

# 城市大气污染物变化规律分析与预测

## 摘要

对于问题一，我们分三方面来分析数据得到结论。第一方面：根据附件 1 中数据，得到各市的月平均空气污染指数变化折线图。分析得知影响因素有：季节因素：冬春季 API 平均值要明显高于夏秋季。政策因素：2014 年空气质量突然开始有所改善的原因是《陕西省大气污染防治条例》在 2014 年 1 月 1 日开始施行。地理因素：西安市、咸阳市均属于关中盆地中部，秦岭北麓，降雨期与风速等气候条件极其相似，所以其 AQI 值走势会相似。第二方面：我们根据“国家空气质量评价标准”对 AQI 月均值进行分类，得到各市空气污染情况的饼状图，并得知铜川的空气质量最好，无重度污染。渭南、西安、咸阳的空气质量最差，达到良的月数最少，重度污染的月数最多。第三方面：我们运用 MATLAB 得到每个城市每一天 6 种污染物的日平均浓度图，分析得知陕西省地处黄土高原南缘，盛行偏北风，北方沙漠区和黄土区地表的细颗粒（常以 PM<sub>10</sub> 为主）易被沙尘天气输送到大气中去，会造成 5 市 PM<sub>10</sub> 本地浓度较高，所以 5 市首要污染物均主要为 PM<sub>10</sub> 与所处特殊地理环境有很大关系。

对于问题二，我们建立了“基于支持向量机的大气污染预测模型”，并运用 SPSS 软件中的 T 检验验证得知预测是合理的。建立“支持向量机大气污染预测模型”的步骤有：构建有效的预测因子；选择支持向量机模型的核函数；选择模型的参数值；确定支持向量基的结构。在求解过程中，我们对数据进行了预处理，此处因为预测数据的类型不一样分为两种情况来讨论输入因子和输出因子。在程序编写中，为了防止数据溢出同时加快运算速度，我们对数据均进行归一化处理：将输入输出数据变换为  $[-1,1]$  区间的值，统计结果时再进行反归一化处理得到具

体的预测值。准备工作完成以后，对 SVM 模型进行学习训练，可得到未来 2 年 6 种主要污染物的日均浓度及空气质量指数的预测数据。我们运用 SPSS 软件对得到的 2019 上半年预测结果与现有的实际数据进行 T 检验得知：我们得到的预测结果是真实可信的，建立的模型是可靠的。

对于问题三，我们建立了“灰色关联分析模型”来分析周边城市对西安大气污染的影响程度。“灰色关联分析模型”主要步骤为：确定序列矩阵；用初始化方法进行无量纲化处理；求差序列；计算关联系数和计算关联度。运用 MATLAB 编程得到西安与周围城市 6 种污染物和空气质量间的关联度。按关联度大小排序为：咸阳>渭南>宝鸡>铜川。从数据上来看，西安与咸阳渭南的关联度最大，这与两市距离最近且相邻区域最大的情况吻合；西安与铜川的关联度最小，这也与两市距离最远且无相邻区域的情况吻合。

**关键词：**基于支持向量机的大气污染预测模型 灰色关联分析模型 T 检验

## 一、 问题重述

完成下列问题：

1 根据附件 1 中数据并结合空气质量分指数，分析各城市及整个关中平原区域各城市（宝鸡、咸阳、西安、渭南、铜川）的大气污染物浓度和空气质量的变化特征、分布规律（包括污染物浓度超标的时空特征等）。

2 对关中平原区域各城市未来 1 个月、2 个月、3 个月、6 个月、1 年内和 2 年六种主要污染物的日均浓度及空气质量指数的变化情况进行预测，并说明预测的合理性。

3 分析各城市不同污染物浓度变化之间的关联性，以及各城市空气质量分指数变化的关联性。并以西安市为例，研究周边城市对其大气污染的影响程度。

## 二、 问题分析

对于问题一：根据附件 1 中数据进行处理得到“2013-2019 上半年 5 市 AQI 空气质量指数月均值变化曲线”、“5 市 2013-2019 上半年各城市空气污染占比情况”、“5 市 2013-2019 上半年 6 种大气污染物浓度日均值变化曲线”三种图，对图进行分析得出各城市及整个关中平原区域各城市的大气污染物浓度和空气质量的变化特征、分布规律（污染物浓度普遍在冬季超标）的结论。

对于问题二：需要预测各城市未来 1 个月、2 个月、3 个月、6 个月、1 年内和 2 年六种主要污染物的日均浓度及空气质量分指数的变化情况，应当建立各城市大气污染指标浓度预测模型。但六种主要大气污染物浓度与空气质量分指数之间的关系为非线性的，因而采用一般的回归模型并不能够准确的反应其之间的关系，所以我们建立了“基于支持向量机得大气污染预测模型”，并运用 SPSS 中的 T 检验得知预测结果是真实可靠的。

对与问题三：我们建立了“灰色关联分析模型”来分析周边城市对西安大气污染的影响程度。“灰色关联分析模型”主要步骤为：确定序列矩阵；用初始化方法进行无量纲化处理；求差序列；计算关联系数和计算关联度。运用 MATLAB 编程得到西安与周围城市 6 种污染物和空气质量间的关联度。按关联度大小排序为：咸阳>渭南>宝鸡>铜川。从数据上来看，西安与咸阳渭南的关联度最大，这与两市距离最近且相邻区域最大的情况吻合；西安与铜川的关联度最小，这也与两市距离最远且无相邻区域的情况吻合。

## 三、 模型假设与符号说明

### 3.1 模型假设

- 1、假设题目所给的数据真实可靠。
- 2、假设无自然灾害影响。
- 3、假设 AQI 月均值能很好的代表该月空气质量，具有比较意义。

### 3.2 符号说明

$x_i$ :第*i*种大气污染物的浓度

$y$ :空气质量分指数

$x_{ij}$ :第*i*个城市第*j*种污染物的浓度值

$\Delta_{m,j}$ :第*m*个城市和西安市的第*j*种污染物浓度无量纲化数据的差值的绝对值

$\gamma(X_{m,j}, X_{1,j})$ :第*m*个城市与西安市的第*j*种影响因素的关联度

$\zeta$ :分辨系数

$\gamma_m$ :第*m*个城市与西安市的关联度

## 四、模型的建立与求解

### 4.1 问题一的思路分析

我们对附件 1 中数据进行处理得到“2013-2019 上半年 5 市 AQI 空气质量指数月均值变化曲线”、“5 市 2013-2019 上半年各城市空气污染占比情况”、“5 市 2013-2019 上半年 6 种大气污染物浓度日均值变化曲线”三种图。并根据国家空气质量评价标准对图进行分析,得到各城市及整个关中平原区域各城市的大气污染物浓度和空气质量的变化特征、分布规律(污染物浓度普遍在冬季超标)。

#### 4.1.1 空气污染指数与国家空气质量评价标准

空气污染指数描述了空气清洁或者污染的程度,以及对健康的影响。根据 2012 年 3 月国家发布的新空气质量评价标准,污染物监测为 6 项:二氧化硫、二氧化氮、PM10、PM2.5、一氧化碳和臭氧, AQI 将这 6 项污染物用统一的评价标准呈现。见表 1:

表 1 国家空气质量评价标准

空气污染指数	空气质量状况	对健康影响情况
0 - 50	一级(优)	空气质量令人满意,基本无空气污染
51 -100	二级(良)	空气质量可接受,但某些污染物可能对极少数异常敏感人群健康有较弱影响
101-150	三级(轻度污染)	易感人群症状有轻度加剧,健康人群出现刺激症状
151-200	四级(中度污染)	进一步加剧易感人群症状,可能对健康人群心脏、呼吸系统有影响
201-300	五级(重度污染)	心脏病和肺病患者症状显著加剧,运动耐受力降低,健康人群普遍出现症状
300+	六级(严重污染)	健康人群运动耐受力降低,有明显强烈症状,提前出现某些疾病

#### 4.1.2 问题一的求解

##### 1. 各市月平均空气污染指数变化分析

根据附件 1 中数据,计算出对陕西省 5 个城市(宝鸡、咸阳、西安、渭南、铜川)的月平均空气污染指数并进行分析。2013—2019 年 5 市月平均空气污染指数

变化见图 1:

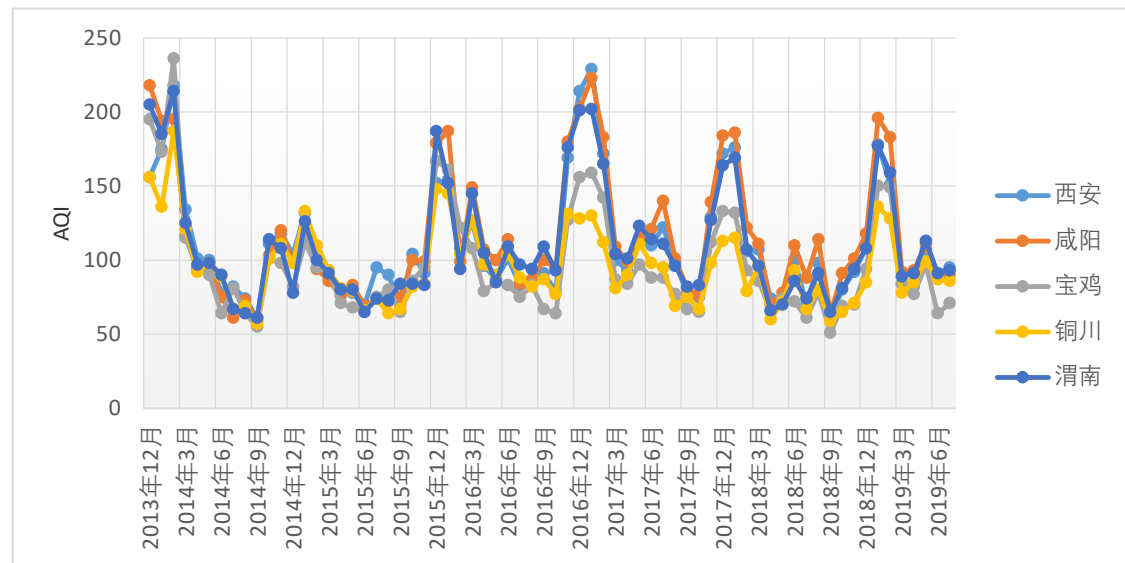


图 1 2013-2019 上半年 5 市空气质量指数 (AQI) 月均值变化曲线

季节因素: 由图 1 可见, 近几来 5 个城市 1—12 月份月平均 AQI 值总体变化规律呈“高一低一高”趋势。即 1—4 月份 AQI 值呈下降趋势, 5—9 月份呈平稳状态, 10—12 月份呈上升趋势。因为我国北方大气污染属于煤烟型污染, 所以这 5 市污染指数变化会有明显季节性变化特征, 冬春季 API 平均值要明显高于夏秋季。

政策因素: 其中自 2014 年 1 月开始, 5 个城市的 AQI 均都呈迅速下降趋势直到 2014 年 9 月达到几年内最低值。并且在该年冬季, AQI 也在 100 到 150 之间, 空气质量良好。空气质量突然开始有所改善的原因是《陕西省大气污染防治条例》在 2014 年 1 月 1 日开始施行。

地理因素: 图中 2013—2019 年 5 市月平均 AQI 值横向比较中, 铜川宝鸡的曲线明显异于其他 3 市, 而其他 3 市来走势基本一致。其中很重要原因的是: 西安市、咸阳市均属于关中盆地中部, 秦岭北麓, 降雨期与风速等气候条件极其相似, 所以其 AQI 值走势会相似。可以看出几个城市中铜川的空气污染程度从 15 年开始一直都较小。

## 2. 各城市近七年空气污染情况

我们根据表 1 中的国家空气质量评价标准, 对图 1 中的 AQI 的月均值进行分类, 得到下图 2:

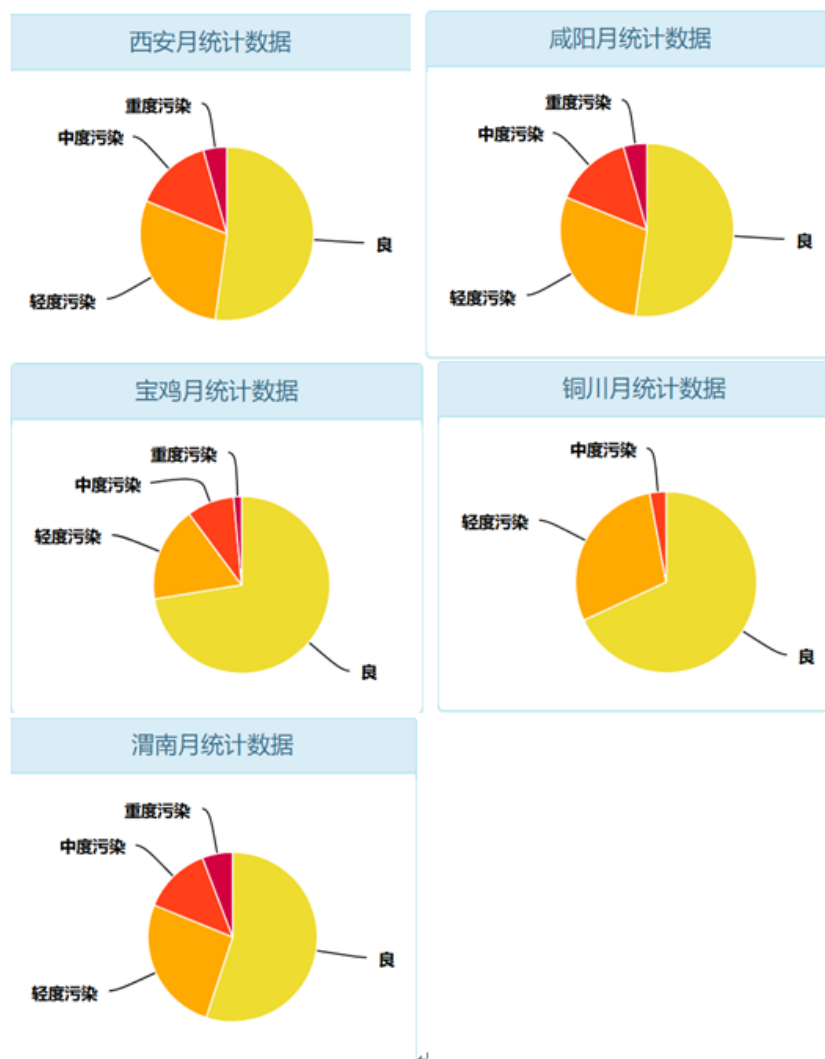


图2 5市2013-2019上半年各城市空气污染情况

从图中可以看出：2013年到2019年间城市空气污染情况：西安≈咸阳≈渭南>宝鸡>铜川。铜川的空气质量最好，无重度污染。渭南、西安、咸阳的空气质量最差，达到良的月数最少，重度污染的月数最多。

### 3. 各市近7年6种大气污染物浓度变化

我们运用MATLAB求出每个城市每一天6种污染物的日平均浓度，并作出下图3：

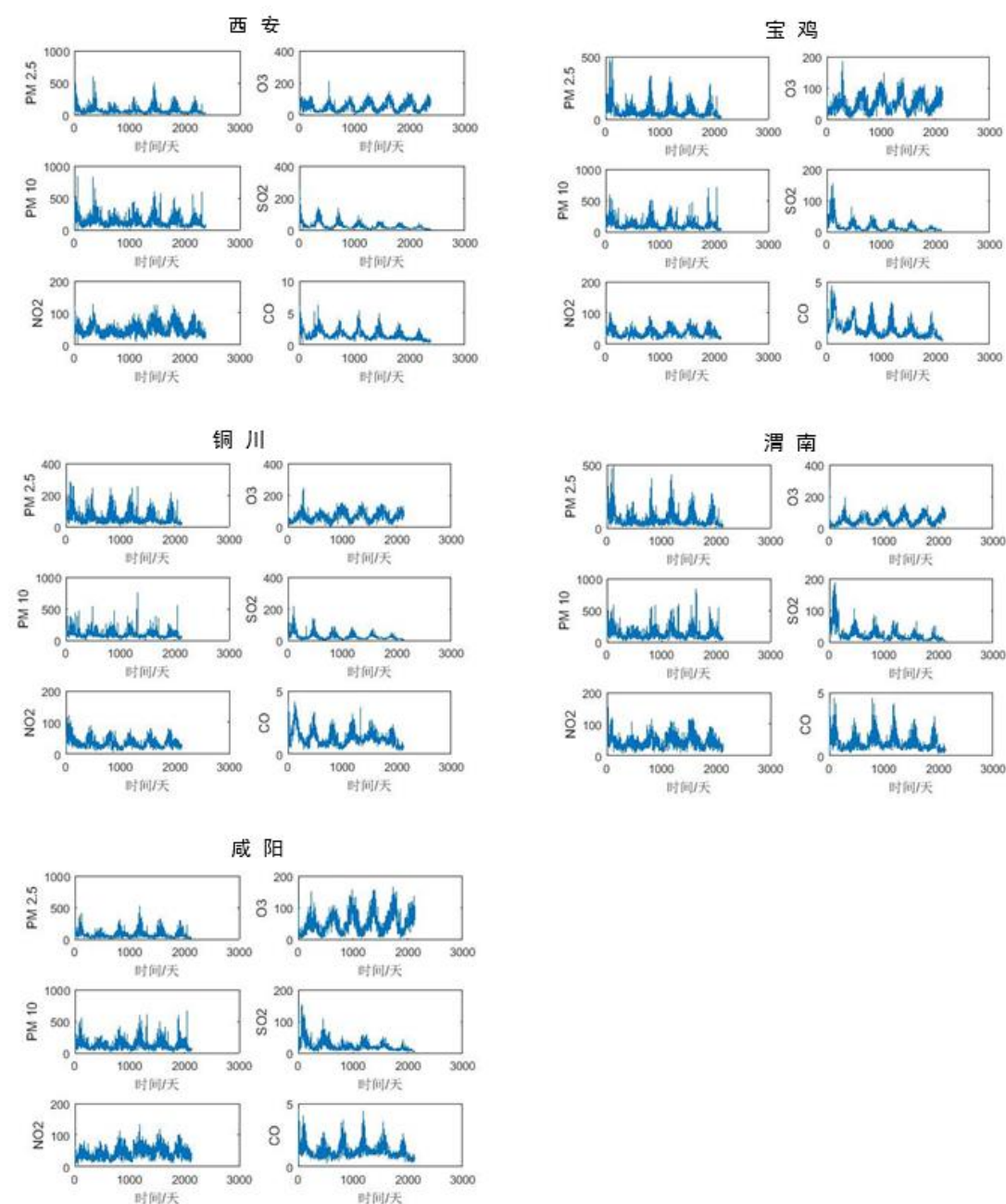


图3 5市2013-2019上半年6种大气污染物浓度日均值变化曲线

从图中我们可以得出：陕西省地处黄土高原南缘，盛行偏北风，北方沙漠区和黄土区地表的细颗粒（常以PM<sub>10</sub>为主）易被沙尘天气输送到大气中去，会造成5市PM<sub>10</sub>本地浓度较高，所以4市的首要污染物均主要为PM<sub>10</sub>与所处特殊地理环境有很大关系。该区域人口密度大，重化产业聚集，能源结构偏煤、产业结构偏重、运输结构偏公路，是我国细颗粒物(PM<sub>2.5</sub>)浓度仅次于京津冀区域第二高的区域，同时又是二氧化硫浓度最高的区域。5市的首要污染物主要为PM<sub>10</sub>。由于陕西省地处黄土高原，有造成PM<sub>10</sub>污染的不可避免的客观原因，但还是可以采取很多措施降低PM<sub>10</sub>污染。例如加快生态环境建设步伐，积极退耕还林，大力绿化、美化城区环境，提高固体废弃物的回收利用率，降低PM<sub>10</sub>的本底值；严格控制电厂、水泥厂等行业的粉尘排放总量；加强对施工扬尘的管理；增加城市绿化面积，提高街道

人工洒水频率;多做环保宣传,提高全民环保意识。这些都可以大大减少 PM10 污染浓度。

#### 4.2 问题二的思路分析

本小题中,需要预测各城市未来 1 个月、2 个月、3 个月、6 个月、1 年内和 2 年六种主要污染物的日均浓度及空气质量分指数的变化情况,应当建立各城市大气污染指标浓度预测模型。但六种主要大气污染物浓度与空气质量分指数之间的关系为非线性的,因而采用一般的回归模型并不能够准确的反应其之间的关系,所以我们建立了“基于支持向量机得大气污染预测模型”,并运用 SPSS 软件中的 T 检验得知建立的模型得到的预测是合理的。

##### 4.2.1 “基于支持向量机的大气污染预测模型”的建立

以下为“支持向量机大气污染预测模型”的建立步骤:

###### 1.构建有效的预测因子。

本题中,需要预测六种主要污染物的日均浓度:PM2.5浓度( $x_1$ )、O<sub>3</sub>浓度( $x_2$ )、PM10浓度( $x_3$ )、SO<sub>2</sub>浓度( $x_4$ )、NO<sub>2</sub>浓度( $x_5$ )、CO浓度( $x_6$ )及空气质量分指数( $y$ )的变化情况,因而,存在如下两种情形:

A.在预测某一种大气污染物的浓度  $x_i$  时,可构建  $x_i$  与其余五种污染物的浓度  $x_j(j \neq i)$  的关系式:

$$x_i = f(x_j), j \neq i, i=1, \dots, 6, j=1, \dots, 6$$

B.在预测空气质量分指数  $y$  时,可构建  $y$  与六种污染物的浓度  $x_m(m=1, \dots, 6)$  之间的关系式:

$$y = \phi(x_m), m=1, \dots, 6$$

###### 2.选择支持向量机模型的核函数。

常用的核函数有四种:线性核函数、多项式核函数、RBF 核函数(高斯核函数)、sigmoid 核函数。因四种核函数的使用条件不同,考虑到该题的特征维数远远小于样本的数量,并且每个因素之间存在着非线性关系,所以我们选择最符合本题的 RBF 核函数作为预测模型的核函数。RBF 核函数表达式如下:

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}$$

其中

$$\gamma = \frac{1}{2\delta^2}, \delta > 0$$

###### 3.选择模型的参数值。

在建立模型时选用的 RBF 核函数,其中影响支持向量机学习性能的两个未知参数分别是 RBF 核函数中的参数  $\delta$ 、惩罚因子  $C$ 。支持向量机推广性能的好坏,相当程度上取决于参数的选择,因此我们采用了交叉验证的方法对参数进行优化选择。最终选取参数为:

$$\delta = 50, C = 1000$$

#### 4.确定支持向量基的结构。

用训练样本训练具有优化参数的支持向量机预测模型，获得支持向量，确定支持向量机的结构。最后使用训练过的支持向量预测器对测试样本进行预测，得到结果。

#### 4.2.2 “基于支持向量机的大气污染预测模型”的求解

##### 1.数据的预处理（存在两种情况）。

A.将附件 1 中 2013 年-2018 年关中平原区域各城市（宝鸡、咸阳、西安、渭南、铜川）的 PM<sub>2.5</sub>浓度、O<sub>3</sub>浓度、PM<sub>10</sub>浓度、SO<sub>2</sub>浓度、NO<sub>2</sub>浓度、CO浓度，除去需要预测的大气污染物浓度，剩下的五个因子作为SVM预测模型的预选输入因子，输出为该城市当日预测的大气污染物的浓度值。

B.将附件 1 中 2013 年-2018 年关中平原区域各城市（宝鸡、咸阳、西安、渭南、铜川）的 PM<sub>2.5</sub>浓度、O<sub>3</sub>浓度、PM<sub>10</sub>浓度、SO<sub>2</sub>浓度、NO<sub>2</sub>浓度、CO浓度共六个因子作为SVM预测模型的预选输入因子，输出为该城市当日的空气质量指数AQI。

在程序编写中，为了防止数据溢出同时加快运算速度，我们对数据均进行归一化处理:将输入输出数据变换为[-1,1]区间的值，统计结果时再进行反归一化处理得到具体的预测值。

##### 2.SVM的学习训练与预测。

将各城市（宝鸡、咸阳、西安、渭南、铜川）在 2013 到 2018 年间的的数据作为训练样本，2019 上半年的数据作为测试样本。每组数据包括了数据预处理中的输入因子，因为需要预测六种主要污染物的日均浓度及空气质量分指数，所以同样分为两种情况。

准备工作完成以后，对SVM模型进行学习训练，可得到未来 2 年六种主要污染物的日均浓度及空气质量指数的预测数据。我们单独做出了各市 AQI 的预测图，见图 4：



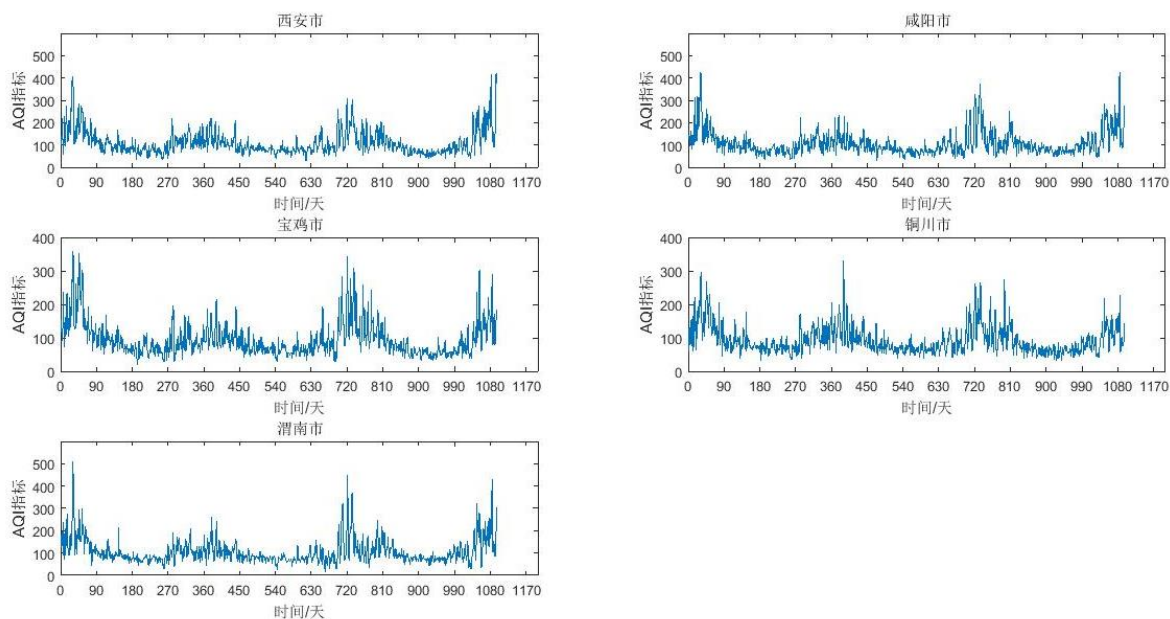


图 4 各市未来两年空气质量指数的预测数据

因为得到数据过多所以放在附件中，此处仅列举了未来一周内的预测数据，见表 2-表 6：

表 2 西安未来一周内六种主要污染物的日均浓度及空气质量指数

日期	AQI	PM2.5	O3	PM10	SO2	NO2	CO
2019/8/1	111.20	77.17	97.19	134.51	12.51	39.57	1.22
2019/8/2	108.57	74.06	89.97	136.22	9.52	39.54	1.12
2019/8/3	79.05	43.94	83.13	104.93	5.58	31.87	0.99
2019/8/4	76.40	43.21	82.50	91.48	6.91	30.85	1.07
2019/8/5	60.45	32.11	49.80	67.46	0.38	31.85	0.93
2019/8/6	43.37	17.60	32.20	41.21	7.75	31.11	1.37
2019/8/7	38.19	13.25	31.58	31.79	7.64	31.14	1.43

表 3 咸阳未来一周内六种主要污染物的日均浓度及空气质量指数

日期	AQI	PM2.5	O3	PM10	SO2	NO2	CO
2019/8/1	93.60	57.48	83.37	103.91	19.97	26.40	1.08
2019/8/2	94.36	54.12	87.88	111.47	20.37	29.59	1.03
2019/8/3	99.46	58.89	87.88	116.13	22.32	32.49	1.03
2019/8/4	106.04	64.30	81.44	125.40	22.50	34.40	1.07
2019/8/5	81.99	43.26	63.76	104.03	21.18	30.40	1.05
2019/8/6	81.91	43.49	64.48	104.01	18.62	31.15	1.10
2019/8/7	56.53	26.05	43.88	72.48	17.14	21.62	1.11

表 4 宝鸡未来一周内六种主要污染物的日均浓度及空气质量指数

日期	AQI	PM2.5	O3	PM10	SO2	NO2	CO
2019/8/1	83.71	55.97	71.97	86.48	10.61	26.30	1.07
2019/8/2	102.39	72.37	77.47	113.48	18.67	28.19	1.28

2019/8/3	99.40	71.59	82.32	104.53	16.18	28.12	1.21
2019/8/4	114.32	83.10	87.45	123.82	16.54	29.15	1.28
2019/8/5	62.08	34.18	65.39	64.96	9.98	19.10	1.04
2019/8/6	62.04	37.11	56.50	63.53	7.90	27.05	1.06
2019/8/7	41.43	18.91	33.01	45.92	7.30	21.78	1.16

表 5 铜川未来一周内六种主要污染物的日均浓度及空气质量指数

日期	AQI	PM2_5	O3	PM10	SO2	NO2	CO
2019/8/1	99.84	74.17	102.97	76.43	5.81	18.08	0.90
2019/8/2	83.83	58.75	95.01	70.41	4.59	26.34	0.98
2019/8/3	93.45	66.39	92.43	83.27	9.81	25.98	1.02
2019/8/4	93.23	66.79	98.17	78.74	6.96	23.05	0.96
2019/8/5	76.15	46.81	87.35	79.94	6.27	26.21	0.91
2019/8/6	62.28	37.91	66.91	63.31	8.34	26.24	0.82
2019/8/7	64.52	39.18	50.27	66.47	12.66	26.56	0.90

表 6 咸阳未来一周内六种主要污染物的日均浓度及空气质量指数

日期	AQI	PM2_5	O3	PM10	SO2	NO2	CO
2019/8/1	82.35	44.64	89.79	100.68	16.61	25.91	0.77
2019/8/2	90.14	52.99	92.92	106.89	18.93	26.90	0.86
2019/8/3	91.62	55.44	88.74	106.86	20.82	27.89	0.85
2019/8/4	93.17	55.66	80.86	113.24	18.61	30.53	0.78
2019/8/5	90.41	50.62	71.24	116.76	20.54	36.70	0.76
2019/8/6	72.61	37.12	72.97	90.37	20.19	30.89	0.79
2019/8/7	59.91	33.89	50.44	66.77	15.00	30.13	0.92

#### 4.2.2 “基于支持向量机的大气污染预测模型”的检验

我们运用 SPSS 软件对得到的 2019 上半年预测结果与现有的实际数据进行 T 检验，结果如下图 5 到图 9：

成对样本检验 <sup>a</sup>								
<sup>a</sup>	成对差分 <sup>a</sup>					t <sup>a</sup>	df <sup>a</sup>	Sig.(双侧) <sup>a</sup>
	均值 <sup>a</sup>	标准差 <sup>a</sup>	均值的标准误 <sup>a</sup>	差分的 95% 置信区 <sup>a</sup>				
				间 <sup>a</sup>				
				下限 <sup>a</sup>	上限 <sup>a</sup>			
对 1 <sup>a</sup> x1 - y1 <sup>a</sup>	.93885	44.63286	1.34818	-1.70647	3.58416	.696	1095	.086
对 2 <sup>a</sup> x2 - y2 <sup>a</sup>	.58664	37.98126	1.14727	-1.66445	2.83772	.511	1095	.009
对 3 <sup>a</sup> x3 - y3 <sup>a</sup>	.19932	11.88062	.35887	-.50483	.90346	.555	1095	.079
对 4 <sup>a</sup> x4 - y4 <sup>a</sup>	-.55357	62.05938	1.87457	-4.23173	3.12459	-.295	1095	.068
对 5 <sup>a</sup> x5 - y5 <sup>a</sup>	.40507	10.31272	.31151	-.20615	1.01629	1.300	1095	.094
对 6 <sup>a</sup> x6 - y6 <sup>a</sup>	1.35135	10.54397	.31849	.72642	1.97627	4.243	1095	.000
对 7 <sup>a</sup> x7 - y7 <sup>a</sup>	-.01406	.48118	.01453	-.04258	.01446	-.967	1095	.034

图 5 西安 2019 上半年预测结果与现有的实际数据的显著性检验

成对样本检验 <sup>a</sup>								
	成对差分 <sup>a</sup>					t <sup>a</sup>	df <sup>a</sup>	Sig.(双 侧) <sup>a</sup>
	均值 <sup>a</sup>	标准差 <sup>a</sup>	均值的标准 误 <sup>a</sup>	差分的 95% 置信区 间 <sup>a</sup>				
				下限 <sup>a</sup>	上限 <sup>a</sup>			
对 1 <sup>a</sup> x1 - y1	38.47931	128.18334	10.50119	17.72767	59.23094	3.664	148	.000
对 2 <sup>a</sup> x2 - y2	38.34196	118.54866	9.71188	19.15008	57.53383	3.948	148	.000
对 3 <sup>a</sup> x3 - y3	-16.63227	28.59120	2.34228	-21.26091	-12.00363	-7.101	148	.000
对 4 <sup>a</sup> x4 - y4	39.70107	148.58789	12.17279	15.64613	63.75600	3.261	148	.001
对 5 <sup>a</sup> x5 - y5	19.58760	39.97065	3.27452	13.11674	26.05846	5.982	148	.000
对 6 <sup>a</sup> x6 - y6	-3.65400	18.35985	1.50410	-6.62628	-.68172	-2.429	148	.016
对 7 <sup>a</sup> x7 - y7	.33519	.86149	.07058	.19572	.47465	4.749	148	.000

图 6 咸阳 2019 上半年预测结果与现有的实际数据的显著性检验

成对样本检验 <sup>a</sup>								
°	成对差分 <sup>a</sup>					t <sup>a</sup>	df <sup>a</sup>	Sig.(双侧) <sup>a</sup>
	均值 <sup>a</sup>	标准差 <sup>a</sup>	均值的标准 误 <sup>a</sup>	差分的 95% 置信区 间 <sup>a</sup>				
				下限 <sup>a</sup>	上限 <sup>a</sup>			
对 1 <sup>a</sup> x1 - y1 <sup>a</sup>	30.49161	126.47493	10.39617	9.94634	51.03687	2.933	147	.004
对 2 <sup>a</sup> x2 - y2 <sup>a</sup>	31.98068	128.31842	10.54771	11.13594	52.82541	3.032	147	.003
对 3 <sup>a</sup> x3 - y3 <sup>a</sup>	-10.70790	22.16002	1.82154	-14.30769	-7.10811	-5.878	147	.000
对 4 <sup>a</sup> x4 - y4 <sup>a</sup>	20.82650	123.48808	10.15066	.76643	40.88656	2.052	147	.042
对 5 <sup>a</sup> x5 - y5 <sup>a</sup>	19.96515	47.23299	3.88253	12.29237	27.63793	5.142	147	.000
对 6 <sup>a</sup> x6 - y6 <sup>a</sup>	5.78790	20.85477	1.71425	2.40014	9.17566	3.376	147	.001
对 7 <sup>a</sup> x7 - y7 <sup>a</sup>	.32633	1.64331	.13508	.05938	.59327	2.416	147	.017

图 7 宝鸡 2019 上半年预测结果与现有的实际数据的显著性检验

成对样本检验 <sup>a</sup>								
	成对差分 <sup>a</sup>					t <sup>a</sup>	df <sup>a</sup>	Sig.(双 侧) <sup>a</sup>
	均值 <sup>a</sup>	标准差 <sup>a</sup>	均值的标准 误 <sup>a</sup>	差分的 95% 置信区 间 <sup>a</sup>				
				下限 <sup>a</sup>	上限 <sup>a</sup>			
对 1 <sup>a</sup> x1 - y1	16.82275	94.77763	7.76449	1.47917	32.16632	2.167	148	.032
对 2 <sup>a</sup> x2 - y2	20.43812	86.70735	7.10334	6.40105	34.47520	2.877	148	.005
对 3 <sup>a</sup> x3 - y3	-15.89628	26.09329	2.13765	-20.12052	-11.67203	-7.436	148	.000
对 4 <sup>a</sup> x4 - y4	10.45197	101.94523	8.35168	-6.05197	26.95591	1.251	148	.213
对 5 <sup>a</sup> x5 - y5	20.43355	57.16958	4.68351	11.17835	29.68875	4.363	148	.000
对 6 <sup>a</sup> x6 - y6	13.64434	24.95705	2.04456	9.60403	17.68464	6.673	148	.000
对 7 <sup>a</sup> x7 - y7	.13300	1.32019	.10815	-.08072	.34673	1.230	148	.221

图 8 铜川 2019 上半年预测结果与现有的实际数据的显著性检验

成对样本检验 <sup>◇</sup>								
◇	成对差分 <sup>◇</sup>					t <sup>◇</sup>	df <sup>◇</sup>	Sig.(双 侧) <sup>◇</sup>
	均值 <sup>◇</sup>	标准差 <sup>◇</sup>	均值的标准 误 <sup>◇</sup>	差分的 95% 置信区 间 <sup>◇</sup>				
				下限 <sup>◇</sup>	上限 <sup>◇</sup>			
对 1 <sup>◇</sup> x1 - y1 <sup>◇</sup>	48.82276 <sup>◇</sup>	124.44945 <sup>◇</sup>	10.16126 <sup>◇</sup>	28.74399 <sup>◇</sup>	68.90154 <sup>◇</sup>	4.805 <sup>◇</sup>	149 <sup>◇</sup>	.000 <sup>◇</sup>
对 2 <sup>◇</sup> x2 - y2 <sup>◇</sup>	49.09356 <sup>◇</sup>	122.74829 <sup>◇</sup>	10.02236 <sup>◇</sup>	29.28926 <sup>◇</sup>	68.89787 <sup>◇</sup>	4.898 <sup>◇</sup>	149 <sup>◇</sup>	.000 <sup>◇</sup>
对 3 <sup>◇</sup> x3 - y3 <sup>◇</sup>	-13.48016 <sup>◇</sup>	30.59242 <sup>◇</sup>	2.49786 <sup>◇</sup>	-18.41597 <sup>◇</sup>	-8.54435 <sup>◇</sup>	-5.397 <sup>◇</sup>	149 <sup>◇</sup>	.000 <sup>◇</sup>
对 4 <sup>◇</sup> x4 - y4 <sup>◇</sup>	51.84242 <sup>◇</sup>	137.27603 <sup>◇</sup>	11.20854 <sup>◇</sup>	29.69419 <sup>◇</sup>	73.99064 <sup>◇</sup>	4.625 <sup>◇</sup>	149 <sup>◇</sup>	.000 <sup>◇</sup>
对 5 <sup>◇</sup> x5 - y5 <sup>◇</sup>	24.37359 <sup>◇</sup>	62.63299 <sup>◇</sup>	5.11396 <sup>◇</sup>	14.26833 <sup>◇</sup>	34.47884 <sup>◇</sup>	4.766 <sup>◇</sup>	149 <sup>◇</sup>	.000 <sup>◇</sup>
对 6 <sup>◇</sup> x6 - y6 <sup>◇</sup>	8.25863 <sup>◇</sup>	26.21629 <sup>◇</sup>	2.14055 <sup>◇</sup>	4.02887 <sup>◇</sup>	12.48838 <sup>◇</sup>	3.858 <sup>◇</sup>	149 <sup>◇</sup>	.000 <sup>◇</sup>
对 7 <sup>◇</sup> x7 - y7 <sup>◇</sup>	.48771 <sup>◇</sup>	1.01603 <sup>◇</sup>	.08296 <sup>◇</sup>	.32379 <sup>◇</sup>	.65164 <sup>◇</sup>	5.879 <sup>◇</sup>	149 <sup>◇</sup>	.000 <sup>◇</sup>

图 9 渭南 2019 上半年预测结果与现有的实际数据的显著性检验

在SPSS软件统计结果中“sig”，意为“显著性”，后面的值就是统计出的P值，如果P值 $0.01 < P < 0.05$ ，则为差异显著，如果 $P < 0.01$ ，则差异极显著。而我们分析表中的数据可以知道：我们得到的预测结果是真实可信的，建立的模型是可靠的。

#### 4.3 问题三的思路分析

本小题中要分析各城市不同污染物浓度变化之间的关联性以及各城市空气质量分指数变化的关联性，我们建立了灰色关联分析模型对此求解。灰色关联度计算主要步骤有确定序列矩阵、用初始化方法进行无量纲化处理、求差序列、计算关联系数和计算关联度。最后根据得到的关联度进行分析，得到以西安市为例，周边城市对其大气污染的影响程度。

##### 4.3.1 “灰色关联分析模型”的建立

###### 1. 确定序列矩阵。

本文中序列矩阵是指五所城市的 2013-2019 上半年的统计数据，表示为：

$$x_i = \{x_{ij} \mid j = 1, \dots, 7\}, i = 1, \dots, 5$$

其中， $i$ 表示五所城市； $j$ 表示影响不同城市之间大气污染物浓度及空气质量分指数变化的因素种类数。

###### 2. 用初始化方法进行无量纲化处理。

用每个城市的某种污染物的浓度值除以五所城市该种污染物浓度值之和，得到每个数据无量纲化结果：

$$X_{ij} = \frac{x_{ij}}{\sum_{i=1}^5 x_{ij}}$$

其中， $x_{ij}$ 指第 $i$ 个城市第 $j$ 种污染物的浓度值； $X_{ij}$ 指 $x_{ij}$ 进行无量纲化后的值

###### 3. 求差序列。

首先定义西安市的各污染物浓度及空气质量分指数的无量纲化数据 $X_{0,j}$ 为

参考序列，其余四城市数据为比较数列，表达式如下：

$$X_1 = (X_{1j}), j = 1, \dots, 7$$

其次对每个城市每种污染物浓度的无量纲化处理数据  $X_{mj}$  减去另一个城市对应污染物浓度的无量纲化处理数据  $X_{1j}$  进行绝对值计算，得到这两个城市盖中污染物浓度的绝对差数列，表达式如下：

$$\Delta_{m,j} = |X_{mj} - X_{1j}| (m = 1, \dots, 5, j = 1, \dots, 7)$$

其中， $\Delta_{m,j}$  指第  $p$  个城市和第  $q$  个城市的第  $j$  种污染物浓度无量纲化数据的差值的绝对值。

最后将  $\Delta_{m,j}$  转化为矩阵，则可以得到一个如下  $5 \times 7$  的绝对值差对称矩阵：

$$A = (\Delta_{m,j})_{5 \times 7}$$

4. 计算关联系数。

计算其余四所城市与西安市的各种污染物浓度的关联度的表达式如下：

$$\zeta(X_{m,j}, X_{1,j}) = \frac{\min_{1 \leq j \leq 7}(\Delta_{m,j}) + \zeta \max_{1 \leq j \leq 7}(\Delta_{m,j})}{|X_{1j} - X_{mj}| + \zeta \max_{1 \leq j \leq 7}(\Delta_{m,j})} (\zeta = 0.5, m = 1, \dots, 5, j = 1, \dots, 7)$$

其中， $\gamma(X_{m,j}, X_{1,j})$  表示第  $m$  个城市与西安市的第  $j$  种影响因素的关联度；

$\min_{1 \leq j \leq 7}(\Delta_{m,j})$  表示第  $m$  个城市与西安市的  $j$  种影响因素的绝对值差序列的最小值；

$\max_{1 \leq j \leq 7}(\Delta_{m,j})$  表示第  $m$  个城市与西安市的  $j$  种影响因素的绝对值差序列的最大值；

$\zeta$  表示分辨系数，一般取  $0 < \zeta < 1$ ，在这里我们取  $\zeta = 0.5$ ； $X_{1j}$  表示西安市的第  $j$  种污染物的浓度值； $X_{mj}$  表示第  $m$  个城市第  $j$  种污染物的浓度值。

5. 计算关联度  $\gamma_m$  的公式如下：

$$\gamma_m = \frac{1}{j} \sum_{j=1}^7 \zeta(X_{m,j}, X_{1,j}), m = 1, \dots, 5$$

#### 4.3.2 “灰色关联分析模型”的求解

我们运用 MATLAB 得到西安与周边城市六种主要污染物和空气质量指数的关联度，如下表 7、表 8：

表 7 西安与周边城市六种主要污染物的关联度

西安与咸阳	西安与宝鸡	西安与铜川	西安与渭南
0.7900763510	0.6777426527	0.6304524958	0.7738758419

表 8 西安与周边城市空气质量指数的关联度

西安与咸阳	西安与宝鸡	西安与铜川	西安与渭南
0.3845066254	0.3389396873	0.3333333333	0.3749156033

得知按关联度大小排序为：咸阳>渭南>宝鸡>铜川。从数据上来看，西安与咸阳渭南的关联度最大，这与两市距离最近且相邻区域最大的情况吻合；西安与铜川的关联度最小，这也与两市距离最远且无相邻区域的情况吻合。

## 五、 模型的评价与优化

### 5.1 模型的优缺点分析

#### 5.1.1 模型的优点

1、由于支持向量机的优良特性，特别适合于那些模糊、随机、不确定性、样本数有限和非线性的复杂问题，支持向量回归模型能很好捕捉大气污染物浓度与其影响因子的非线性关系。

2、支持向量机预测大气污染模型的核函数我们选取了 RBF 核函数（高斯径向基函数），通过对核函数和惩罚参数 C 的交叉测试，C 取 1000。RBF 核函数适用于特征数量少，而样本数量正常的情形，在对预测值和实际值进行配对样本 T 检验时，其统计显著性 sig 的值均很小，也说明了此函数的优越性。

#### 5.1.2 模型的缺点

求解问题三时，我们选择的灰色关联度分析法，没有考虑到空间维度对各城市污染物浓度的影响，会造成一些误差。

## 六、 模型的推广

SVM 支持向量机是一种新型机器学习理论。它是基于统计理论，根据结构风险最小化而设计的，目前已成功应用于多个领域，特别是对于大数据的处理，不会出现维数灾难现象。SVM 采用结构最小化原则，学习目标是得到现有数据下的最优解，而不是样本趋于无穷时的最优；此方法是将机器学习巧妙地转换成二次规划学习问题，理论上凸集下二次规划可得到全局最优，避免了局部极小值问题。SVM 在各领域的模式识别问题中有应用，包括人像识别、文本分类、手写字符识别、生物信息学等。

## 参考文献

- [1] 万旭.川南六市区域主要大气污染物关联性研究(M).四川：四川师范大学，2016.
- [2] 白永亮,郭珊,孙涵.大气污染的空间关联与区域间防控协作——基于全国 288 个地市工业 SO<sub>2</sub> 污染数据的空间统计分析(M).武汉：中国地质大学学报(社会科学版)，2016.
- [3] 陈恩水,王峰.数学建模与实验(M)。北京：科学出版社，2008.
- [4] 陈俏.支持向量机应用于大气污染物浓度预测(M).西安：西安科技大学，2010.
- [5] 余倩楠.基于多源数据的长三角地区主要大气污染物时空演变特征及其重污染时段成因与模拟分析(M).上海：华东师范大学，2019.

## 附件

运行环境 matlab2015b , 并安装 libsvm 工具包 (Matlab 版)。

### 1. 数据处理

将问题中的 Excel 中表的数据进行初步处理。将各城市中各站点进行编号, 将日期转换为数值数据 (如 2019/8/29 转换为 2019829)。

### 2. 数据录入

将处理后表的数据, 通过复制、粘贴建立数据文件 (\*.txt), 分别命名为 xian、xianyang、baoji、tongchuan、weinan, 并保存至 matlab 的工作文件夹中, 等待程序读取数据。

根据查找后的各城市各站点的经纬度建立数据文件 (\*.txt), 分别命名为 xian01、xianyang01、baoji01、tongchuan01、weinan01, 并保存至 matlab 的工作文件夹中, 等待程序读取数据。

### 3. 建立 .m 文件, 并命名为 disici01。

以西安市数据为例, 只需将第二行中的第二个 xian 换为 xianyang、baoji、tongchuan、weinan, 即可依次得到其余四座城市的相关数据及图。

```
load xian.txt,load xianyang.txt,load baoji.txt,load tongchuan.txt,load weinan.txt;
xian=xian;
time=size(xian);
day=0;
XA0=zeros();
for j=1:1:7
    data=2013;
    for i=1:1:time(1)
        if data~=xian(i,8)
            data=xian(i,8);
            num=0;
            a=0;
        end
        if data==xian(i,8)
            num=num+1;
            a=a+xian(i,j);
            if num==1
                day=day+1;
            end
        end
        XA0(day,j)=a/num;
    end
end
b=size(XA0);
XA1=zeros((b(1)/7),8);
for i=1:1:(b(1)/7)
    for j=1:1:7
```



```

        XA1(i,j)=XA0((b(1)/7)*(j-1)+i,j);
    end
    XA1(i,8)=i;
end
plot(XA1(:,8),XA1(:,1));
hold on
xlabel('时间/天');
ylabel('AQI 指标');
title('西安市');
set(gca,'Xtick',[0:200:2300]);
time=size(xian);
day=0;
XA0=zeros();
for j=1:1:7
    data=2013;
    for i=1:1:time(1)
        if data~=xian(i,8)
            data=xian(i,8);
            num=0;
            a=0;
        end
        if data==xian(i,8)
            num=num+1;
            a=a+xian(i,j);
            if num==1
                day=day+1;
            end
        end
        XA0(day,j)=a/num;
    end
end
b=size(XA0);
XA1=zeros((b(1)/7),8);
for i=1:1:(b(1)/7)
    for j=1:1:7
        XA1(i,j)=XA0((b(1)/7)*(j-1)+i,j);
    end
    XA1(i,8)=i;
End
figure
subplot(3,2,1);
plot(XA1(:,8),XA1(:,2));
set(gca,'Xtick',[0:200:2300]);
xlabel('时间/天');ylabel('PM 2.5');

```

```

subplot(3,2,2);
plot(XA1(:,8),XA1(:,3));
set(gca,'Xtick',[0:200:2300]);
xlabel('时间/天');ylabel('O3');
subplot(3,2,3);
plot(XA1(:,8),XA1(:,4));
set(gca,'Xtick',[0:200:2300]);
xlabel('时间/天');ylabel('PM 10');
subplot(3,2,4);
plot(XA1(:,8),XA1(:,5));
set(gca,'Xtick',[0:200:2300]);
xlabel('时间/天');ylabel('SO2');
subplot(3,2,5);
plot(XA1(:,8),XA1(:,6));
set(gca,'Xtick',[0:200:2300]);
xlabel('时间/天');ylabel('NO2');
subplot(3,2,6);
plot(XA1(:,8),XA1(:,7));
set(gca,'Xtick',[0:200:2300]);
xlabel('时间/天');ylabel('CO');

```

建立.m 文件，并命名为 disici02\_xian

```

function xianpjz=disici02_xian()
load xian.txt;
cs=xian;
a=size(cs);
dian=max(cs(:,9));
pjz=zeros(7,dian);%每个城市几个点
for j=1:1:7
    num1=0;num2=0;num3=0;num4=0;num5=0;num6=0;num7=0;num8=0;num9=0;
num10=0;num11=0;num12=0;num13=0;
    for i=1:1:a(1)
        b=cs(i,9);
        switch b
            case 1,
                num1=num1+1;
                pjz(j,b)=pjz(j,b)+cs(i,j);
            case 2,
                num2=num2+1;
                pjz(j,b)=pjz(j,b)+cs(i,j);
            case 3,
                num3=num3+1;
                pjz(j,b)=pjz(j,b)+cs(i,j);

```

```

        case 4,
            num4=num4+1;
            pjz(j,b)=pjz(j,b)+cs(i,j);
        case 5,
            num5=num5+1;
            pjz(j,b)=pjz(j,b)+cs(i,j);
        case 6,
            num6=num6+1;
            pjz(j,b)=pjz(j,b)+cs(i,j);
        case 7,
            num7=num7+1;
            pjz(j,b)=pjz(j,b)+cs(i,j);
        case 8,
            num8=num8+1;
            pjz(j,b)=pjz(j,b)+cs(i,j);
        case 9,
            num9=num9+1;
            pjz(j,b)=pjz(j,b)+cs(i,j);
        case 10,
            num10=num10+1;
            pjz(j,b)=pjz(j,b)+cs(i,j);
        case 11,
            num11=num11+1;
            pjz(j,b)=pjz(j,b)+cs(i,j);
        case 12,
            num12=num12+1;
            pjz(j,b)=pjz(j,b)+cs(i,j);
        case 13,
            num13=num13+1;
            pjz(j,b)=pjz(j,b)+cs(i,j);
    end
end
pjz(j,1)=pjz(j,1)/num1;pjz(j,2)=pjz(j,2)/num2;pjz(j,3)=pjz(j,3)/num3;pjz(j,4)=pjz(
j,4)/num4;
pjz(j,5)=pjz(j,5)/num5;pjz(j,6)=pjz(j,6)/num6;pjz(j,7)=pjz(j,7)/num7;pjz(j,8)=pjz(
j,8)/num8;
pjz(j,9)=pjz(j,9)/num9;pjz(j,10)=pjz(j,10)/num10;pjz(j,11)=pjz(j,11)/num11;
    pjz(j,12)=pjz(j,12)/num12;pjz(j,13)=pjz(j,13)/num13;
end
xianpjz=pjz;
建立.m 文件， 并命名为 disici02_xianyang
function xianyangpjz=disici02_xianyang()
load xianyang.txt;
cs=xianyang;

```

```

a=size(cs);
dian=max(cs(:,9));
pjz=zeros(7,dian);%每个城市几个点
for j=1:1:7
    num1=0;num2=0;num3=0;num4=0;
    for i=1:1:a(1)
        b=cs(i,9);
        switch b
            case 1,
                num1=num1+1;
                pjz(j,b)=pjz(j,b)+cs(i,j);
            case 2,
                num2=num2+1;
                pjz(j,b)=pjz(j,b)+cs(i,j);
            case 3,
                num3=num3+1;
                pjz(j,b)=pjz(j,b)+cs(i,j);
            case 4,
                num4=num4+1;
                pjz(j,b)=pjz(j,b)+cs(i,j);
        end
    end
    pjz(j,1)=pjz(j,1)/num1;pjz(j,2)=pjz(j,2)/num2;pjz(j,3)=pjz(j,3)/num3;pjz(j,4)=pjz(
j,4)/num4;
end
xianyangpjz=pjz;
建立.m 文件，并命名为 disici02_baoji
function baojipjz=disici02_baoji()
load baoji.txt;
cs=baoji;
a=size(cs);
dian=max(cs(:,9));
pjz=zeros(7,dian);%每个城市几个点
for j=1:1:7
    num1=0;num2=0;num3=0;num4=0;num5=0;num6=0;num7=0;num8=0;
    for i=1:1:a(1)
        b=cs(i,9);
        switch b
            case 1,
                num1=num1+1;
                pjz(j,b)=pjz(j,b)+cs(i,j);
            case 2,
                num2=num2+1;
                pjz(j,b)=pjz(j,b)+cs(i,j);

```

```

        case 3,
            num3=num3+1;
            pjz(j,b)=pjz(j,b)+cs(i,j);
        case 4,
            num4=num4+1;
            pjz(j,b)=pjz(j,b)+cs(i,j);
        case 5,
            num5=num5+1;
            pjz(j,b)=pjz(j,b)+cs(i,j);
        case 6,
            num6=num6+1;
            pjz(j,b)=pjz(j,b)+cs(i,j);
        case 7,
            num7=num7+1;
            pjz(j,b)=pjz(j,b)+cs(i,j);
        case 8,
            num8=num8+1;
            pjz(j,b)=pjz(j,b)+cs(i,j);
    end
end
pjz(j,1)=pjz(j,1)/num1;pjz(j,2)=pjz(j,2)/num2;pjz(j,3)=pjz(j,3)/num3;pjz(j,4)=pjz(
j,4)/num4;
pjz(j,5)=pjz(j,5)/num5;pjz(j,6)=pjz(j,6)/num6;pjz(j,7)=pjz(j,7)/num7;pjz(j,8)=pjz(
j,8)/num8;
end
baojipjz=pjz;

```

建立.m 文件，并命名为 disici02\_tongchuan

```
function tongchuanpjz=disici02_tongchuan()
```

```
load tongchuan.txt;
```

```
cs=tongchuan;
```

```
a=size(cs);
```

```
dian=max(cs(:,9));
```

```
pjz=zeros(7,dian);%每个城市几个点
```

```
for j=1:1:7
```

```
    num1=0;num2=0;num3=0;num4=0;
```

```
    for i=1:1:a(1)
```

```
        b=cs(i,9);
```

```
        switch b
```

```
            case 1,
```

```
                num1=num1+1;
```

```
                pjz(j,b)=pjz(j,b)+cs(i,j);
```

```
            case 2,
```

```
                num2=num2+1;
```

```

        pjz(j,b)=pjz(j,b)+cs(i,j);
    case 3,
        num3=num3+1;
        pjz(j,b)=pjz(j,b)+cs(i,j);
    case 4,
        num4=num4+1;
        pjz(j,b)=pjz(j,b)+cs(i,j);
    end
end
pjz(j,1)=pjz(j,1)/num1;pjz(j,2)=pjz(j,2)/num2;pjz(j,3)=pjz(j,3)/num3;pjz(j,4)=pjz(
j,4)/num4;
end
tongchuanpjz=pjz;

```

建立.m 文件，并命名为 disici02\_weinan

```

function weinanpjz=disici02_weinan()
load weinan.txt;
cs=weinan;
a=size(cs);
dian=max(cs(:,9));
pjz=zeros(7,dian);% 每个城市几个点
for j=1:1:7
    num1=0;num2=0;num3=0;num4=0;
    for i=1:1:a(1)
        b=cs(i,9);
        switch b
            case 1,
                num1=num1+1;
                pjz(j,b)=pjz(j,b)+cs(i,j);
            case 2,
                num2=num2+1;
                pjz(j,b)=pjz(j,b)+cs(i,j);
            case 3,
                num3=num3+1;
                pjz(j,b)=pjz(j,b)+cs(i,j);
            case 4,
                num4=num4+1;
                pjz(j,b)=pjz(j,b)+cs(i,j);
        end
    end
    pjz(j,1)=pjz(j,1)/num1;pjz(j,2)=pjz(j,2)/num2;pjz(j,3)=pjz(j,3)/num3;pjz(j,4)=pjz(
j,4)/num4;
end
weinanpjz=pjz;

```

建立.m 文件，并命名为 IDW。（反距离加权插值算法）

```
function [Z]=IDW(x,y,z,X,Y)
[m0,n0]=size(x);
[m1,n1]=size(X);
%生成距离矩阵 r
for i=1:m1
    for j=1:n1
r(m0*n1*(i-1)+m0*(j-1)+1:m0*n1*(i-1)+m0*(j),:)=sqrt((X(i,j)-x).^2+(Y(i,j)-y).^2);
    end
end
%定义插值函数
for i=1:m1
    for j=1:n1
        if find(r(m0*n1*(i-1)+m0*(j-1)+1:m0*n1*(i-1)+m0*(j),:)==0)

[m2,n2]=find(r(m0*n1*(i-1)+m0*(j-1)+1:m0*n1*(i-1)+m0*(j),:)==0);
            Z(i,j)=z(m2,n2);
        else
            numerator=sum(sum(z./r(m0*n1*(i-1)+m0*(j-1)+1:m0*n1*(i-1)+m0*(j),:)));
            denominator=sum(sum(1./r(m0*n1*(i-1)+m0*(j-1)+1:m0*n1*(i-1)+m0*(j),:)));
            Z(i,j)=numerator/denominator;
        end
    end
end
end
```

建立.m 文件，并命名为 mysvm

以西安市数据为例，只需将第二行中的第二个 xian 换为 xianyang、baoji、tongchuan、weinan，即可依次得到其余四座城市的相关数据及图。

```
load xian.txt,load xianyang.txt,load baoji.txt,load tongchuan.txt,load weinan.txt;
xian=weinan;
time=size(xian);
day=0;
XA0=zeros();
for j=1:1:7
    data=2013;
    for i=1:1:time(1)
        if data~=xian(i,8)
            data=xian(i,8);
            num=0;
            a=0;
        end
        if data==xian(i,8)
            num=num+1;
        end
    end
end
```

```

        a=a+xian(i,j);
        if num==1
            day=day+1;
        end
    end
    XA0(day,j)=a/num;
end
end
b=size(XA0);
XA1=zeros((b(1)/7),8);
for i=1:1:(b(1)/7)
    for j=1:1:7
        XA1(i,j)=XA0((b(1)/7)*(j-1)+i,j);
    end
    XA1(i,8)=i;
end
x=XA1(1:end-212,1:7);
martx=XA1(:,2:7);label=XA1(:,1);
% 随机产生训练集和测试集
n = randperm(size(matrix,1));
% 训练集
train_matrix = matrix(n(1:2000),:);
train_label = label(n(1:2000),:);
% 测试集
test_matrix = matrix(n(2000:end),:);
test_label = label(n(2000:end),:);
[Train_matrix,PS] = mapminmax(train_matrix');
Train_matrix = Train_matrix';
Test_matrix = mapminmax('apply',test_matrix',PS);
Test_matrix = Test_matrix';
% 寻找最佳 c/g 参数——交叉验证方法
[c,g] = meshgrid(-10:0.2:10,-10:0.2:10);
[m,n] = size(c);
cg = zeros(m,n);
eps = 10^(-4);
v = 5;
bestc = 1;
bestg = 0.1;
bestacc = 0;
for i = 1:m
    for j = 1:n
        cmd = ['-v ',num2str(v),' -t 2',' -c ',num2str(2^c(i,j)),' -g ',num2str(2^g(i,j))];
        cg(i,j) = svmtrain(train_label,Train_matrix,cmd);
    end
end

```



```

        if cg(i,j) > bestacc
            bestacc = cg(i,j);
            bestc = 2^c(i,j);
            bestg = 2^g(i,j);
        end
        if abs( cg(i,j)-bestacc )<=eps && bestc > 2^c(i,j)
            bestacc = cg(i,j);
            bestc = 2^c(i,j);
            bestg = 2^g(i,j);
        end
    end
end
cmd = ['-t 2','-c ',num2str(bestc),'-g ',num2str(bestg)];
% 创建/训练 SVM 模型
model = svmtrain(train_label,Train_matrix,cmd);
[predict_label_1,accuracy_1] = svmpredict(train_label,Train_matrix,model);
[predict_label_2,accuracy_2] = svmpredict(test_label,Test_matrix,model);
result_1 = [train_label predict_label_1];
result_2 = [test_label predict_label_2];
figure
plot(1:length(test_label),test_label,'r-*')
hold on
plot(1:length(test_label),predict_label_2,'b:o')
grid on
legend('真实类别','预测类别')
xlabel('测试集样本编号')
ylabel('测试集样本类别')
string = {'测试集 SVM 预测结果对比(RBF 核函数)';[accuracy = '
num2str(accuracy_2(1)) '%']};
title(string)
xianout0=xianout(349:349+1095);out0=cell2mat(xianout0);
xianyangout0=xianyangout(91:91+1095);out1=cell2mat(xianyangout0);
baojiout0=baojiout(91:91+1095);out2=cell2mat(baojiout0);
tongchuanout0=tongchuanout(91:91+1095);out3=cell2mat(tongchuanout0);
weinanout0=weinanout(91:91+1095);out4=cell2mat(weinanout0);
out0=out0';out1=out1';out2=out2';out3=out3';out4=out4';
subplot(3,2,1);
for i=1:1:size(out0)
    out0(i,8)=i;
end
plot(out0(:,8),out0(:,1));
set(gca,'Xtick',[0:90:1200]);
set(gca,'Ytick',[0:100:500]);
xlabel('时间/天');ylabel('AQI 指标');title('西安市');

```

```

subplot(3,2,2);
for i=1:1:size(out1)
    out1(i,8)=i;
end
plot(out1(:,8),out1(:,1));
set(gca,'Xtick',[0:90:1200]);
set(gca,'Ytick',[0:100:500]);
xlabel('时间/天');ylabel('AQI 指标');title('咸阳市');
subplot(3,2,3);
for i=1:1:size(out2)
    out2(i,8)=i;
end
plot(out2(:,8),out2(:,1));
set(gca,'Xtick',[0:90:1200]);
set(gca,'Ytick',[0:100:500]);
xlabel('时间/天');ylabel('AQI 指标');title('宝鸡市');
subplot(3,2,4);
for i=1:1:size(out3)
    out3(i,8)=i;
end
plot(out3(:,8),out3(:,1));
set(gca,'Xtick',[0:90:1200]);
set(gca,'Ytick',[0:100:500]);
xlabel('时间/天');ylabel('AQI 指标');title('铜川市');
subplot(3,2,5);
for i=1:1:size(out4)
    out4(i,8)=i;
end
plot(out4(:,8),out4(:,1));
set(gca,'Xtick',[0:90:1200]);
set(gca,'Ytick',[0:100:500]);
xlabel('时间/天');ylabel('AQI 指标');title('渭南市');

```

建立.m 文件，并命名为 disici03

%原始数据

```

xx1=disici02_xian;num1=size(xx1);
xx2=disici02_xianyang;num2=size(xx2);
xx3=disici02_baoji;num3=size(xx3);
xx4=disici02_tongchuan;num4=size(xx4);
xx5=disici02_weinan;num5=size(xx5);
for i=1:1:7
    a1=sum(xx1(i,:));
    xianzpj(1,i)=a1/num1(2);
    a2=sum(xx2(i,:));

```

```

        xianyangzpj(1,i)=a2/num2(2);
        a3=sum(xx3(i,:));
        baojizpj(1,i)=a3/num3(2);
        a4=sum(xx4(i,:));
        tongchuanzpj(1,i)=a4/num4(2);
        a5=sum(xx5(i,:));
        weinanzpj(1,i)=a5/num5(2);
    end
    x(1,:)=xianzpj;
    x(2,:)=xianyangzpj;
    x(3,:)=baojizpj;
    x(4,:)=tongchuanzpj;
    x(5,:)=weinanzpj;
    b=size(x);
    m=b(1);n=b(2);
    %参考序列
    x0=xianzpj;
    %均值初始化
    for i=1:n
        avg(i)=0;
    end
    for i=1:m
        for j=1:n
            avg(j)=avg(j)+x(i,j);
        end
    end
    %求均值序列
    for i=1:n
        avg(i)=avg(i)/m;
    end
    %均值化
    for j=1:m
        for i=1:n
            x(i,j)=x(j,i)/avg(i);
        end
    end
    %参考序列均值化
    for i=1:n
        x0(i)=x0(i)/avg(i);
    end
    %求序列差
    for j=1:m
        for i=1:n
            delta(j,i)=abs(x(j,i)-x0(i));

```

```

        end
    end
    max=delta(1,1);
    for j=1:m
        for i=1:n
            if delta(j,i)>max
                max=delta(j,i);
            end
        end
    end
    end
    %求两级差
    min=0;
    %计算关联系数及相关度
    for j=1:m
        xgd(j)=0;
        for i=1:n
            glxs(j,i)=0.5*max/(0.5*max+delta(j,i));
            xgd(j)=xgd(j)+glxs(j,i);
        end
        xgd(j)=xgd(j)/n;
    end
    xgd

```