

文章编号: 1001-0920(2006)05-0513-04

## 基于离差优化的信息不完全确定的多准则分类方法

王 坚 强

(中南大学 商学院, 长沙 410083)

**摘 要:** 提出一种准则权系数信息不完全确定且有训练集的多准则分类决策方法。该方法利用训练集的分类结果和准则权系数的不完全确定信息, 基于类间离差最大和类内离差最小构建非线性规划模型, 再利用遗传算法求解所得优化模型, 得到训练集中各方案在各准则下的效用值和准则权系数, 通过插值得到方案集中方案在各准则下的效用值, 确定方案与各类间的离差, 进而得到方案的分类。实例计算表明了该方法的有效性和可行性。

**关键词:** 多准则决策; 分类; 离差; 信息不完全确定; 遗传算法

中图分类号: C934

文献标识码: A

## Multi-criteria Classification Approach with Incomplete Certain Information Based on Optimizing Deviation of Categories

WANG Jian-qiang

(School of Business, Central South University, Changsha 410083, China. E-mail: jqwang@mail.csu.edu.cn)

**Abstract:** A multi-criteria classification approach with incomplete certain information and the reference set base on optimizing deviation of categories is proposed. Using classification to reference set and incomplete certain information on weights, nonlinear programming is constructed based on maximizing deviations between categories and minimizing deviations on categories. The criteria weights and utilities of alternative in reference set under criteria are obtained by using genetic algorithms to solve the nonlinear programming model. The utility of alternatives is gained by using interpolation. The deviations of alternatives and categories are computed, and the alternatives are classified. An example shows the feasibility and availability of the proposed approach.

**Key words:** Multi-criteria decision making; Classification; Deviation; Incomplete uncertain information; Genetic algorithm

### 1 引 言

在社会经济生活中, 存在大量多准则分类决策问题。目前求解这类问题的方法很多<sup>[1-5]</sup>, 如 ELECTRE TRI, MHDIS, UTADIS 等。在这些方法中, 绝大部分是假定准则权系数确定或准则权系数是由训练集完全推导出来的, 如 UTADIS 及其改进方法是假设方案的效用为各准则值的效用之和<sup>[3]</sup>, 通过训练集利用线性规划求得准则的权系数和训练集在各准则值的效用以及效用分类阈值, 进而对备选方案进行分类。其中决策者的偏好是通过训练集反映出来的, 且准则的权系数也是通过训练集推导

出来的。但在实际决策中, 决策者不仅可能对训练集进行决策, 而且可能给出各准则权系数的一些信息(不完全确定信息)。目前, 研究准则权系数信息不完全确定且有训练集的分类决策问题并不多见<sup>[6]</sup>。为此, 本文通过定义类间离差和类内离差, 提出一种基于类离差优化的信息不完全确定的多准则分类方法来求解这类分类决策问题。

### 2 基于类离差优化的信息不完全确定的多准则分类方法

#### 2.1 问题描述

收稿日期: 2005-03-31; 修回日期: 2005-08-15

基金项目: 国家自然科学基金项目(70572060); 教育部博士点基金项目(20040533057); 湖南省社会科学基金项目(05YB74)。

作者简介: 王坚强(1963—), 男, 湖南湘潭人, 教授, 博士, 从事决策理论与应用、物流管理等研究。

设某多准则分类决策问题有  $t$  个准则, 记为  $C_1, C_2, \dots, C_t, m$  个方案, 记为  $A = \{a_1, a_2, \dots, a_m\}$ , 决策者根据自己的偏好对一部分方案(训练集)  $B = \{b_1, b_2, \dots, b_n\}$  进行分类, 设分为  $q$  类, 记为  $D_1, D_2, \dots, D_q$ , 且各类之间有严格偏好关系, 即  $D_1 > D_2 > \dots > D_q$ . 方案  $a_i$  在准则  $C_j$  的值为  $a_{ij}$ , 方案  $b_i$  在准则  $C_j$  的值为  $b_{ij}$ . 根据上述假设确定方案  $a_1, a_2, \dots, a_m$  的分类

假设准则是效益型准则, 将非效益型准则转化为效益型准则

### 2.2 准则权系数的不完全确定信息

在实际决策中, 决策者很难准确地给出准则权系数的确定值, 或不能对准则间的重要性程度进行两两比较, 进而不能由 AHP, ANP, CHP 等方法确定准则权系数, 但通常能以不完全确定信息的形式给出准则权系数间的关系, 如某一准则的权系数在某一区间内变化; 一个准则比另一准则更重要; 几个准则的权系数确定, 而其他准则的权系数未知等. 在此, 假定准则权系数的不完全确定信息可以是线性不等式和线性等式的形式, 它可分为以下 3 类:

- 1)  $\{\omega A_1 \omega > b, \omega > 0, b > 0\}$ ;
- 2)  $\{\omega A_1 \omega < b, \omega > 0, b > 0\}$ ;
- 3)  $\{\omega A_1 \omega = b, \omega > 0, b > 0\}$ .

其中:  $A_1$  是一个  $l \times t$  的矩阵,  $\omega = (\omega_1, \omega_2, \dots, \omega_t)^T$ . 上述 3 类不完全确定信息是不完全信息、不确定信息、部分确定信息等的扩展. 令  $H$  表示准则权系数的不完全确定信息的集合.

### 2.3 决策过程

设  $X = \{a_1, a_2, \dots, a_m, b_1, b_2, \dots, b_n\}$ , 则任给  $x \in X$ , 方案  $x$  的效用为<sup>[7]</sup>

$$U(x) = \sum_{i=1}^t \omega_i U_i(x_i).$$

其中:  $\omega_i$  为准则  $C_i$  的相对权重;  $\sum_{i=1}^t \omega_i = 1$ ;  $U_i(x_i)$  是一非减的且值在 0 和 1 之间的实函数(称为效用函数), 且  $U_i(x_i^*) = 0, U_i(x_i^{\dagger}) = 1$ ;  $x_i^*$  和  $x_i^{\dagger}$  为准则  $C_i$  值的最小值和最大值.

#### 2.3.1 类间离差

对任意两个方案  $a_i$  和  $a_j$ , 方案  $a_i$  和  $a_j$  的离差定义为

$$d(a_i, a_j) = \sum_{k=1}^t [\omega_k |U_k(a_{ik}) - U(b_{jk})|^2].$$

类  $D_r$  和  $D_p$  间的离差定义为

$$d(D_r, D_p) = \min_{a_i \in D_r, a_j \in D_p} d(a_i, a_j).$$

$D_r$  类的离差定义为

$$d(D_r) = \min_{a_i \in D_r, a_j \in D_r} d(a_i, a_j).$$

方案  $a_i$  与类  $D_r$  的离差定义为

$$d(a_i, D_r) = \min_{a_j \in D_r} d(a_i, a_j).$$

#### 2.3.2 模型的建立与求解

由于假设所有准则均为效益型准则, 因而当  $x > y$  时, 有  $U_k(x) > U_k(y)$ .

为更好地分类, 需要两类之间离差尽可能大, 各类内离差尽可能小. 为此建立下列规划模型:

$$\begin{aligned} \max z &= \frac{1}{q-1} \sum_{n=1}^{q-1} d(D_n, D_{n+1}), \\ \text{s.t.} &\begin{cases} \sum_{k=1}^t \omega_k (U_k(b_{ik}) - U_k(b_{jk})) > 0, \\ b_i \in D_n, b_j \in D_{n+1}, \\ n = 1, 2, \dots, q-1; \\ U_k(x) > U_k(y), x > y; \\ \omega \in H; \\ \sum_{k=1}^t \omega_k = 1. \end{cases} \end{aligned} \quad (1)$$

$$\begin{aligned} \min z &= \frac{1}{q} \sum_{n=1}^q d(D_n), \\ \text{s.t.} &\text{同式(1)}. \end{aligned} \quad (2)$$

综合式(1)和(2)可得

$$\begin{aligned} \max z &= \frac{1}{q-1} \sum_{n=1}^{q-1} d(D_n, D_{n+1}) - \\ &\frac{1}{q} \sum_{n=1}^q d(D_n), \\ \text{s.t.} &\text{同式(1)}. \end{aligned} \quad (3)$$

上述模型是非线性的

令  $\mathcal{Q}(x) = \sum \omega U_k(x)$ , 此时有  $\mathcal{Q}(x^*) = \omega$ , 其中  $x^*$  为在准则  $C_k$  下最大值. 则式(3)变为

$$\begin{aligned} \max z &= \frac{1}{q-1} \sum_{n=1}^{q-1} d(D_n, D_{n+1}) - \\ &\frac{1}{q} \sum_{n=1}^q d(D_n), \\ \text{s.t.} &\begin{cases} \sum_{k=1}^t (\mathcal{Q}(b_{ik}) - \mathcal{Q}(b_{jk})) > 0, \\ b_i \in D_n, b_j \in D_{n+1}, \\ n = 1, 2, \dots, q-1; \\ \mathcal{Q}(x) > \mathcal{Q}(y), x > y; \\ (\mathcal{Q}(x_1^*), \mathcal{Q}(x_2^*), \dots, \mathcal{Q}(x_t^*)) \in H; \\ \sum_{k=1}^t \mathcal{Q}(x_k^*) = 1. \end{cases} \end{aligned} \quad (4)$$

其中

$$\begin{aligned} d(a_i, a_j) &= \sum_{k=1}^t |\mathcal{Q}(a_{ik}) - \mathcal{Q}(a_{jk})|^2, \\ d(D_r, D_p) &= \min_{a_i \in D_r, a_j \in D_p} d(a_i, a_j), \end{aligned}$$

$$d(D_r) = \max_{a_i, D_r, a_j, D_r} d(a_i, a_j).$$

2.3.3 模型的求解

由于非线性规划模型(4)的目标函数是由一系列二次函数取极大值和极小值构成的,因此采用传统优化方法求解较为困难,这里采用遗传算法求解对带有约束条件的非线性规划问题的染色体进行遗传运算时,通常会产生不可行的后代,因此主要问题是如何满足约束条件.本文利用惩罚策略,通过构造惩罚函数惩罚不可行解,将约束问题转化为无约束问题.在遗传算法中,每代种群都会保持部分不可行解,因而可从可行域和不可行域两边搜索找到最优解<sup>[8]</sup>.其算法流程如图1所示

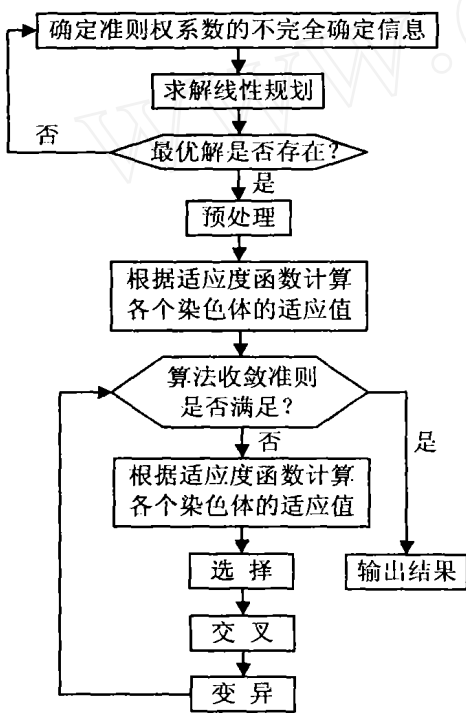


图1 遗传算法流程

算法的关键环节设计如下:

1) 初始值的选择

通过求解下列线性规划:

$$\begin{aligned} \max & 0, \\ \text{s.t.} & \text{同式(4)}, \end{aligned} \quad (5)$$

确定其最优解为遗传算法的初始解.如果式(5)的最优解不存在,则说明决策者给出的准则权系数的不完全确定信息和/或决策者对训练集的分类信息中存在矛盾,需要重新调整,调整后再继续

2) 预处理

采用实数编码,将  $\varphi(x_1^*), \varphi(x_2^*), \dots, \varphi_1(x_{i-1}^*), \{\varphi(x_1) | x_1 \in \{b_n\} \setminus \{x_1^*\}\}, \{\varphi(x_2) | x_2 \in \{b_2\} \setminus \{x_2^*\}\}, \dots, \{\varphi(x_i) | x_i \in \{b_n\} \setminus \{x_i^*\}\}$  构成一个染色体  $\varphi(x_i^*) = 1 - \varphi(x_1^*) - \varphi(x_2^*) - \dots -$

$\varphi_1(x_{i-1}^*)$  由计算得出,这样才能保证经过交叉、变异等运算,使得  $\sum_{k=1}^i \varphi(x_k^*) = 1$ .

3) 根据适应度函数计算种群的适应值

为了防止早熟现象,不应直接根据目标函数值来确定染色体的适应度,而是在目标函数的基础上进行某种变换后得到适应度函数  $F(x) = g(f(x))$ ,其中  $f$  是目标函数,  $g$  将目标函数值转换成一个非负数,  $F$  就是相应的适应度

在图1中调用目标函数  $f(x)$ ,即转入图2的处理.这里适应度函数采用  $F(\text{pos}) = 2 - \text{SP} + 2(\text{SP} - 1)(\text{pos} - 1) / (\text{N ind} - 1)$ ,其中:  $\text{pos}$  是对染色体  $x$  的目标函数值  $f(x)$  经过降序排列后的位置,  $\text{SP}$  为选择压力,可取  $[1, 1, 2]$  之间的数,  $\text{N ind}$  是种群的大小<sup>[8]</sup>.

4) 算法的收敛条件

根据是否找到搜索结果和是否超出指定的迭代次数而定.如果找到,即可得到权系数等参数的值

5) 遗传算子

根据适应度来复制下一代染色体,这里选择算子采用  $\text{SU S}$  (Stochastic Universal Sampling); 交叉算子采用算术交叉,由两个染色体的线性组合产生两个新的个体;变异算子采用非均匀变异,使之能够较好地搜索原个体附近的微小区域

6) 目标函数的计算流程

对于目标函数,其计算流程如图2所示.首先从传进来的种群中取出一条染色体,计算出权系数  $\varphi(x_i^*)$ ,求出是否满足约束条件的惩罚量.当惩罚量为零时,求出  $d(D_r, D_p)$  和  $d(D_r)$ ,得到目标函数值,保存这条染色体并返回适应度函数.由适应度函数判断算法收敛条件是否满足.当惩罚量不为零时,

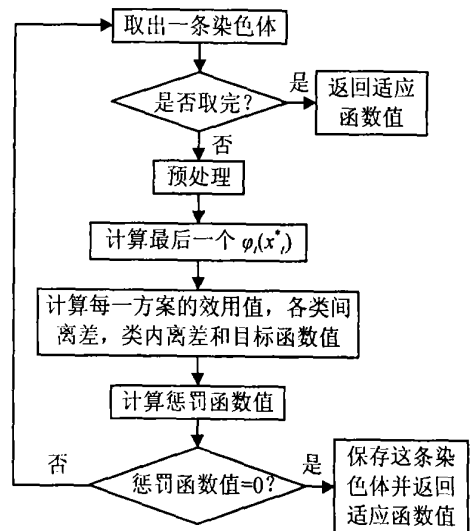


图2 目标函数值计算流程

说明为不可行解,令求出的适应度加上这个惩罚量,继续进行运算,从而达到从可行域和不可行域两边搜索最优解的目的

按前面介绍的遗传算法求解模型(4),便可得到最优解和最优值

### 2.3.4 方案的分类

通过线性插值或样条插值得到每一方案在各准则下的效用值,进而得到每一方案与已分类方案的离差.如果方案与某类的离差最小,则该方案属于该类.即计算  $d_{ir} = \min_{b_j} (a_i, b_j)$ ,  $i = 1, 2, \dots, m; r = 1, 2, \dots, q$ . 若  $d_{ik} = \min_{r=1,2,\dots,q} \{d_{ir}\}$ , 则  $a_i \in D_k$ . 这样就得到了方案集A 的分类

## 3 实例计算

考虑有5个准则,10个方案的分类决策问题<sup>[6]</sup>. 5个准则均为效益型准则,已知前8个方案分为3类  $D_1 > D_2 > D_3$ , 其中:  $a_4, a_6 \in D_1; a_1, a_2, a_8 \in D_2; a_3, a_5, a_7 \in D_3$ . 准则权系数满足下列条件:  $0.1 \leq \omega_1 \leq 0.35, 0.15 \leq \omega_2 \leq 0.45, 0.05 \leq \omega_3 \leq 0.2, 0.1 \leq \omega_4 \leq 0.3, 0.08 \leq \omega_5 \leq 0.15; \omega_1 > 2\omega_2, \omega_1 > 3\omega_3, \omega_2 > 2\omega_4, \omega_2 > \omega_5 > \omega_3 > \omega_4 > \omega_5 > 0$ . 各方案在准则下的值如表1所示

表1 各方案的准则值

	准则1	准则2	准则3	准则4	准则5
$a_1$	30	6	5	3.5	18
$a_2$	70	3	8	4.5	24
$a_3$	50	1	4	5.5	15
$a_4$	100	4	6	8.0	20
$a_5$	60	2	8	7.5	19
$a_6$	80	5	7	6.0	21
$a_7$	30	2	4	4.5	17
$a_8$	70	4	6	6.5	20
$a_9$	60	5	7	7	22
$a_{10}$	45	3	4	5	19

确定其他方案的分类.利用式(4)建立优化模型,在Matlab 6.5中编程求解得到如下结果:

准则的权系数为

$$\omega_1 = 0.307, \omega_2 = 0.308, \omega_3 = 0.152, \omega_4 = 0.153, \omega_5 = 0.080$$

方案  $a_9$  与各分类的离差分别为

$$0.2282, 0.2690, 0.4668,$$

所以  $a_9 \in D_1$ .

方案  $a_{10}$  与各分类的离差分别为

$$0.6357, 0.3139, 0.2105,$$

所以  $a_{10} \in D_3$ .

上面分类结果与文献[6]中的分类结果一致.对表1中数据进行分析可知,上述分类结果是合理的.

## 4 结论

针对准则权系数信息不完全确定且有训练集的多准则分类决策问题,建立了基于类离差优化的非线性规划模型,利用遗传算法求解该优化模型,并详细讨论了实现过程.该方法充分反映了决策者的偏好,同时又能满足决策者给出的信息不完全确定的准则权系数的要求.实际应用表明,本文方法操作性较强,可用于各类相关决策.

## 参考文献(References)

[1] Jacquet-Lagrange E, Siskos Y. Preference Disaggregation: 20 Years of MCDA Experience [J]. *European J of Operational Research*, 2001, 130(2): 233-245

[2] Zopounidis C, Doumpos M. Multi-criteria Classification and Sorting Methods: A Literature Review [J]. *European J of Operational Research*, 2002, 138(2): 229-246

[3] Beuthe M, Scannella G. Comparative Analysis of UTA Multicriteria Methods [J]. *European J of Operational Research*, 2001, 130(2): 246-262

[4] Zopounidis C, Doumpos M. A Preference Disaggregation Decision Support System for Financial Classification Problems [J]. *European J of Operational Research*, 2001, 130(2): 402-413

[5] Zopounidis C, Doumpos M. Multi-group Discrimination Using Multi-criteria Analysis: Illustrations From the Field of Finance [J]. *European J of Operational Research*, 2002, 139(2): 371-389

[6] 王坚强. 信息不完全的多准则层次分类方法[J]. *控制与决策*, 2004, 19(11): 1237-1240  
(Wang J Q. Study on Hierarchical Discrimination Approach of Multi-criteria Classification with Incomplete Information [J]. *Control and Decision*, 2004, 19(11): 1237-1240)

[7] Keeney R L, Raiffa H. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs* [M]. New York: Wiley, 1976

[8] 玄光男, 程润伟. *遗传算法与工程设计* [M]. 北京: 清华大学出版社, 2004  
(Gen M S, Cheng R W. *Genetic Algorithms and Engineering Optimization* [M]. Beijing: Tsinghua University Press, 2004)