP 权值的主函数  
function net=GABPNET(XX,YY)  
% GABPNET.m  
% 使用遗传算法对 BP 网络权值阈值进行优化，再用 BP 算法训练网络  
%数据归一化预处理  
nntwarn off
```

```

XX=[1:19;2:20;3:21;4:22];
YY=[1:4];
XX=premnmx(XX);
YY=premnmx(YY);
YY

%创建网络
net=newff(minmax(XX), [19, 25, 1], {'tansig','tansig','purelin'}, 'trainlm');

%下面使用遗传算法对网络进行优化
P=XX;
T=YY;
R=size(P, 1);
S2=size(T, 1);
S1=25;%隐含层节点数
S=R*S1+S1*S2+S1+S2;%遗传算法编码长度
aa=ones(S, 1)*[-1, 1];
popu=50;%种群规模
save data2 XX YY % 是将 xx,yy 二个变数的数值存入 data2 这个
MAT-file,
initPpp=initializega(popu, aa, 'gabpEval');%初始化种群
gen=100;%遗传代数
%下面调用 gaot 工具箱, 其中目标函数定义为 gabpEval
[x, endPop, bPop, trace]=ga(aa, 'gabpEval', [], initPpp, [1e-6
1], 'maxGenTerm', gen, ...
'normGeomSelect', [0.09], ['arithXover'], [2], 'nonUnifMutation', [2
gen 3]);
%绘收敛曲线图
figure(1)
plot(trace(:, 1), 1./trace(:, 3), 'r-');
hold on
plot(trace(:, 1), 1./trace(:, 2), 'b-');

```

```

xlabel('Generation');
ylabel('Sum-Squared Error');
figure(2)
plot(trace(:, 1), trace(:, 3), 'r-');
hold on
plot(trace(:, 1), trace(:, 2), 'b-');
xlabel('Generation');
ylabel('Fitness');

%下面将初步得到的权值矩阵赋给尚未开始训练的 BP 网络
[W1, B1, W2, B2, P, T, A1, A2, SE, val]=gadecod(x);
net.LW{2, 1}=W1;
net.LW{3, 2}=W2;
net.b{2, 1}=B1;
net.b{3, 1}=B2;
XX=P;
YY=T;

%设置训练参数
net.trainParam.show=1;
net.trainParam.lr=1;
net.trainParam.epochs=50;
net.trainParam.goal=0.001;

%训练网络
net=train(net, XX, YY);

%程序二：适应值函数
function [sol, val] = gabpEval(sol, options)
% val - the fitness of this individual
% sol - the individual, returned to allow for Lamarckian evolution
% options - [current_generation]
load data2
nnwarn off
XX=premnmx(XX);

```

```

YY=premnmx (YY) ;
P=XX;
T=YY;
R=size (P, 1) ;
S2=size (T, 1) ;
S1=25;%隐含层节点数
S=R*S1+S1*S2+S1+S2;%遗传算法编码长度
for i=1:S, x(i)=sol(i);
end;
[W1, B1, W2, B2, P, T, A1, A2, SE, val]=gadecod(x);
%程序三：编解码函数
function [W1, B1, W2, B2, P, T, A1, A2, SE, val]=gadecod(x)
load data2
nnwarn off
XX=premnmx (XX) ;
YY=premnmx (YY) ;
P=XX;
T=YY;
R=size (P, 1) ;
S2=size (T, 1) ;
S1=25;%隐含层节点数
S=R*S1+S1*S2+S1+S2;%遗传算法编码长度
% 前 R*S1 个编码为 W1
for i=1:S1, for k=1:R, W1(i, k)=x(R*(i-1)+k);
end
end
% 接着的 S1*S2 个编码（即第 R*S1 个后的编码）为 W
for i=1:S2, for k=1:S1, W2(i, k)=x(S1*(i-1)+k+R*S1);
end
end
% 接着的 S1 个编码（即第 R*S1+S1*S2 个后的编码）为 B1

```

```
for i=1:S1, B1(i, 1)=x((R*S1+S1*S2)+i);  
end  
% 接着的 S2 个编码 (即第 R*S1+S1*S2+S1 个后的编码) 为 B2  
for i=1:S2, B2(i, 1)=x((R*S1+S1*S2+S1)+i);  
end  
% 计算 S1 与 S2 层的输出  
A1=tansig(W1*P, B1);  
A2=purelin(W2*A1, B2);  
% 计算误差平方和  
SE=sumsqr(T-A2);  
val=1/SE; % 遗传算法的适应
```