

文章编号: 1004-4574(2010) 04-0213-06

基于熵权的灰色聚类在洪灾评估中的应用

刘 力¹, 周建中², 杨 莉³, 张勇传²

(1 湖南省电力公司调度通信中心, 湖南 长沙 410007; 2 华中科技大学 水电与数字化工程学院, 湖北 武汉 430074;
3 湖南省电力试验研究院, 湖南 长沙 410007)

摘 要: 洪水灾害灾情识别的实质就是建立各个洪水灾害灾情决策指标与洪灾等级之间的非线性关系。鉴于洪灾成因机制的复杂性和发生过程的随机性, 在灰色聚类的基础上引入信息熵概念, 提出了具有典型指数白化权函数, 并采用加权平均原则的灰色信息熵聚类。该方法有效解决了“零权重”问题, 通过引用熵权所反应的数据本身的效用值来修正指标的权重系数, 充分利用了样本遗留信息, 并极大地保留了聚类系数的蕴涵信息。实例证明, 灰色信息熵聚类的评价过程直观简单, 结果合理有效, 能有效扩大灰色聚类在实际工程中的应用范围。

关键词: 灰色聚类; 信息熵; 洪水灾害灾情; 白化权函数; 加权平均原则

中图分类号: P426 616

文献标识码: A

Application of information entropy-based grey clustering to flood disaster evaluation

LIU Li¹, ZHOU Jian-zhong², YANG Li³, ZHANG Yong-chuan²

(1 Dispatch & Communication Center Hunan Electric Power Company, Changsha 410007, China; 2 College of Hydroelectric and Digitalization Engineering
Huazhong University of Science and Technology Wuhan 430074, China; 3 Hunan Electric Power Test and Research Institute, Changsha 410007, China)

Abstract Flood disaster loss evaluation is essentially to evaluate the damage degree caused by flood disaster according to flood disaster loss evaluation criterions, existing flood disaster loss evaluation index values and disaster loss evaluation model. In view of the complexity of the causes and the randomness of the occurring process of the flood disaster, we proposed a new grey clustering method by introducing the concept of information entropy into the conventional grey method. After constructing a typical exponential whitening weight function, this new method replaces the maximum clustering coefficient principle by using the weighted average principle. Based on the above characteristic, this method can effectively solve the “zero weight” problem, make full use of the simple data and largely reserve the information implied in the clustering weight by modifying the clustering weight with the values reflected by the entropy. Finally, an illustrative example was given to verify that this method is simple and reasonable and can extend application range of the grey clustering in engineering.

Key words grey clustering; information entropy; flood disaster loss; whitening weight function; weighted average principle

作为洪水灾害风险管理的重要内容之一, 洪水灾害灾情识别是指根据各类灾害灾情决策指标和灾害灾

收稿日期: 2009-03-20 修订日期: 2010-04-25

基金项目: 国家重点基础研究计划(973)课题(2007CB714107); 水利部公益性行业科研专项经费项目(200701008)

作者简介: 刘力(1983-), 男, 博士, 主要从事洪水灾害研究。E-mail: liul0324@163.com

通讯作者: 周建中, 男, 教授, 博士生导师。E-mail: jz_zhou@hust.edu.cn

情评价方法,对因洪水灾害造成的破坏程度进行评价,其结果可为洪水灾害管理提供风险决策依据,对防灾减灾实践具有重要的指导意义^[1]。洪水灾害灾情识别涉及自然、技术、经济、社会、政治等诸多因素,且实际洪灾中各单项指标的评价结果往往不确定、不相容,甚至是相互矛盾,直接利用灾情评价标准进行灾情评价十分困难。因此,学者们相继提出了灾度判别法、模糊综合评判、物元分析、神经网络、逻辑斯蒂曲线等洪水灾情评价模型^[2~6]。这些评价方法一般需要根据对所研究系统的认识程度预先确定模型的结构形式,因而具有各自的局限性。

灰色系统理论是我国著名学者邓聚龙教授于 20 世纪 80 年代提出并创立的一门新兴的系统工程学科^[7],它基于数学理论,以“部分信息已知,部分信息未知”的“小样本”、“贫信息”不确定性系统为研究对象,主要通过对“部分”已知信息的生成和开发,提取有价值的信息,利用灰色聚类分析、趋势预测及系统控制等方法实现对系统运行行为、演化规律的正确描述和有效监控。

针对洪水灾害灾情决策指标与洪灾等级之间的不确定性和洪灾样本信息可能存在的不完全性,本文采用灰色系统理论为基础,结合信息熵概念,建立了基于灰色信息熵聚类的洪水灾害识别方法,为洪水灾害灾情的识别提供依据,从而得到了一种较为科学合理而又切实可行的洪水灾害识别方法。

1 灰色聚类

所谓灰色聚类就是根据多个不同决策指标对待评事物的影响对其进行综合评价,以确定该事物是否满足给定的取舍准则^[8]。

1.1 聚类过程

构建 n 个待评事物 m 个决策指标的样本 $x_{ij} (i = 1, 2, \dots, n; j = 1, 2, \dots, m)$ 和 m 个决策指标 c 个聚类灰类的判断标准 $y_{jk} (k = 1, 2, \dots, c)$, 灰色聚类过程如下:

(1) 由于决策指标特征值的量纲不尽相同,在进行识别前要先将待评事物各指标的实测值和灰类进行无量纲化处理:

$$d_{ij} = x_{ij} / Y_{0j} \tag{1}$$

$$r_{jk} = y_{jk} / Y_{0j} \tag{2}$$

式中: d_{ij} 和 r_{jk} 分别为 x_{ij} 和 y_{jk} 的无量纲化处理值; Y_{0j} 为第 j 个指标的参考标准。

(2) 建立第 j 个指标属于第 k 个灰类的白化权函数 f_{jk} 。

(3) 聚类权值描述了决策指标对评价结果的影响程度,第 j 个指标关于第 k 个灰类的聚类权 ω_{jk} 计算公式如下:

$$\omega_{jk} = r_{jk} \setminus \sum_{j=1}^m r_{jk} \tag{3}$$

(4) 聚类系数刻画了待评事物对灰类的亲疏程度,第 i 个事物关于第 k 个灰类的聚类系数 ϵ_{ik} 为:

$$\epsilon_{ik} = \sum_{j=1}^m f_{jk}(d_{ij}) g \omega_{jk} \tag{4}$$

式中, $f_{jk}(d_{ij})$ 为由样本值 d_{ij} 求得的白化函数值。

(5) 按最大聚类系数原则确定样本类别,聚类系数行向量为 $\epsilon_{ik} = (\epsilon_{i1}, \epsilon_{i2}, \dots, \epsilon_{ic})$, 在各个行向量中,聚类系数最大者所对应的灰类即为此待评事物所属的类别。

1.2 白化权函数

白化权函数反映了决策指标对灰类的亲疏关系,其函数类型的确定需要研究者根据已知信息选择设计。灰色聚类通常采用的白化权函数类型有典型白化权函数、适中测度白化权函数、上限测度白化权函数和下限测度白化权函数,如图 1-4 所示:

使用以上白化权函数时,不相邻灰类的权重均视为零,即存在“零权重”现象,这就相当于人为的提高了相邻灰类的权重,有可能使得评价结果出现偏差。鉴于此并考虑到洪水灾害灾情的评价标准一般以区间形式表示,本文构造了典型指数白化权函数:

$$f_{j,1}(d_{ij}) = \begin{cases} 1 & d_{ij} \in (0, \lambda_{j,1}^a], \\ e^{w \times (\lambda_{j,1}^a - d_{ij}) / \lambda_{j,1}^a} & d_{ij} \in (\lambda_{j,1}^a, +\infty); \end{cases} \quad (5)$$

$$f_{j,k}(d_{ij}) = \begin{cases} e^{w \times (d_{ij} - \lambda_{j,k}^a) / d_{i,j}} & d_{ij} \in (0, \lambda_{j,k}^a], \\ 1 & d_{ij} \in (\lambda_{j,k}^a, \lambda_{j,k}^b], \\ e^{w \times (\lambda_{j,k}^b - d_{ij}) / \lambda_{j,k}^b} & d_{ij} \in (\lambda_{j,k}^b, +\infty); \end{cases} \quad (6)$$

$$f_{j,c}(d_{ij}) = \begin{cases} e^{w \times (d_{ij} - \lambda_{j,c}^a) / d_{i,j}} & d_{ij} \in (0, \lambda_{j,c}^a], \\ 1 & d_{ij} \in (\lambda_{j,c}^a, +\infty). \end{cases} \quad (7)$$

式中: $[\lambda_{j,b}^a, \lambda_{j,k}^b]$ 为灰类的阈值域; $[\lambda_{j,1}^a, \lambda_{j,c}^a]$ 为全体灰类阈值边界; w 为衰减因子, 其大小决定了相邻灰类关联程度弱化的速度。

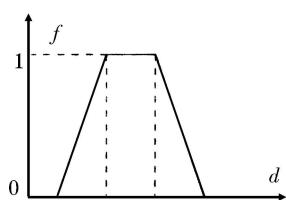


图 1 典型白化权函数
Fig 1 Typical whitening weight function

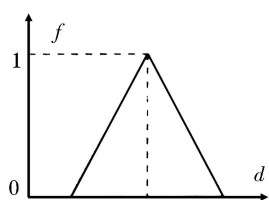


图 2 适中测试白化权函数
Fig 2 Moderate measure whitening weight function

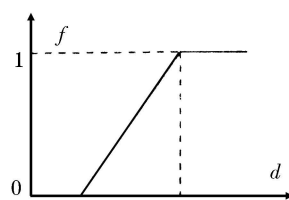


图 3 上限测度白化权函数
Fig 3 Upper measure whitening weight function

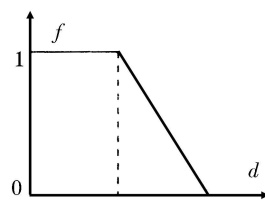


图 4 下限测度白化权函数
Fig 4 Lower measure whitening weight function

典型指数白化权函数摒弃传统灰色聚类忽略不相邻等级之间影响的做法, 样本值与各级非零权重之间存在一一对应关系, 能有效解决“零权重”问题。

1.3 加权平均原则

最大聚类系数原则是一种常用的分析方法, 在决策和评价方面有广泛的应用, 然而, 最大聚类系数原则本身存在大量信息丢失的缺陷, 容易因此失效而引起判别结果出现偏差。

加权平均原则的基本思想是以等级 $z = (z_1, z_2, \dots, z_c)$ 作为变量, 如, $z = (1, 2, \dots, c)$ 评价结果 $\varepsilon_{i,k}$ 作为权重, 计算所得的 A_i 值即为事物 i 所隶属的等级特征值, 公式如下:

$$A_i = \frac{\sum_{k=1}^c (z_k g \varepsilon_{i,k})}{\sum_{k=1}^c \varepsilon_{i,k}} \quad (8)$$

加权平均原则尽可能地保留了聚类系数向量中的全部信息或大部分信息, 可以一定程度上解决最大聚类系数原则失效的问题。

2 熵权法确定权重系数

熵的概念源于热力学, 表示任何一种能量在空间中分布的均匀程度, 是描述和研究自然界中广泛存在的运动形式转化的不可逆性的一个极其重要的概念, 后由数学家 Shannon 于 20 世纪 40 年代将其引入信息论, 形成了信息熵的概念, 以反映信息无序化的程度^[9]。从信息角度考虑, 信息熵代表了各指标在问题中提供有效信息的多寡程度, 可以用于度量数据所提供的有效信息程度和确定权重。但需要说明的是, 信息熵并不是表示指标在决策评估问题中重要程度的系数, 而是指在事物各评价指标值确定的情况下, 各指标在竞争意义上的相对激烈程度。因此, 熵权法确定权重系数, 其本质是利用指标信息的效用值, 即指标值在各事物间的差异程度来修正指标的权重, 具体步骤如下:

(1) 原始数据归一化处理以消除指标量纲不同带来的影响:

$$b_{ij} = \left[x_{ij} - \min_{i \in \{1, \dots, n\}} \{x_{ij}\} \right] / \left[\max_{i \in \{1, \dots, n\}} \{x_{ij}\} - \min_{i \in \{1, \dots, n\}} \{x_{ij}\} \right] \quad (9)$$

式中, $\max_{i \in \{1, \dots, n\}} \{x_{ij}\}$ 和 $\min_{i \in \{1, \dots, n\}} \{x_{ij}\}$ 分别为所有待评事物同一决策指标实测值中的最大值和最小值。

(2) 根据信息熵的定义, 在有 n 个待评事物 m 个决策指标的评估问题中, 可以确定第 j 个决策指标的熵

值:

$$h_j = -pg \sum_{i=1}^m (g_{i,j} g \ln g_{i,j}) . \tag{10}$$

式中 $g_{i,j} = b_{i,j} / \sum_{i=1}^n b_{i,j}$, $p = 1/\ln n$. 当 $g_{i,j} = 0$ 时, 令 $g_{i,j} \ln g_{i,j} = 0$

(3) 计算第 j 个决策指标的熵权:

$$\theta_j = (1 - h_j) \left/ \left(m - \sum_{j=1}^m h_j \right) \right. , \tag{11}$$

其中, $\theta_j \in [0, 1]$, 且 $\sum_{j=1}^m \theta_j = 1$

(4) 最后, 利用熵权修正指标权重系数矩阵 ω :

$$\omega'_{i,j} = \theta_j g \omega_{i,j} \left/ \sum_{j=1}^m (\theta_j g \omega_{i,j}) \right. . \tag{12}$$

3 基于熵权的灰色聚类方法

综上所述, 基于灰色信息熵聚类的洪水灾害识别步骤可归纳为:

- 步骤 1 根据实测资料构建样本 $x_{i,j}$ 和判断标准 $y_{j,k}$, 并设置各决策指标参考标准 $Y_{0,j}$;
- 步骤 2 根据式 (1)、式 (2) 分别对样本指标值 $x_{i,j}$ 和洪灾等级标准 $y_{j,k}$ 无量纲化, 得到 $d_{i,j}$ 和 $r_{j,k}$;
- 步骤 3 将 $d_{i,j}$ 代入式 (5) - (7) 得到白化权函数 $f_{j,k}$;
- 步骤 4 $r_{j,k}$ 将代入式 (3) 得到聚类权 $\omega_{j,k}$;
- 步骤 5 利用式 (9) 计算出 $x_{i,j}$ 的样本归一化值 $b_{i,j}$;
- 步骤 6 通过式 (10) 计算出各个评价指标的熵值 h_j , 并进而由式 (11) 得出各自的熵权 θ_j ;
- 步骤 7 将 θ_j 代入式 (12) 即可修正聚类权 $\omega_{j,k}$ 得到 $\omega'_{i,j}$;
- 步骤 8 将 $f_{j,k}$ 和 $\omega'_{i,j}$ 代入式 (4) 后即得聚类系数 $\varepsilon_{i,k}$;
- 步骤 9 通过式 (8) 计算得出聚类结果, 对结果进行判断分析以识别洪灾等级。

4 算例分析

河南省地处中纬度, 介于亚热带和亚热带边缘地区, 全省年平均气温在 12~ 15℃ 之间, 多年平均无霜期为 180~ 240d, 多年平均降水量 800mm, 总面积为 16.7 万 km², 其中耕地面积 694.4 万 hm², 是我国重要的农业大省。同时, 河南省农业自然灾害也较多。据 1950- 1990 年成灾面积资料 (缺 1966- 1977 年资料) 统计, 河南省各种自然灾害成灾面积累计达 9184 万 hm², 其中洪水灾害占 36.5%。1950- 1990 年 41a 期间, 洪水灾害多年平均成灾面积为 77.3 万 hm², 占全省耕地面积的 11.1%, 多年平均粮食损失为 12 亿 kg, 多年平均经济损失达 20.3 亿元。

文献 [10] 把成灾面积 (hm²) 和直接经济损失 R (亿元) 作为洪水灾害灾情等级决策指标, 并利用河南省 1950- 1990 年共 41a 的实际系列资料作频率分析, 得到该省洪灾等级标准和重现期如表 1 所示:

表 1 河南省洪水灾害灾情等级标准和重现期

Table 1 Gradation criterion and return period of flood disaster in Henan Province of China

洪灾指标	普通洪灾	较大洪灾	大洪灾	特大洪灾
重现期 /a	< 2	2~ 5	5~ 20	> 20
成灾面积 /hm ²	< 46.7	46.7~ 136.7	136.7~ 283.3	> 283.3
直接经济损失 R /亿元	< 9.5	9.5~ 31.0	31~ 85.0	> 85.0

根据文献 [2] 中提供的方法随机产生 23 组各洪灾指标值及其对应的标准洪灾等级样本系列, 如表 2 所示。同时, 采用灰色信息熵聚类对表 2 中提供的样本进行洪水灾害识别, 聚类结果见表 2

表 2 样本等级标准值和灰色信息熵聚类结果比较

Table 2 Comparison results between evaluation criterion for samples of flood disaster loss and values of grey clustering

样本 序号	评价指标		洪灾等级		样本 序号	评价指标		洪灾等级	
	S	R	标准值	聚类结果		S	R	标准值	聚类结果
1	38.7	7.9	1.0	1.306	13	259.1	76.1	3.0	3.382
2	38.5	7.8	1.0	1.297	14	200.1	54.4	3.0	3.028
3	32.1	6.5	1.0	1.138	15	280.1	83.8	3.0	3.480
4	24.2	4.9	1.0	1.023	16	236.1	67.6	3.0	3.255
5	36.4	7.4	1.0	1.244	17	157.3	38.6	3.0	2.701
6	46.7	9.5	1.5	1.500	18	283.3	85.0	3.5	3.494
7	97.6	21.7	2.0	2.152	19	556.9	167.1	4.0	3.979
8	60.4	12.8	2.0	1.785	20	649.5	194.9	4.0	3.994
9	112.6	25.2	2.0	2.291	21	602.3	180.7	4.0	3.989
10	56.2	11.8	2.0	1.709	22	446.5	134.0	4.0	3.909
11	80.6	17.6	2.0	2.009	23	694.9	208.5	4.0	3.997
12	136.7	31.0	2.5	2.505					

表 3 给出了该样本系列标准等级值与灰色信息熵聚类评价结果之间的误差分析结果, 并与文献 [11] 给出的投影寻踪评价模型、逻辑斯蒂曲线评价模型以及经典灰色聚类分析得出的结果进行了比较。表 2 和表 3 说明, 灰色信息熵聚类所描述的河南省各洪水灾害灾情决策指标与洪灾等级间的非线性关系令人满意。

表 3 样本等级评价结果比较

Table 3 Error analysis between evaluation criterion grades of flood disaster loss and grades of different methods

评价 方法	误差绝对值 (级) 落在下列区间的百分比 %						平均绝对 误差 / 级	平均相对 误差 %
	[0, 0.1]	[0, 0.2]	[0, 0.3]	[0, 0.4]	[0, 0.5]	[0, 0.6]		
投影寻踪模型	34.78	43.48	60.87	82.61	95.65	100.00	0.22	13.31
逻辑斯蒂曲线	39.13	47.83	60.87	78.26	95.65	100.00	0.21	12.21
经典灰色聚类	26.09	34.78	52.17	91.30	100.00	100.00	0.24	12.56
灰色信息熵聚类	47.83	56.52	86.96	95.65	100.00	100.00	0.16	8.72

河南省 1950-1990 年 41a 中发生大的洪水灾年有 1950、1954、1956、1957、1963、1964、1975、1982、1984 等 9a 用灰色信息熵聚类对这 9 次大的洪水灾害损失资料进行灾情等级评估, 结果见表 4。表 4 也同时给出了其它方法的相应评价结果。

表 4 洪水灾害灾情等级评价结果

Table 4 Evaluation results of flood disaster loss samples by using different methods

洪灾 年份	评价指标		评价方法			
	S	R	投影寻踪模型	逻辑斯蒂曲线	经典灰色聚类	灰色信息熵聚类
1950	72.92	9.900	1.574	1.704	1.668	1.683
1954	148.13	20.656	2.156	2.527	2.311	2.385
1956	203.92	27.521	2.559	3.046	2.665	2.651
1957	179.10	24.858	2.389	2.831	2.514	2.54
1963	375.46	94.927	3.726	3.830	3.531	3.681
1964	301.24	47.836	3.233	3.619	3.197	3.254
1975	141.97	116.439	3.368	2.553	2.972	3.131
1982	279.84	121.127	3.699	3.559	3.51	3.632
1984	172.06	51.619	2.750	2.789	2.806	2.897

对照表 1 的洪水灾情评价标准, 灰色信息熵聚类的评价结果较其它评价方法更为合理、精确。例如, 1957 年成灾面积为 179.1 km², 直接经济损失为 24.858 亿元, 分别落在第 3 等级和第 2 等级区间内, 两者均较接近第 2 等级与第 3 等级的边界值, 且距离比值基本相当, 因此灰色信息熵聚类判断 1957 年的灾情等级为 2.54 是比较合理的, 而投影寻踪模型和逻辑斯蒂曲线判断该年的灾情等级则分别为 2.389 和 2.831。又

如, 1982 的成灾面积为 279.84km^2 , 直接经济损失为 121.127 亿元, 分别落在第 3 等级和第 4 等级区间内, 前者非常接近第 3 等级与第 4 等级的边界值, 后者则超过第 3 等级与第 4 等级的边界值较多, 因此灰色信息熵聚类判断 1982 年的灾情等级为 3.632 较经典灰色聚类判断的 3.51 更为精确。

5 结论

(1) 为了模拟洪水灾害灾情决策指标与洪灾等级之间的非线性关系, 文中分别采用经典灰色聚类、投影寻踪模型、逻辑斯蒂曲线与灰色信息熵聚类进行洪水灾害灾情评价, 最后结果表明灰色信息熵聚类的评价结果与其它方法的评价结果基本相同甚至更为精确合理, 可以用于解决洪水灾害灾情识别以及类似问题;

(2) 根据灰色信息熵聚类的评价结果, 河南省 1950–1990 年 41a 中发生的这 9 次洪水灾害按照灾情等级从大到小的顺序依次为 1963 1982 1964 1975 1984 1956 1957 1954 年和 1950 年洪水灾害。其中, 1963 1982 年的洪水灾害属于特大洪灾, 1956 1957 1964 1975 1984 年的洪水灾害属于大洪灾, 1950 1954 年的洪水灾害属于较大洪灾;

(3) 灰色信息熵聚类构建的典型指数白化权函数有效解决了“零权重”问题; 信息熵的引入, 通过熵权所反应的数据效用值来修正决策指标的权重系数, 充分利用了样本的遗留信息; 采用加权平均原则代替最大聚类系数原则, 极大地保留了聚类系数的蕴涵信息, 有利于聚类结果的合理化。以上实例表明, 灰色信息熵聚类的方法简单, 概念清晰, 逻辑严密, 计算量小, 应用方便, 结果合理且精度较高, 具有一定的工程实用价值。

参考文献:

- [1] 王栋, 潘少明, 吴吉春, 等. 洪水风险分析的研究进展与展望[J]. 自然灾害学报, 2006 15 (1): 103–109
- [2] 金菊良, 魏一鸣, 杨晓华. 基于遗传算法的洪水灾情评估神经网络模型探讨[J]. 灾害学, 1998 13(2): 6–11
- [3] 于庆东. 灾度等级判别方法的局限性及其改进[J]. 自然灾害学报, 1993, 2(2): 8–10
- [4] 赵黎明, 王康, 邱佩华. 灾害综合评估研究[J]. 系统工程理论与实践, 1997, (3): 63–69
- [5] 李祚泳, 邓新民. 自然灾害的物元分析灾情评估模型初探[J]. 自然灾害学报, 1994 3(2): 28–33
- [6] 金菊良, 金保明, 杨晓华, 等. 建立洪水灾情等级模型的实用方案[J]. 灾害学, 2000 15(2): 1–61
- [7] 邓聚龙. 灰理论基础[M]. 武汉: 华中科技大学出版社, 2002
- [8] 陈亚宁. 灾害损失评估的灰色聚类分析[J]. 西北大学学报: 自然科学版, 1999, 29(6): 551–555.
- [9] Shannon C E. A mathematical theory of communication[J]. Bell System Technical Journal 1948 27(3): 379–423
- [10] 蒋金才, 季新菊, 刘良, 等. 河南省 1950~1990 年水旱灾害分析[J]. 灾害学, 1996 11(4): 69–73
- [11] 金菊良, 丁晶. 水资源系统工程[M]. 成都: 四川科学技术出版社, 2002.