

指标合成的客观权重方法

通过一个实例讲解指标权重确定的方法。

我们根据采集到的指标对高校进行排名。

1、数据预处理

(1)统计数据的指标介绍

为全面反映各高校实际情况，选取了包括人才培养、科学研究及成果方面的 18 个指标。这 18 个指标具体为： X_1 – 授予博士学位， X_2 – 授予硕士学位， X_3 – 优博入选数， X_4 – 发明专利数， X_5 – 实用新型专利数， X_6 – 国家一等奖励， X_7 – 国家二等奖励数量， X_8 – 国家社科基金项目奖一等数量， X_9 – 国家社科基金项目奖二等数量， X_{10} – 国家社科基金项目奖三等数量， X_{11} – 教育部人文社科奖一等数量， X_{12} – 教育部人文社科奖二等数量，

X_{13} – 教育部人文社科奖三等数量, X_{14} – 国家基地总数和国家重点学科(国家重点实验室、国家工程研究中心、人文社科基地数之和), X_{15} – 经费总数(万元), X_{16} – SCI 总数, X_{17} – EI 总数, X_{18} – CSCD、CSSCI 总数。

表 1 部分高校某年的 18 个指标信息

学校	x1	x2	x3	x4	x5	x6	x7	x8	x9
北京大学	967	2212	45	139	10	1	16	4	3
中国人民大学	377	1634	16	0	0	0	0	1	4
清华大学	578	2990	48	947	177	2	27	0	0
北京交通大学	68	535	0	16	6	1	0	0	0
北京工业大学	37	429	1	69	17	0	1	0	0
北京航空航天大学	133	805	8	69	19	0	2	0	0
北京理工大学	201	757	2	14	8	0	0	0	0
北京科技大学	80	496	2	130	16	1	2	0	0

续表 1 各高校的 18 个指标信息(部分)

学校	x10	x11	x12	x13	x14	x15	x16	x17	x18
北京大学	8	4	9	15	99	31954.1	3837	468	4526
中国人民大学	4	3	7	15	36	3665	23	5	1340
清华大学	2	2	3	6	61	77163.2	5073	1698	4647
北京交通大学	0	0	0	0	5	11331	282	49	372
北京工业大学	0	0	0	0	2	15126.2	250	31	493
北京航空航天大学	0	0	0	0	12	47526.05	403	33	1189
北京理工大学	0	0	0	0	12	29713.81	390	200	1035
北京科技大学	0	0	0	1	8	21358.3	668	5	641

(2)数据的归一化处理

由于各个指标的取值范围不同，量纲与意义不同，为消除这些影响，需要对数据进行归一化处理。

设共有 n 所学校，每个学校共有 m 个指标，采集到的观测数据为： x_{ij}
($i = 1, 2, \dots, n; j = 1, 2, \dots, m$)，每个数值显然越大对排名越有利，因此归一化处理方法
可采用下式：

$$y_{ij} = \frac{x_{ij} - x_{jm}^*}{x_{jM}^* - x_{jm}^*} \quad (i = 1, 2, \dots, n; j = 2, \dots, m) \quad (1)$$

其中 $x_{jM}^* = \max_{1 \leq i \leq n} x_{ij}$ ， $x_{jm}^* = \min_{1 \leq i \leq n} x_{ij}$ 。

经过上面变换，所有数据都变到[0,1]，便于后续工作进行统一处理。

2. 客观权重确定的三种方法

(1) 熵权法

设 n 个学校的 m 个指标已经归一化处理, 数据为 y_{ij} ($i = 1, 2, \dots, n; j = 1, 2, \dots, m$), 其第 j 项指标的信息熵计算公式为:

$$E_j = -\frac{\sum_{i=1}^n p_{ij} \ln p_{ij}}{\ln n} \quad j = 1, 2, \dots, m \quad (2)$$

$$0 \leq E_j \leq 1 \quad \text{其中 } p_{ij} = \frac{y_{ij}}{\sum_{i=1}^n y_{ij}}, \text{ 若 } p_{ij} = 0, \text{ 则定义 } p_{ij} \ln p_{ij} = 0。$$

E_j 越小, 表明数据间差异越大, 因此提供的信息越大, 该指标权重就越大;

E_j 越大, 表明数据间彼此越接近, 因此提供的信息越少, 该指标权重就越小。

$$\text{客观权重计算式: } W_j = \frac{1 - E_j}{m - \sum_{j=1}^m E_j} \quad j = 1, 2, \dots, m \quad (3)$$

(2) 标准离差法

如果某个指标的标准差大，因此提供的信息越大，该指标权重就越大；反之，某个指标的标准差小，因此提供的信息越少，该指标权重就越小。利用标准差来计算各指标的客观权重，其计算式为：

$$W_j = \frac{\sigma_j}{\sum_{j=1}^m \sigma_j} \quad j = 1, 2, \dots, m \quad (4)$$

(3) CRITIC法确定权重

CRITIC法是Diakoulaki提出的一种客观赋权方法，确定权值以两个基本概念为基础：一是对比度，标准差越大权重相对越大。二是评价指标间的冲突性，当两个指标间有较强的正相关，说明两个指标冲突性低，两个指标反映的信息具有较大的相似性；当两个指标间有较强的负相关，说明两个指标冲突性大，两个指标反映的信息具有较大的不同。

确定第 j 个指标包含的信息量为:

$$c_j = \sigma_j \sum_{i=1}^m (1 - r_{ij}) \quad (j = 1, 2, \dots, m) \quad (5)$$

第 j 个指标权重为:

$$w_j = \frac{c_j}{\sum_{i=1}^m c_i} \quad (j = 1, 2, \dots, m) \quad (6)$$

3. 综合排名方法

采用(1)式对数据归一化处理后, 采用三种不同的客观权重, 对各学校的所有指标进行加权平均, 可以得到各学校的综合得分, 计算式为:

$$f_i = \sum_{j=1}^m w_j \cdot y_{ij} \quad (i = 1, 2, \dots, n) \quad (7)$$

其中权重 w_j 可由(3)、(4)或(6)式计算得到。

表 2 三种不同排名法得到的前 10 名高校

名次	熵权法	标准离差法	CRITIC 法
1	北京大学	北京大学	北京大学
2	清华大学	清华大学	清华大学
3	复旦大学	浙江大学	复旦大学
4	中国人民大学	复旦大学	浙江大学
5	武汉大学	武汉大学	武汉大学
6	北京师范大学	中国人民大学	中国人民大学
7	浙江大学	北京师范大学	北京师范大学
8	南京大学	南京大学	南京大学
9	吉林大学	吉林大学	吉林大学

从表2所排前10名结果来，不同客观权重确定方法排名结果存在一定差异，这说明排名结果是相对的，跟采用方法有关。从差异大小来看，标准离差法和CRITIC法排名结果更接近。

4. 不同排名结果差异的度量方法

设有两种不同方法得到的排名序列分别为 $B_1 = (k_1, k_2, \dots, k_n)$, $B_2 = (z_1, z_2, \dots, z_n)$ 。

其中 k_1, k_2, \dots, k_n 和 z_1, z_2, \dots, z_n 都是 $1, 2, \dots, n$ 的一个排列, k_i 代表第一种排名方法中第 i 个学校的名次, z_i 代表第二种排名方法中第 i 个学校的名次。

则第 i 个学校的名次在两种排名方法中的差异为:

$$d_i = |k_i - z_i| \quad (i = 1, 2, \dots, n) \quad (8)$$

两种排名方法的排名平均差异定义为:

$$d = \sum_{i=1}^n d_i / n \quad (9)$$

为反映两种排名方法名次差异的波动程度，采用排名差异构成的序列 $d_i (i = 1, 2, \dots, n)$ 的标准差来度量。其计算式如下：

$$v = \sqrt{\sum_{i=1}^n (d_i - d)^2 / n} \quad (10)$$

表 3 三种方法的排名差异

方法	(1,2)	(1,3)	(2,3)
平均差异度 d	5.1200	4.4400	1.7400
差异度的标准差	5.6089	4.8686	2.3034

其中 1 代表熵权法，2 代表标准偏差法，3 代表 CRITIC 法。

从表3的结果来看，标准偏差法与CRITIC法的排名差异最小，为1.74，各学校采用这两种方法的排名差异的波动程度也最小，为2.3034。从结果看，标准偏差法和CRITIC法确定的权重排名结果最为接近。

Matlab程序univer.m

%采用三种客观权重的方法对高校排名

%1.熵权法,2.标准离差法,3.CRITIC法

load university.txt;

x=university;

[m,n]=size(x);

%m---数据样本数,n---指标数

%数据的标准化

xmin=min(x);

xmax=max(x);

dis=xmax-xmin;

for i=1:n;

 x(:,i)=(x(:,i)-xmin(i))/dis(i);

end; %归一化

%1.熵权法

p=[]; E=[];

for k=1:n

 s=sum(x(:,k));

 p=x(:,k)/s;

 s=0.0;

 for i=1:m

 if(p(i)>0) s=s+p(i)*log(p(i)); end

 end

 E(k)=-s/log(m); %获得熵

end;

s=sum(E);

W1=(1-E)/(n-s); %熵权法权重

%2.标准离差法

au=mean(x););%均值

sig=std(x);%);%标准差

s=sum(sig);

w2=sig/s; %标准离差法获得的权重

%3.CRITIC法

r=corrcoef(x); %求相关系数

w=[];

for i=1:n

s=0.0;

for j=1:n

s=s+(1.0-r(i,j));

end

w3(i)=sig(i)*s;

end

s=sum(W3);

W3=W3/s; %CRITIC法获得的权重

```
w=W1; %选取一种权重
wg=100*w; %权重归一化
f=x*wg'; %计算各学校得分
ff=f;
s=1:m;
for i=1:m-1
    for j=i+1:m
        if(ff(j)>ff(i))
            temp=ff(i);    ff(i)=ff(j);    ff(j)=temp;
            temp=s(i);    s(i)=s(j);
            s(j)=temp; %s(i)为第i名的学校序号
        end
    end
end
```

```
fid=fopen('result.txt','w');
%输出前20名的: 名次,学校序号,得分
fprintf(fid,' 名次,学校序号,得分\r\n');
for i=1:20
    fprintf(fid,'%5d %5d %6.2f\r\n',i,s(i),ff(i));
end
fclose(fid);
```


谢 谢！