

灰色系统模型及预测

灰色系统模型在数据处理和预测中经常使用。

灰色系统理论建模特点：原始数据必须等时间间距。

处理思路：首先对原始数据进行累加，弱化原始时间序列数据的随机因素.然后建立生成数的微分方程。

GM(1,1)模型是灰色系统理论中的单序列一阶灰色微分方程。

设已知序列为 $x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)$

做一次累加 AGO (Accumulated Generating Operation) 生成新序列：

$x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)$

其中 $x^{(1)}(1) = x^{(0)}(1)$, $x^{(1)}(2) = x^{(1)}(1) + x^{(0)}(2)$, \cdots , $x^{(1)}(n) = x^{(1)}(n-1) + x^{(0)}(n)$

$$\text{也即 } x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i) \quad k = 1, 2, \dots, n$$

生成均值序列:

$$z^{(1)}(k) = \alpha x^{(1)}(k) + (1 - \alpha)x^{(1)}(k-1) \quad k = 2, 3, \dots, n \quad (1)$$

其中 $0 \leq \alpha \leq 1$ 。通常可取 $\alpha = 0.5$

建立灰微分方程:

$$x^{(0)}(k) + az^{(1)}(k) = b \quad k = 2, 3, \dots, n \quad (2)$$

相应的 GM(1,1)白化微分方程为:

$$\frac{dx^{(1)}}{dt} + ax^{(1)}(t) = b \quad (3)$$

将方程(2)变形为:

$$-az^{(1)}(k)+b=x^{(0)}(k) \quad k=2,3,\dots,n \quad (4)$$

其中 a, b 为待定模型参数。

将方程组(4)采用矩阵形式表达为:

$$\begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \dots & \dots \\ -z^{(1)}(n) & 1 \end{bmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(n) \end{pmatrix} \quad (5)$$

$$\text{即: } X\beta = Y \quad (6)$$

$$\text{其中 } X = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \dots & \dots \\ -z^{(1)}(n) & 1 \end{bmatrix}, \quad \beta = \begin{pmatrix} a \\ b \end{pmatrix}, \quad Y = \begin{pmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(n) \end{pmatrix}$$

解方程(6)得到最小二乘解为:

$$\hat{\beta} = (a, b)^T = (X^T X)^{-1} X^T Y \quad (7)$$

求解微分方程(3)得到 GM(1,1)模型的离散解:

$$\hat{x}^{(1)}(k) = [x^{(0)}(1) - \frac{b}{a}]e^{-\alpha(k-1)} + \frac{b}{a} \quad k = 2, 3, \dots, n \quad (8)$$

还原为原始数列, 预测模型为:

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1) \quad k = 2, 3, 4, \dots, n \quad (9)$$

将式(8)代入式(9)得

$$\hat{x}^{(0)}(k) = [x^{(0)}(1) - \frac{b}{a}]e^{-a(k-1)}(1 - e^a) \quad k = 2, 3, 4, \dots, n \quad (10)$$

GM(1.1)模型与统计模型相比,具有两个显著优点:

一是灰色模型即使在少量数据情况下建立的模型,精度也会很高;
二是灰色模型从其机理上讲,越靠近当前时间点精度会越高,因此灰色模型的预测功能优于统计模型。

灰色系统建模实际上是一种以数找数的方法,从系统的一个或几个离散数列中找出系统的变化关系,试图建立系统的连续变化模型。

实例计算

2003年的SARS疫情对中国部分行业的经济发展产生了一定的影响，特别是对部分疫情严重省市的相关行业所造成的影响是明显的。经济影响分为直接经济影响和间接影响。很多方面难以进行定量评估。现就某市SARS疫情对商品零售业的影响进行定量的评估分析。

表 1 商品零售额(单位：亿元)

年代	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
1997	83.0	79.8	78.1	85.1	86.6	88.2	90.3	86.7	93.3	92.5	90.9	96.9
1998	101.7	85.1	87.8	91.6	93.4	94.5	97.4	99.5	104.2	102.3	101.0	123.5
1999	92.2	114.0	93.3	101.0	103.5	105.2	109.5	109.2	109.6	111.2	121.7	131.3
2000	105.0	125.7	106.6	116.0	117.6	118.0	121.7	118.7	120.2	127.8	121.8	121.9
2001	139.3	129.5	122.5	124.5	135.7	130.8	138.7	133.7	136.8	138.9	129.6	133.7
2002	137.5	135.3	133.0	133.4	142.8	141.6	142.9	147.3	159.6	162.1	153.5	155.9
2003	163.2	159.7	158.4	145.2	124	144.1	157.0	162.6	171.8	180.7	173.5	176.5

解答：

SARS发生在2003年4月。因此我们可根据1997年到2002年的数据，预测2003年的各月的零售额，并与实际的零售额进行。从而判断2003年到底哪几个月受到SARS影响，并给出影响大小的评估。

将 1997--2002 年的数据记作矩阵 $A_{6 \times 12}$ ，代表 6 年 72 个数据。

$$\text{计算各年平均值 } x^{(0)}(i) = \frac{1}{12} \sum_{j=1}^{12} a_{ij} \quad i = 1, 2, \dots, 6$$

$$\text{得到 } x^{(0)} = (87.6167, 98.5000, 108.4750, 118.4167, 132.8083, 145.4083)$$

$$\text{计算累加序列 } x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i) \quad k = 1, 2, \dots, 6$$

$$\text{得到 } x^{(1)} = (87.6167, 186.1167, 294.5917, 413.0083, 545.8167, 691.2250)$$

生成均值序列:

$$z^{(1)}(k) = \alpha x^{(1)}(k) + (1 - \alpha)x^{(1)}(k - 1) \quad k = 2, 3, \dots, n$$

这里取 $\alpha = 0.4$ 。

$$z^{(1)} = (0, 127.0167, 229.5067, 341.9583, 466.1317, 603.9800)$$

建立灰微分方程:

$$x^{(0)}(k) + \alpha z^{(1)}(k) = b \quad k = 2, 3, \dots, 6$$

相应的 GM(1,1)白化微分方程为:

$$\frac{dx^{(1)}}{dt} + \alpha x^{(1)}(t) = b$$

求解微分方程得到 $a = -0.0993$, $b = 35.5985$

GM(1,1)模型的离散解:

$$\hat{x}^{(1)}(k) = [x^{(0)}(1) - \frac{b}{a}]e^{-a(k-1)} + \frac{b}{a} \quad k = 2, 3, \dots, 6$$

还原为原始数列预测模型为:

$$\hat{x}^{(0)}(k) = [x^{(0)}(1) - \frac{b}{a}]e^{-a(k-1)}(1 - e^a) \quad k = 2, 3, 4, \dots, 6$$

取 $k = 7$, 得到 2003 年销售额平均值的预测值为: $\hat{x}^{(0)}(7) = 162.8793$ 。

则全年总销售额为 $T = 12 \times \hat{x}^{(0)}(7) = 1954.55$ 。

下面估计2003年各月的销售额。

根据前 6 年数据估计各月销售额的比例 r_1, r_2, \dots, r_{12} 。其中 $r_j = \frac{\sum_{i=1}^6 a_{ij}}{\sum_{i=1}^6 \sum_{j=1}^{12} a_{ij}}$

计算得到 $r = (0.0794, 0.0807, 0.0749, 0.0786, 0.0819, 0.0818, 0.0845, 0.0838, 0.0872, 0.0886, 0.0866, 0.0920)$

从而 2003 年各月销售额预测为：

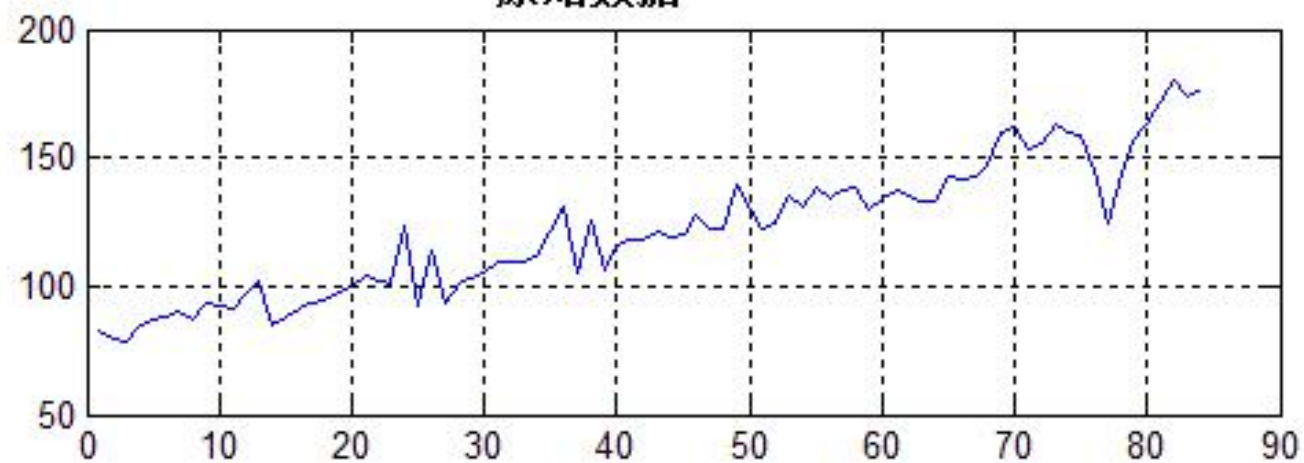
155.2, 157.7, 146.4, 153.5, 160.1, 159.8, 165.1, 163.8, 170.5, 173.1, 169.3, 179.8

表 2 2003 年商品实际销售额和预测（亿元）

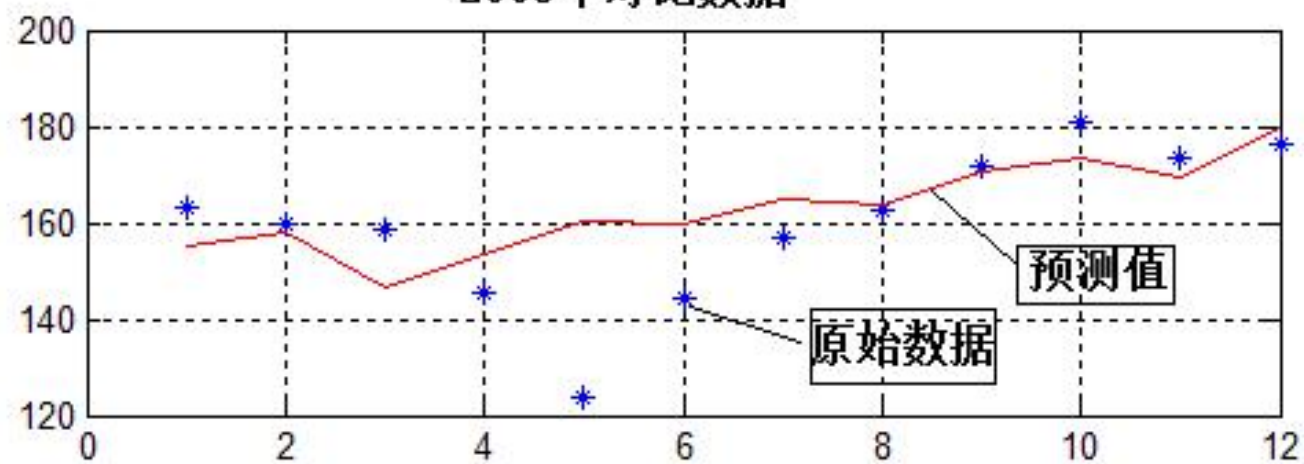
月份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
预测	155.2	157.7	146.4	153.5	160.1	159.8	165.1	163.8	170.5	173.1	169.3	179.8
实际	163.2	159.7	158.4	145.2	124	144.1	157.0	162.6	171.8	180.7	173.5	176.5

结果分析：2003年4、5、6月实际销售额为145.2、124、144.1亿元，统计部门这三个月受SRAS影响最严重，估计为62亿元。
这三个月我们的预测值总和与实际值总和之差为60.22亿元。与统计吻合，说明我们所建模型合理。

原始数据



2003年对比数据



Matlab实现程序
%1997--2003年数据

```
A=[83.0, 79.8,78.1,85.1,86.6,88.2,90.3,86.7,93.3,92.5,90.9,96.9;  
101.7,85.1,87.8,91.6,93.4,94.5,97.4,99.5,104.2,102.3,101.0,123.5;  
92.2,114.0,93.3,101.0,103.5,105.2,109.5,109.2,109.6,111.2,121.7,131.3;  
105.0,125.7,106.6,116.0,117.6,118.0,121.7,118.7,120.2,127.8,121.8,121.9;  
139.3,129.5,122.5,124.5,135.7,130.8,138.7,133.7,136.8,138.9,129.6,133.7;  
137.5,135.3,133.0,133.4,142.8,141.6,142.9,147.3,159.6,162.1,153.5,155.9;  
163.2,159.7,158.4,145.2,124,144.1,157.0,162.6,171.8,180.7,173.5,  
176.5];  
T=A(1:6,1:12);  
x0=mean(T');%对前6年求平均  
x1=zeros(size(x0));  
n=length(x0);  
x1(1)=x0(1);
```

```
for i=2:n
    x1(i)=x1(i-1)+x0(i); %累积求和
end
z=zeros(size(x0));
af=0.4; %参数
for i=2:n
    z(i)=af*x1(i)+(1-af)*x1(i-1);
end
Y=zeros(n-1,1); B=zeros(n-1,2);
for i=2:n
    Y(i-1,1)=x0(i);
    B(i-1,1)=-z(i);
    B(i-1,2)=1;
end
```

```
Para=inv(B'*B)*B'*Y; %计算参数
a=Para(1);
b=Para(2);
Pred=(x0(1)-b/a)*exp(-a*n)*(1-exp(a));
%预测第n+1年数值(2003年)
Total=12*Pred; %2003年总平均值
%估计各月所占比重;
r=sum(T)/sum(sum(T));
%预测2003年各月销售量
Px=Total*r;
fprintf('输出2003年预测值与实际值.\n');
```

```
for i=1:12
    fprintf('%5d ',i);
end
fprintf('\n');
for i=1:12
    fprintf('%6.1f ',Px(i)); %输出2003年预测值
end
fprintf('\n');
for i=1:12
    fprintf('%6.1f ',A(7,i)); %输出2003年实际值
end
fprintf('\n');
Error=sum(Px(4:6))-sum(A(7,4:6));
fprintf('2003年4,5,6月SARS导致减少销售额%6.2f亿元\n',Error);
```


%作图

```
subplot(2,1,1);
```

```
PA=[A(1,:),A(2,:),A(3,:),A(4,:),A(5,:),A(6,:),A(7,:)];%变为一行数据
```

```
plot(PA); grid on
```

```
title('原始数据');
```

```
subplot(2,1,2);
```

```
plot(1:12,A(7,:), 'b*', 1:12, Px, 'r');
```

```
title('2003年对比数据');
```

```
grid on
```

谢 谢！