文章编号: 1001-0920 (2012) 09-1411-04

基于判断矩阵的专家模糊核聚类组合赋权方法

李 闯, 端木京顺, 蔡忠义, 高建国

(空军工程大学 工程学院，西安 710038)

摘 要: 研究群决策中专家赋权问题. 实际决策问题中, 由于客体信息自身存在的不完备性和不确定性以及人们描述过程中的模糊性, 更适合采用模糊聚类的分析方法, 为此提出一种基于判断矩阵的专家模糊核聚类赋权方法. 该方法运用模糊核聚类理论对专家排序向量进行分类, 根据分类结果、判断矩阵一致性和排序向量的熵对各专家进行组合赋权. 算例表明, 所提出的方法是可行且有效的.

关键词: 专家赋权；判断矩阵；模糊核聚类分析

中图分类号: C943 **文献标志码:** A

Method for combination weighting experts based on judgment matrix and fuzzy kernel clustering analysis

LI Chuang, DUANMU Jing-shun, CAI Zhong-yi, GAO Jian-guo

(College of Engineering，Air Force Engineering University，Xi’an 710038，China．Correspondent：LI Chuang, E-mail：lichgcxy@163.com)

Abstract：The problem of weighting experts in group decision-making is studied. In the decision-making problems, because the information of object itself is incomplete and uncertainty, and people describe fuzzily in the process of decision-making, the fuzzy clustering analysis is more suitable for such problems. Therefore, a method for deriving experts’ combination weights based on judgment matrix and fuzzy kernel cluster analysis is provided, in which the collating vectors of an individual expert are classified by using fuzzy kernel clustering principle. The experts’ combination weights are determined according to the result of classification, the judgment matrix’s consistency and the entropy of collating vectors. Finally, a numerical example shows the feasibility and effectiveness of the proposed method.

Key words：experts’ weights；judgment matrix；fuzzy kernel clustering

1 引 言

在多属性、多目标群决策问题中, 由于采用群体专家决策的方法, 减小了单个评价专家主观偏见和模糊认识对最终评估结果的影响, 提高了决策的正确性. 其中专家赋权问题的研究在决策过程中占有重要的地位, 是关键环节之一, 其合理性直接影响着决策结果的准确性.

目前, 人们针对专家赋权方法进行了许多研究. 但如何合理地确定专家权值是迄今仍然没有得到很好解决的问题. 在群决策过程中, 基于判断矩阵的专家赋权方法大致可分为两种: 一是根据专家判断矩阵一致性的方法; 二是利用系统聚类分析的思想对专家先分类再赋权的方法[1]. 传统的系统聚类方法有基于距离和基于相似系数的方法等, 是一种硬性聚类. 这种聚类的特点是界限分明, 具有“非此即彼”的性质. 但在实际决策问题中, 由于客体信息自身存在的不完备性和不确定性以及人们描述过程中的模糊性, 更适合采用模糊聚类[2]的分析方法, 通过建立模糊相似关系, 对客观事物进行聚类, 得到样本隶属于各个类的不确定性程度, 建立隶属于类的不确定性描述, 更能客观地反映现实世界[3]. 模糊核聚类是对传统聚类方法的改进, 将排序向量映射到高维特征空间, 增加了个体间的可分概率, 较好地实现了对差别微弱的向量间的聚类[4-5], 同时能够对非超球体数据、被噪声污染的数据、多种模式原型混合的数据以及不对称数据等多种数据结构进行聚类分析, 其聚类性能明显高于传统模糊聚类算法[6]. 同时, 现有的研究成果虽然分别从判断矩阵的一致性[1,7]和排序向量的熵[8-10]的角度

收稿日期: 2011-04-07；**修回日期:** 2011-08-29.

基金项目: 国家自然科学基金项目(60773209).

作者简介: 李闯(1982–), 男, 博士生, 从事装备安全管理与决策等研究；端木京顺(1955–), 男, 教授, 博士生导师, 从事装备安全管理与决策、装备安全控制等研究.

对类内专家进行赋权, 但尚未考虑二者对专家权值的综合影响.

针对以上问题, 本文提出一种基于判断矩阵的专家模糊核聚类组合赋权方法. 通过模糊核 C 均值算法对专家进行聚类分析, 根据聚类结果中的专家人数、判断矩阵一致性比例和排序向量的熵为各专家组合赋权. 使得在整体评价中专家数目多、共识较好的类别具有更高的权重, 在同一类内的专家, 逻辑清晰、思维严密的专家赋予更高的类内权重, 从而获得更加合理的专家权值.

2 基于判断矩阵的专家模糊核聚类组合赋权方法

基于判断矩阵的专家模糊核聚类组合赋权方法分两步: 首先, 对各专家排序向量进行模糊核 C 均值聚类; 然后, 根据聚类结果中的专家人数、判断矩阵一致性比例 CR 和排序向量的熵为各专家组合赋权.

2.1 专家模糊核聚类分析

模糊核聚类的思想是, 通过非线性映射函数 $\Phi(\cdot)$, 将各排序向量非线性映射到一个高维特征空间 R^q, 在特征空间中对专家进行模糊聚类. 通过引入满足 Mercer 条件核函数, 在特征空间的点积运算可以转化为对应排序向量核函数的计算, 而与映射函数 $\Phi(\cdot)$ 的具体形式以及特征空间的维数无关. 采用排序向量隶属度和类间模糊相关度[11]作为聚类的评判标准.

设有 p 个专家在某一准则下对 n 个对象进行评价, 第 i 个专家给出的判断矩阵为 A_i, 由判断矩阵 A_i 求出的评价个体排序向量为 $U_i=(u_{i1},u_{i2},\cdots,u_{in})$, $i=1,2,\cdots,p$, 以 $\Phi(U_i)$ 表示排序向量 $U_i\,(i=1,2,\cdots,p)$ 在高维特征空间的像. 若选用核函数 K, 则在特征空间 R^q, U_i 和 U_j 之间的 Euclid 距离定义为

$$d_{Kij}=\|\Phi(U_i)-\Phi(U_j)\|_E=$$
$$[K(U_i,U_i)-2K(U_i,U_j)+K(U_j,U_j)]^{1/2}. \tag{1}$$

鉴于在基于目标函数的聚类算法中, 模糊 C 均值算法 (FCM) 最为完善, 为此在特征空间扩展 Bezdek 的 FCM 算法中, 将对 $\Phi(U_i)\,(i=1,2,\cdots,p)$ 进行模糊聚类的问题归为带约束的目标函数非线性规划问题, 其目标函数为

$$J_{Km}(U;R,V)=$$
$$\sum_{j=1}^{C}\sum_{i=1}^{p}r_{ji}^{m}\|\Phi(U_i)-\Phi(V_j)\|_E^2=$$
$$\sum_{j=1}^{C}\sum_{i=1}^{p}r_{ji}^{m}(K(U_i,U_i)-2K(U_i,V_j)+K(V_j,V_j))=$$
$$\sum_{j=1}^{C}\sum_{i=1}^{p}r_{ji}^{m}d_{Kij}^{2}(U_i,V_j),\ 2\leqslant C\leqslant p;$$

$$\sum_{j=1}^{C}r_{ji}=1,\forall i=1,2,\cdots,p,\ 0\leqslant r_{ji}\leqslant 1. \tag{2}$$

其中: $V=\{V_1,V_2,\cdots,V_C\}^{\mathrm{T}}, V_j\in R^n$ 为聚类中心, $R=(r_{ji})_{C\times p}$ 为隶属度矩阵, 元素 $r_{ji}\in[0,1]$ 表示排序向量 U_i 隶属于第 j 类的程度, C 为类数, $m\in[0,\infty)$ 为模糊加权指数.

隶属度公式为

$$r_{ji}=\frac{(1/d_{Kji}^2(U_i,V_j))^{1/(m-1)}}{\displaystyle\sum_{j=1}^{C}(1/d_{Kji}^2(U_i,V_j))^{1/(m-1)}}, \tag{3}$$

聚类中心的迭代公式为

$$\Phi(\hat{V}_j)=\frac{\displaystyle\sum_{i=1}^{p}r_{ji}^{m}\Phi(U_i)}{\displaystyle\sum_{i=1}^{p}r_{ji}^{m}},\ j=1,2,\cdots,C. \tag{4}$$

则可计算得到

$$K(U_i,\hat{V}_j)=\Phi(U_i)\Phi(\hat{V}_j)=\frac{\displaystyle\sum_{k=1}^{p}r_{jk}^{m}K(U_k,U_i)}{\displaystyle\sum_{k=1}^{p}r_{jk}^{m}}, \tag{5}$$

$$K(\hat{V}_j,\hat{V}_j)=\Phi(\hat{V}_j)\times\Phi(\hat{V}_j)=$$
$$\frac{\displaystyle\sum_{k=1}^{p}\sum_{l=1}^{p}r_{jk}^{m}r_{jl}^{m}K(U_k,U_l)}{\Big[\displaystyle\sum_{k=1}^{p}r_{jk}^{m}\Big]^2}. \tag{6}$$

隶属度迭代公式为

$$\hat{r}_{ji}=\frac{(1/d_{Kji}^2(U_i,\hat{V}_j))^{1/(m-1)}}{\displaystyle\sum_{j=1}^{C}(1/d_{Kji}^2(U_i,\hat{V}_j))^{1/(m-1)}}=$$
$$\frac{\Big(\dfrac{1}{K(U_i,U_i)-2K(U_i,\hat{V}_j)+K(\hat{V}_j,\hat{V}_j)}\Big)^{1/(m-1)}}{\displaystyle\sum_{j=1}^{C}\Big(\dfrac{1}{K(U_i,U_i)-2K(U_i,\hat{V}_j)+K(\hat{V}_j,\hat{V}_j)}\Big)^{1/(m-1)}}. \tag{7}$$

目标函数求解的基本思想是, 通过反复修改聚类中心 V 和隶属度矩阵 R 实现动态的迭代聚类, 求解获得一个最优模糊隶属度矩阵 R^*, 使得被划分到同一个类的对象之间相似度最大, 而不同类之间的相似度最小, 实现排序向量的模糊划分和聚类.

聚类分析步骤如下:

Step 1: 求出专家给出的排序向量.

Step 2: 选择类数 C, 迭代停止条件 $\varepsilon\in(0,1)$ 或迭代次数 T; 选择核函数 K 及其参数; 初始化类中心 V_j, $j=1,2,\cdots,C$.

Step 3: 按式 (3) 计算排序向量在特征空间的隶属

度 $r_{ji}, j=1,2,\cdots,C,\ i=1,2,\cdots,p$.

Step 4: 由式 (5) 和 (6) 计算新的核函数值 $K(U_i, \hat{V}_j)$ 和 $K(\hat{V}_j,\hat{V}_j)$, 并按式 (7) 更新隶属度 r_{ji} 为 $\hat{r}_{ji}$.

Step 5: 若 $\max\limits_{j,i}|u_{ji}-\hat{u}_{ji}|<\varepsilon$ 或迭代次数等于预定迭代次数 T, 则算法停止, 输出隶属度矩阵 R^*, 确定最终的专家聚类结果; 否则, 转 Step 4.

2.2 专家组合赋权分析

为专家赋权, 不仅要考虑专家所在类中的人数, 还应同时考虑专家判断矩阵的一致性以及排序向量所蕴含的信息量.

对于人数多的类, 其专家给出的评价信息符合较多专家的意见, 对应专家应赋予较大的权值; 反之, 对于人数较少的类中的专家, 应赋予较小的权值.

类内专家权值的确定. 传统的赋权方法有平均赋权法和基于判断矩阵一致性的赋权方法, 但都存在不足: 前者没有考虑排序向量之间的差异性; 后者虽然考虑了判断矩阵的一致性, 但没有考虑排序向量所表达的信息量和不确定程度[10]. 事实上, 判断矩阵的一致性越好, 说明专家思路越清晰, 矩阵数据内隐含的矛盾越小, 对应的专家应该赋予更大的权值. 同时, 仅考虑判断矩阵的一致性是不全面的, 还应综合考虑排序向量的信息量和不确定程度. 熵作为信息不确定性以及信息量的度量, 反映了信息的混乱程度, 排序向量的熵代表了专家由于个体原因在描述客观事物的过程中损失的信息量, 即由此产生对客观事物认识的混乱性[8-10,12-14]. 如果专家给出的排序向量的熵越小, 说明该专家对客观事物认识的混乱程度越小, 提供的信息量越大, 不确定因素越小, 则计算专家权重时应赋予更大的权值. 基于上述不足, 下面提出基于判断矩阵的一致性以及排序向量的熵对类内专家进行组合赋权.

假设 p 位专家被分为 L 类, 第 i 类中专家的数量为 Ψ_i. λ_i 为类间权值, α_{ij} 为第 i 类中第 j 位专家的类内组合权值, β_{ij} 为第 i 类中第 j 位专家的类内一致性权值, γ_{ij} 为第 i 类中第 j 位专家的类内熵权值; ω_{ij} 为第 i 类中第 j 位专家的总体权值.

$U_{ij}=(u_{ij1},u_{ij2},\cdots,u_{ijn})$ 为第 i 类中第 j 位专家给出的排序向量, U_{ij} 的熵定义[12]为

$$H(U_{ij})=-\sum_{k=1}^{n}u_{ijk}\log_2(u_{ijk}),\quad i=1,2,\cdots,l,\ j=1,2,\cdots,\Psi_i. \tag{8}$$

专家的类内熵权值为

$$\gamma_{ij}=(1-H(U_{ij}))\Big/\sum_{j=1}^{\Psi_i}[1-H(U_{ij})], \tag{9}$$

专家的类内一致性权值为[7]

$$\beta_{ij}=\left(\frac{1}{1+bCR_{ij}}\right)\Big/\sum_{j=i}^{\Psi_i}\frac{1}{1+bCR_{ij}}. \tag{10}$$

通过求得的 γ_{ij} 和 β_{ij}, 对专家的类内权值进行组合赋权, 即

$$\alpha_{ij}=(\beta_{ij}+\gamma_{ij})\Big/\sum_{j=1}^{\Psi_i}(\beta_{ij}+\gamma_{ij}). \tag{11}$$

专家类间权值为

$$\lambda_i=\Psi_i^2\Big/\sum_{j=1}^{l}\Psi_j^2. \tag{12}$$

通过求得的 α_{ij} 和 λ_i, 可得到专家的总体权值

$$\omega_{ij}=\lambda_i\times\alpha_{ij}. \tag{13}$$

3 算例分析

本文引用文献 [1] 的算例, 有 7 位专家在某一准则下对 3 个元素进行评价.

1) 根据专家给出的排序向量进行专家模糊核聚类分析. 因为高斯核函数对应的特征空间大多是无穷维的, 有限样本在该特征空间中肯定是线性可分的, 所以本文采用高斯核函数

$$K(x,y)=\exp(-\|x-y\|^2/b^2),$$

取参数 $b=4.3$. 迭代停止条件参数的选取, 主要取决于对聚类精度和聚类速度的要求. 本文主要考虑聚类精度问题, 因此取误差精度 ε 作为迭代停止条件, 其数值选取过大会导致算法过早收敛, 聚类结果不稳定, 数值选取过小可能会导致过度计算, 既影响收敛速度又可能发生无法收敛的问题. 文献 [14] 指出, $10^{-4}\leqslant\varepsilon\leqslant10^{-5}$ 时迭代次数较少, 同时模糊聚类精度较高, 因此本文取 $\varepsilon=10^{-5}$. 有 7 位专家参与决策, 聚类数的可能取值为 $C=i, i=2,3,4,5,6$.

C 取 5 或 6 时, 由于存在零专家类的情况, 予以舍弃, 得到的聚类结果以及相应的模糊相关度 RV 和模糊隶属度 R 分别如下:

$C=2:\{(3,5,7),(1,2,4,6)\}$ 时

$$\mathrm{RV}_2=\begin{bmatrix}1.00 & 0.83\\ 0.83 & 1.00\end{bmatrix},$$

$$R_2=\begin{bmatrix}0.16 & 0.17 & 0.79 & 0.23 & 0.73 & 0.23 & 0.87\\ 0.84 & 0.83 & 0.21 & 0.77 & 0.27 & 0.77 & 0.13\end{bmatrix};$$

$C=3:\{(2,7),(3,5),(1,4,6)\}$ 时

$$\mathrm{RV}_3=\begin{bmatrix}1.00 & 0.91 & 0.67\\ 0.91 & 1.00 & 0.55\\ 0.67 & 0.55 & 1.00\end{bmatrix},$$

$$R_3=\begin{bmatrix}0.01 & 0.46 & 0.18 & 0.01 & 0.28 & 0.01 & 0.68\\ 0.00 & 0.23 & 0.81 & 0.01 & 0.69 & 0.01 & 0.31\\ 0.99 & 0.31 & 0.01 & 0.98 & 0.03 & 0.98 & 0.02\end{bmatrix};$$

$C=4:\{(2),(3,7),(5),(1,4,6)\}$ 时

$$\mathrm{RV}_4=\begin{bmatrix}1.00&0.75&0.78&0.63\\0.75&1.00&0.99&0.34\\0.78&0.99&1.00&0.37\\0.63&0.34&0.37&1.00\end{bmatrix},$$

$$R_4=\begin{bmatrix}0.01&0.97&0.03&0.06&0.07&0.06&0.05\\0.00&0.01&0.63&0.01&0.44&0.01&0.58\\0.00&0.01&0.33&0.02&0.45&0.01&0.36\\0.99&0.01&0.01&0.91&0.03&0.92&0.02\end{bmatrix},$$

或 $C=4:\{(1,4,6),(7),(2),(3,5)\}$ 时

$$\mathrm{RV}'_4=\begin{bmatrix}1.00&0.38&0.53&0.33\\0.38&1.00&0.75&0.93\\0.53&0.75&1.00&0.69\\0.33&0.93&0.69&1.00\end{bmatrix},$$

$$R'_4=\begin{bmatrix}0.95&0.02&0.00&0.94&0.03&0.94&0.02\\0.01&0.04&0.05&0.01&0.23&0.01&0.52\\0.04&0.92&0.01&0.04&0.07&0.04&0.06\\0.00&0.01&0.94&0.01&0.67&0.01&0.40\end{bmatrix}.$$

当 $C=2$ 时, 不同类间模糊相关度为 0.83, 表明将专家聚为两类时其区分度很小, 因此予以舍弃; 当 $C=3$ 时, 虽与文献 [1] 聚类结果一致, 但聚类结果的模糊相关度均大于 0.5, 从赋权准确性方面考虑, 应尽可能提高专家区分度及聚类精度, 因此予以舍弃; 当 $C=4$ 时, 得到两个聚类结果, 其中之一与文献 [10] 聚类结果一致, 但从结果的对比分析可知, 其主要差别在于将专家 3 与专家 7 聚为一类或将专家 3 与专家 5 聚为一类. 由于 RV_4 不同类间的模糊相关度值普遍大于 RV'_4, 且将专家 7 单独聚为一类时, 其隶属于本类的隶属度值为 0.52, 隶属于类 $(3,5)$ 的隶属度值为 0.40; 将专家 5 单独聚为一类时, 其隶属于本类的隶属度值为 0.45, 隶属于类 $(3,7)$ 的隶属度值为 0.44. 因此将专家 3 与专家 5 聚为一类, 专家 7 单独聚为一类更加合理, 从而确定最优聚类结果为 $\{(1,4,6),(7),(2),(3,5)\}$.

2) 根据聚类结果中的专家人数、判断矩阵一致性比例 CR 和排序向量的熵为专家进行组合赋权分析. 由式 (12) 得到专家的类间权值为

$$\lambda_1=9/15,\ \lambda_2=\lambda_3=1/15,\ \lambda_4=4/15;$$

由式 (10) 得到专家的类内一致性权值为

$$\{\beta_{11}=0.41,\beta_{12}=0.29,\beta_{13}=0.30\},\ \{\beta_{21}=1\},$$
$$\{\beta_{31}=1\},\ \{\beta_{41}=0.55,\beta_{42}=0.45\};$$

由式 (8) 得到专家排序向量的信息熵为

$$H(U_1)=0.70,\ H(U_2)=0.76,\ H(U_3)=0.84,$$
$$H(U_4)=0.67,\ H(U_5)=0.81,$$
$$H(U_6)=0.70,\ H(U_7)=0.82,$$

进而, 由式 (9) 得到专家的类内熵权值为

$$\{\gamma_{11}=0.40,\gamma_{12}=0.22,\gamma_{13}=0.38\},$$
$$\{\gamma_{21}=1\},\ \{\gamma_{31}=1\},\ \{\gamma_{41}=0.54,\gamma_{42}=0.46\};$$

由式 (11) 得到专家类内组合权值为

$$\{\alpha_{11}=0.41,\alpha_{12}=0.26,\alpha_{13}=0.33\},$$
$$\{\alpha_{21}=1\},\ \{\alpha_{31}=1\},\ \{\alpha_{41}=0.55,\alpha_{42}=0.45\}.$$

3) 由式 (13) 对专家权值进行集结, 可以得到专家总体权值向量为

$$W=(0.24,0.07,0.14,0.16,0.12,0.20,0.07)^{\mathrm{T}}.$$

综合 7 位专家的权值和个人排序向量, 可得最终的综合排序向量为 $U=(0.67,0.24,0.09)$.

从算例分析过程可以发现: 首先, 对专家进行模糊核聚类, 增加了排序向量的可分度, 得到了排序向量隶属于各类的不确定性程度, 聚类过程综合考虑了模糊隶属度和模糊相关度, 比传统用单一指标衡量聚类结果更加合理, 准确度更高; 其次, 类内专家的一致性权值与熵权值的不同, 进一步表明了综合考虑判断矩阵的一致性以及排序向量的熵对专家进行赋权的合理性, 在专家人数多的类别中, 一致性比例越高、熵值越小的专家被赋予的权值越大, 从而对整体排序向量的影响程度越大, 更能客观地反映决策信息, 符合决策的实际需求.

4 结 论

根据对现有专家赋权方法的研究总结, 提出了一种采用模糊核聚类理论对专家进行分类, 依据聚类结果、判断矩阵一致性和排序向量的熵对各专家进行组合赋权的方法. 算例分析表明, 所提出的方法是可行且有效的.

参考文献(References)

[1] 高阳, 罗贤新, 胡颖. 基于判断矩阵的专家聚类赋权研究[J]. 系统工程与电子技术, 2009, 31(3): 593-596.
(Gao Y, Luo X X, Hu Y. Research on methods for deriving experts' weights based on judgment matrix and cluster analysis[J]. Systems Engineering and Electronics, 2009, 31(3): 593-596.)

[2] Bezdek J C, Ehrlich R, Full W. FCM: The fuzzy c-means clustering algorithm[J]. Computers and Geosciences, 1984, 10(23): 191-203.

[3] 伍忠东, 高新波, 谢维信. 基于核方法的分类型属性数据集模糊聚类算法[J]. 华南理工大学学报: 自然科学版, 2004, 32(9): 23-28.
(Wu Z D, Gao X B, Xie W X. A new fuzzy clustering algorithm of categorical data set based on the kernel method[J]. J of South China University of Technology: Natural Science Edition, 2004, 32(9): 23-28.)

(下转第1420页)