灰色系统理论及其应用

1.数据变换技术

为保证建模的质量与系统分析的正确结果，对收集来的原始数据必须进行数据变换和处理，使其消除量纲和具有可比性。

设有序列

$$x = x(x(1), x(2), \ldots, x(n))$$

则称映射

$$f : x \to y$$

$$f(x(k)) = y(k), k = 1,2,\ldots, n$$

为序列 y 的数据变换。

1）当

$$f(x(k)) = \frac{x(k)}{x(1)} = y(k), x(1) \neq 0$$

称 f 是初始化变换

2）当

$$f(x(k)) = \frac{x(k)}{\overline{x}} = y(k), \overline{x} = \frac{1}{n}\sum_{k=1}^{n} x(k)$$

称 f 是为均值化变换

3）当

$$f(x(k)) = \frac{x(k)}{\max\limits_{k} x(k)} = y(k)$$

称 f 是百分比变换

4）当

$$f(x(k)) = \frac{x(k)}{\min\limits_{k} x(k)} = y(k), \min_{k} x(k) \neq 0$$

称 f 是倍数变换

5）当

$$f(x(k)) = \frac{x(k)}{x_0} = y(k)$$

其中 x_0 为大于零的某个值，称 f 是归一化变换

6）当

$$f(x(k)) = \frac{x(k) - \min\limits_{k} x(k)}{\max\limits_{k} x(k)} = y(k)$$

称 f 是极差最大值化变换

7）当

$$f(x(k)) = \frac{x(k) - \min\limits_{k} x(k)}{\max\limits_{k} x(k) - \min\limits_{k} x(k)} = y(k)$$

称 f 是区间值化变换

2.关联分析

选取参考数列

$$x_0 = \{x_0(k) \mid k = 1,2,\ldots,n\} = (x_0(1), x_0(2), \ldots, x_0(n))$$

其中 k 表示时刻。假设有 m 个数列

$$x_i = \{x_i(k) \mid k = 1,2,\ldots,n\} = (x_i(1), x_i(2), \ldots, x_i(n)), i = 1,2,\ldots,m$$

则称：

$$\zeta_{0i}(k) = \frac{\min\limits_i \min\limits_k \mid x_0(t) - x_a(t) \mid + \rho \max\limits_i \max\limits_k \mid x_0(t) - x_a(t) \mid}{\mid x_0(t) - x_i(t) \mid + \rho \max\limits_i \max\limits_k \mid x_0(t) - x_a(t) \mid} \qquad （1）$$

为比较数列 x_i 对参考数列 x_0 在 k 时刻的关联系数，其中 $\rho \in [0,1]$ 为分辨系数。称（1）式中 $\min\limits_s \min\limits_t |x_0(t) - x_s(t)|$、$\max\limits_s \max\limits_t |x_0(t) - x_s(t)|$ 分别为两级最小差及两级最大差。

一般来讲，分辨系数 ρ 越大，分辨率越大；ρ 越小，分辨率越小。

（1）式定义的关联系数是描述比较数列与参考数列在某时刻关联程度的一种指标，由于各个时刻都有一个关联数，因此信息显得过于分散，不便于比较，为此我们给出

定义 3　称

$$r_i = \frac{1}{n}\sum_{k=1}^{n} \xi_i(k) \qquad （2）$$

为数列 x_i 对参考数列 x_0 的关联度。

由（2）易看出，关联度是把各个时刻的关联系数集中为一个平均值，亦即把过于分散的信息集中处理。利用关联度这个概念，我们可以对各种问题进行因素分析。考虑下面的问题。

案例

例 1　通过对某健将级女子铅球运动员的跟踪调查，获得其 1982 年至 1986 年每年最好成绩及 16 项专项素质和身体素质的时间序列资料，见表 2，试对此铅球运动员的专项成绩进行因素分析。

	1982	1983	1984	1985	1986
铅球专项成绩 x_0	13.6	14.01	14.54	15.64	15.69
4kg 前抛 x_1	11.50	13.00	15.15	15.30	15.02
4kg 后抛 x_2	13.76	16.36	16.90	16.56	17.30
4kg 原地 x_3	12.41	12.70	13.96	14.04	13.46
立定跳远 x_4	2.48	2.49	2.56	2.64	2.59
高　翻 x_5	85	85	90	100	105
抓　举 x_6	55	65	75	80	80
卧　推 x_7	65	70	75	85	90
3kg 前抛 x_8	12.80	15.30	16.24	16.40	17.05
3kg 后抛 x_9	15.30	18.40	18.75	17.95	19.30
3kg 原地 x_{10}	12.71	14.50	14.66	15.88	15.70
3kg 滑步 x_{11}	14.78	15.54	16.03	16.87	17.82
立定三级跳远 x_{12}	7.64	7.56	7.76	7.54	7.70
全　蹲 x_{13}	120	125	130	140	140
挺　举 x_{14}	80	85	90	90	95
30 米起跑 x_{15}	4”2	4”25	4”1	4”06	3”99
100 米 x_{16}	13”1	13”42	12”85	12”72	12”56

将上面的数据录入文本文件中，本程序将文本文件命名为“x”

程序：灰色关联分析	运行环境：Matlab2011a

```
clc,clear
load x.txt %把原始数据存放在纯文本文件 x.txt 中，其中把数据的"替换替换成.
for i=1:15
x(i,:)=x(i,:)/x(i,1); %标准化数据（对于正向指标而言）
end
for i=16:17
x(i,:)=x(i,1)./x(i,:); %标准化数据（对于负向指标而言）
end
data=x;
n=size(data,2); %求矩阵的列数，即观测时刻的个数
ck=data(1,:); %提出参考数列
bj=data(2:end,:); %提出比较数列
m2=size(bj,1); %求比较数列的个数
for j=1:m2
t(j,:)=bj(j,:)-ck;
end
mn=min(min(abs(t'))); %求最小差
mx=max(max(abs(t'))); %求最大差
```

rho=0.5; %分辨系数设置 ksi=(mn+rho*mx)./(abs(t)+rho*mx); %求关联系数 r=sum(ksi')/n %求关联度 [rs,rind]=sort(r,'descend') %对关联度进行排序

下面是通用的程序模板，将文本文件省去了，数据少就用下面这种，如果数据量多的话就用上面的程序，将数据录入文本文件中

程序：关联度	运行环境：Matlab2011a

```
x=[        ];
for i=1:?       %按要求设置，？取值为 x 的行数
        x(i,:)=x(i,:)/x(i,1);%标准化数据（这里）
end
data=x;
n=size(data,2);%求矩阵列数
ck=data(1,:);%求矩阵参考数列
bj=data(2:end,:);%找出比较数列
m2=size(bj,1);%求比较数列个数
for j=1:m2
        t(j,:)=bj(j,:)-ck;
end
mn=min(min(abs(t')));%求最小值
mx=max(max(abs(t')));%求最小差
rho=0.5;%分辨系数设置
ksi=(mn+rho*mx)./(abs(t)+rho*mx);%求关联系数
r=sum(ksi')/n%求关联度
[rs,rind]=sort(r,'descend')%对关联度进行排序
```

注：

应该指出的是，公式（1）中的$|x_0(k)-x_i(k)|$不能区别因素关联是正关联还是负关联，可采取下述办法解决这个问题。记

$$\sigma_i=\sum_{k=1}^{n}kx_i(k)-\sum_{k=1}^{n}x_i(k)\sum_{k=1}^{n}\frac{k}{n},\quad i=1,2,\cdots,n$$

则：

（1）当$\mathrm{sign}(\sigma_i)=\mathrm{sign}(\sigma_j)$，则$x_i$和$x_j$为正关联；

（2）当$\mathrm{sign}(\sigma_i)=-\mathrm{sign}(\sigma_j)$，则$x_i$和$x_j$为负关联。

3. 优势分析

当参考数列不止一个，被比较的因素也不止一个时，则需进行优势分析。假设有 m 个参考数列（宜称母因素），记为 $y_1, y_2, \cdots, y_m$，再假设有 l 个比较数列（亦称子因素），记为 $x_1, x_2, \cdots, x_l$。显然，每一个参考数列对 l 个比较数列有 l 个关联度，设 r_{ij} 表示比较数列 x_j 对参考数列 y_i 的关联度，可构造关联（度）矩阵 $R=(r_{ij})_{m\times l}$。根据矩阵 R 的各个元素的大小，可分析判断出哪些因素起主要影响，哪些因素起次要影响。起主要影响的因素称之为优势因素。再进一步，当某一列元素大于其它列元素时，称此列所对应的子因素为优势子因素；若某一行元素均大于其它行元素时，称此行所对应的母元素为优势母元素。例如，矩阵 R 的第 3 列元素大于其它各列元素，

$$r_{i3} > r_{ij}, \quad i=1,2,\cdots,m; \quad j \neq 3$$

则称 x_3 为优势子因素。

如果矩阵 R 的某个元素达到最大，则该行对应的母因素被认为是所有母因素中影响最大的。

为简单起见，先来讨论一下“对角线”以上元素为零的关联矩阵，例如

$$R=\begin{bmatrix} 0.8 & 0 & 0 & 0 & 0 & 0 \\ 0.6 & 0.5 & 0 & 0 & 0 & 0 \\ 0.7 & 0.7 & 0.3 & 0 & 0 & 0 \\ 0.4 & 0.6 & 0.7 & 0.9 & 0 & 0 \\ 0.3 & 0.8 & 0.2 & 0.7 & 0.504 & 0 \end{bmatrix}$$

因为第 1 列元素是满的，故称第 1 个子元素为潜在优势子因素。第 2 列元素中有一个元素为零，故称第 2 个子因素为次潜在优势子因素。余下类推。

当关联矩阵的“对角线”以下全都是零元素，则称第 1 个母因素为潜在优势母因素……，为了分析方便，我们经常把相对较小的元素近似为零，从而使关联矩阵尽量稀疏。

例题

例 2　某地区有 6 个母因素 y_i（$i=1,2,\cdots,6$），5 个子因素 x_j（$j=1,2,\cdots,5$）如下：

x_1：固定资产投资　　y_1：国民收入
x_2：工业投资　　y_2：工业收入
x_3：农业投资　　y_3：农业收入
x_4：科技投资　　y_4：商业收入
x_5：交通投资　　y_5：交通收入
y_6：建筑业收入

	1979	1980	1981	1982	1983
x_1	308.58	310	295	346	367
x_2	195.4	189.9	187.2	205	222.7

x_3	24.6	21	12.2	15.1	14.57
x_4	20	25.6	23.3	29.2	30
x_5	18.98	19	22.3	23.5	27.655
y_1	170	174	197	216.4	235.8
y_2	57.55	70.74	76.8	80.7	89.85
y_3	88.56	70	85.38	99.83	103.4
y_4	11.19	13.28	16.82	18.9	22.8
y_5	4.03	4.26	4.34	5.06	5.78
y_6	13.7	15.6	13.77	11.98	13.95

程序：优势分析	运行环境：Matlab2011a

```
clc,clear
load data.txt %把原始数据存放在纯文本文件 data.txt 中
n=size(data,1); %求矩阵的行数，即求所有因素的个数
m=size(data,2); %求矩阵的列数，即求观测时刻的个数
for i=1:n
data(i,:)=data(i,:)/data(i,1); %标准化数据
end
m1=6; m2=5; %m1 母因素的个数，m2 子因素的个数
ck=data(m2+1:n,:); %提出母因素数据
bj=data(1:m2,:); %提出子因素数据
for i=1:m1
for j=1:m2
t(j,:)=bj(j,:)-ck(i,:);
end
mn=min(min(abs(t'))); %求母因素 i 的最小差
mx=max(max(abs(t'))); %求母因素 i 的最大差
rho=0.5; %各个子因素对母因素的关联度（这里取ρ=0.5）
ksi=(mn+rho*mx)./(abs(t)+rho*mx); %求母因素 i 对所有因素的关联系数
rt=sum(ksi')/m; %求母因素 i 对所有因素的关联度
r(i,:)=rt;
end
r%结果矩阵
```

运行得到结果：

$$R=\begin{bmatrix} 0.802 & 0.761 & 0.557 & 0.810 & 0.936 \\ 0.689 & 0.666 & 0.529 & 0.885 & 0.800 \\ 0.891 & 0.858 & 0.579 & 0.577 & 0.675 \\ 0.678 & 0.663 & 0.568 & 0.780 & 0.731 \\ 0.811 & 0.774 & 0.565 & 0.804 & 0.921 \\ 0.743 & 0.766 & 0.562 & 0.607 & 0.632 \end{bmatrix}$$

结果分析（以该矩阵为例）：

从关联矩阵 R 可以看出：

（1）第 4 行元素都比较小，表明各种投资对商业收入影响不大，即商业是一个不太需要依赖外资而能自行发展的行业。从消耗投资上看，这是劣势，但从少投资多收入的效益观点看，商业是优势。

（2）$r_{15}=0.936$ 最大，表明交通投资的多少对国民收入的影响最大。也可以从此看出交通的影响。

（3）$r_{55}=0.921$ 仅次于 r_{15}，表明交通收入主要取决于交通投资，这是很自然的。

（4）在第 4 列中 $r_{24}=0.885$ 最大，表明科技对工业影响最大；而 $r_{34}=0.577$ 是该列中最小的，表明从全面来衡量，还没有使科技投资与农业经济挂上钩，即科技投资针对的不是农村需要的科技。

（5）第三行的前 3 个元素比价大，表明农业是个综合性行业，需其它方面的配合，例如，$r_{31}=0.891$ 表明固定资产投资能够较大地促进农业的发展。另外，$r_{32}=0.858$ 表明农业发展与工业投资也是密切相关的。