第36卷第4期 海南师范大学学报(自然科学版) Vol.36 No.4
2023年12月 Journal of Hainan Normal University(Natural Science) Dec.2023

Doi:10.12051/j.issn.1674-4942.2023.04.002

基于改进相关系数的犹豫模糊集聚类分析

范玉蕊，王惠文*
(云南师范大学 数学学院,云南 昆明 650500)

摘 要:针对犹豫模糊集相关系数计算公式中可能存在分母为0的特殊情况,对相关系数公式重新定义,并对犹豫模糊集之间的隶属度差异和犹豫度差异2个部分进行综合考虑,进一步将其推广至区间值犹豫模糊集上,同时弥补了需要对数据进行主观填充这一不足之处,最后证明其良好性质,并应用于数值算例中说明合理有效性。

关键词:犹豫模糊集;区间值犹豫模糊集;相关系数;犹豫度;聚类分析

中图分类号:O159 **文献标志码**:A **文章编号**:1674-4942(2023)04-0369-07

Cluster Analysis of Hesitant Fuzzy Sets Based on Improved Correlation Coefficient

FAN Yurui, WANG Huiwen*
(*School of Mathematics*, *Yunnan Normal University*, *Kunming* 650500, *China*)

Abstract: Aiming at the special situation that there may be a denominator of 0 in the calculation formula of the correlation coefficient of the hesitant fuzzy set, the correlation coefficient formula is redefined, and the membership difference and the hesitation difference between the hesitant fuzzy sets are comprehensively considered, and it is extended to interval-valued hesitant fuzzy sets, which makes up for the need to subjectively fill the data, and finally its good properties are proved, and it is applied to to numerical studies to illustrate its reasonable validity.

Keywords: hesitant fuzzy set; interval-valued hesitant fuzzy set; correlation coefficient; hesitant degree; cluster analysis

在处理模糊不确定性的问题时,模糊集理论显示出其优越性和有效性。在日常的经济生产活动中,遇到决策者在几个数值间徘徊的情况时,直觉模糊集不能够进行精确描述。为了解决这个问题,Torra[1]将犹豫这一概念融入到模糊集理论中,并给出犹豫模糊集的概念,用于反映出决策者的犹豫情况。之后,Chen等[2]又提出了区间值犹豫模糊集,用区间数来表示隶属度情况。在此基础上,文献[3-4]研究了其相关性质和基本运算。

相关系数一直是模糊理论中的重要研究对象,在聚类分析中有着广泛的应用。2013年,Chen等[2]定义了犹豫模糊集的相关性度量,并讨论了相关性质,但是默认犹豫模糊元具有相同的长度,其值按递增顺序排列;次年将相应公式推广到区间值犹豫模糊集上。2015年,Liao等[5]提出犹豫模糊集的相关系数也应具有一定的犹豫性,在一定的区间内并非是一个准确的数。2018年,Guan等[6]提出综合相关系数,从均值、方差、长度3个角度综合考量,克服了之前要求对应犹豫模糊元具有相同长度的局限性,但是覆盖了一定的内部信息。2019年,Yang等[7]提出的相关系数公式同样不要求对应犹豫模糊元长度相等,但是存在的特殊情况(分母为0)没有被考虑在内。另外,文献[8-12]细节性地阐述了相关系数在聚类分析中的应用,文献[13-14]展示了2种其他类型的相关系数。本文继承了Yang等人的思想,并对其进行补充,提出带有犹豫度的犹豫模糊集相关系数计算公式,并推

收稿日期:2022-11-12
基金项目:云南师范大学研究生科研训练基金项目(YJSJJ22-B93)
第一作者:范玉蕊(1998—),河南新乡人,硕士研究生,研究方向为模糊多属性决策。E-mail:2922070067@qq.com
*通信作者:王惠文(1975—),江苏镇江人,副教授,研究方向为模糊决策。E-mail:mathynnu@163.com

广到区间值犹豫模糊集上,最后证明其良好性质,并应用于数值算例说明其合理性和有效性。

1 基本概念

定义1[8] 设 $X=\{x_1,x_2,\cdots,x_n\}$ 是一非空集合,称 $A=\left\{\left\langle x,h_A(x)\right\rangle | x\in X\right\}$ 为犹豫模糊集(HFS),其中 $h_A(x)$ 表示 x 对于 A 的所有可能隶属度的集合,隶属度的取值范围为[0,1],$h=h_A(x)$ 为犹豫模糊元。

定义2[3] 设 $X=\{x_1,x_2,\cdots,x_n\}$ 是一非空集合,称 $A=\left\{\left\langle x,\tilde{h}_A(x)\right\rangle | x\in X\right\}$ 为区间值犹豫模糊集(HIVFS),$\tilde{h}_A(x)$ 表示 x 对于 A 的所有可能隶属度区间,其端点值取值范围为[0,1],$h=\tilde{h}_A(x)$ 为区间值犹豫模糊元。

定义3[2] 设 $\boldsymbol{C}=(\rho_{ij})_{m\times m}$ 是一个相关系数矩阵,如果 $\boldsymbol{C}^2=\boldsymbol{C}\circ\boldsymbol{C}=(\bar{\rho}_{ij})_{m\times m}$,其中 $\bar{\rho}_{ij}=\max\limits_k\left\{\min\left\{\rho_{ik},\rho_{kj}\right\}\right\}$,$i,j=1,2,\cdots,m$,称 $\boldsymbol{C}^2$ 是 $\boldsymbol{C}$ 的一个组成矩阵。

定义4[2] 设 $\boldsymbol{C}=(\rho_{ij})_{m\times m}$ 是一个相关系数矩阵,如果 $\boldsymbol{C}^2\subseteq\boldsymbol{C}$,使得 $\max\limits_k\left\{\min\left\{\rho_{ik},\rho_{kj}\right\}\right\}\leqslant\rho_{ij}$, $i,j=1,2,\cdots,m$,称 $\boldsymbol{C}$ 是一个等效相关矩阵。

定义5[2] 设 $\boldsymbol{C}=(\rho_{ij})_{m\times m}$ 是一个相关系数矩阵,通过有限次数的合成:$\boldsymbol{C}\rightarrow\boldsymbol{C}^2\rightarrow\boldsymbol{C}^4\rightarrow\cdots\rightarrow\boldsymbol{C}^{2^k}\rightarrow\cdots$,存在一个正整数 k,使得 $\boldsymbol{C}^{2^k}=\boldsymbol{C}^{2^{k+1}}$,则 $\boldsymbol{C}^{2^k}$ 是一个等效相关矩阵。

定义6[2] 设 $\boldsymbol{C}=(\rho_{ij})_{m\times m}$ 是一个等效相关矩阵,称 $\boldsymbol{C}_\lambda=(\lambda\rho_{ij})_{m\times m}$ 是 $\boldsymbol{C}$ 的一个 λ -截矩阵,其中

$$\lambda\rho_{ij}=\begin{cases}0, & 如果\quad \rho_{ij}<\lambda\\ 1, & 如果\quad \rho_{ij}\geqslant\lambda\end{cases}\quad i,j=1,2,\cdots,m,$$

$\lambda\in[0,1]$,为可信度水平。

2 改进的犹豫模糊集相关系数及算例分析

2.1 改进的犹豫模糊集相关系数

定义7 考虑到各属性权重不同以及决策者对犹豫模糊集隶属度差异部分和犹豫度差异部分的偏好程度不同,定义相关系数公式为

$$\rho_{\mathrm{WHFSX}}(A,B)=\alpha\rho_{\mathrm{WHFSX1}}(A,B)+\beta\rho_{\mathrm{WHFSX2}}(A,B), \tag{1}$$

其中,

$$\rho_{\mathrm{WHFSX1}}(A,B)=\frac{\sum\limits_{i=1}^{n}\omega_i\left[\frac{1}{l_{A_i}l_{B_i}}\sum\limits_{j=1}^{l_{A_i}}\sum\limits_{k=1}^{l_{B_i}}\left(h_{A_j}(x_i)h_{B_k}(x_i)\right)\right]}{\sqrt{\sum\limits_{i=1}^{n}\omega_i\left[\frac{1}{l_{A_i}^2}\sum\limits_{j_1=1}^{l_{A_i}}\sum\limits_{j_2=1}^{l_{A_i}}\left(h_{A_{j_1}}(x_i)h_{A_{j_2}}(x_i)\right)\right]}\sqrt{\sum\limits_{i=1}^{n}\omega_i\left[\frac{1}{l_{B_k}^2}\sum\limits_{k_1=1}^{l_{B_i}}\sum\limits_{k_2=1}^{l_{B_i}}\left(h_{B_{k_1}}(x_i)h_{B_{k_2}}(x_i)\right)\right]}},$$

$$\rho_{\mathrm{WHFSX2}}(A,B)=\frac{\sum\limits_{i=1}^{n}\omega_i\left[S\left(A\left(x_i\right)\right)S\left(B\left(x_i\right)\right)\right]}{\sqrt{\sum\limits_{i=1}^{n}\omega_i\left(S^2(A(x_i))\right)}\sqrt{\sum\limits_{i=1}^{n}\omega_i\left(S^2(B(x_i))\right)}},$$

$$S(A(x_i))=\sqrt{\frac{\sum\limits_{j=1}^{l(h_A(x_i))}\left|h_{A_j}(x_i)-\bar{h}_A(x_i)\right|^2}{l(h_A(x_i))}},$$

$\alpha,\beta\geqslant0$ 且 $\alpha+\beta=1$。$\rho_{\mathrm{WHFSX1}}(A,B)$ 和 $\rho_{\mathrm{WHFSX2}}(A,B)$ 分别代表犹豫模糊集隶属度部分的相关性和犹豫度部分的相关性。

为保证其普适性,需要考虑分母为0的特殊情况,即2个犹豫模糊集中各个属性对应元中的隶属度值的个数都为1或其中一个为1的两种特殊情况。当属性相对应的两个犹豫模糊元中数值个数都为1时,则规定在犹豫度部分其相关性为1;当各个属性对应元 $h_A(x_i)$ 中的数值个数都为1,在 $h_B(x_i)$ 中其数值个数不都为1

时，其犹豫度部分的相关系数定义如下：

$$\rho_{\mathrm{WHFS}X_2}(A,B)=\sum_{i=1}^{n}\omega_i\frac{1}{l_{B_i}}=\rho_{\mathrm{WHFS}X_2}(B,A),$$

当各个属性对应元$h_B(x_i)$中的数值个数都为1，在$h_A(x_i)$中其数值个数不都为1时，其犹豫度部分的相关系数定义如下：

$$\rho_{\mathrm{WHFS}X_2}(A,B)=\sum_{i=1}^{n}\omega_i\frac{1}{l_{A_i}}=\rho_{\mathrm{WHFS}X_2}(B,A),$$

其中l_{A_i}代表$h_A(x_i)$中数值的个数，l_{B_i}代表$h_B(x_i)$中数值的个数。

定理 1[2]　设A、B、C是非空集合$X=\{x_1,x_2,\cdots,x_n\}$上的3个犹豫模糊集，$\rho(A,B)$为A和B之间的相关系数，则$\rho(A,B)$满足以下3个性质：1）$0\leqslant\rho(A,B)\leqslant 1$；2）$\rho(A,B)=\rho(B,A)$；3）$\rho(A,B)=1$，如果$A=B$。

证明　改进的犹豫模糊集相关系数计算公式显然满足定理1中性质的后两条，根据柯西–施瓦茨不等式，性质（1）证明如下：

$$C^2_{\mathrm{WHFS}X1}(A,B)=\left[\sum_{i=1}^{n}\omega_i\left[\frac{1}{l_{A_i}l_{B_i}}\sum_{j=1}^{l_{A_i}}\sum_{k=1}^{l_{B_i}}\left(h_{A_j}(x_i)h_{B_k}(x_i)\right)\right]\right]^2=$$

$$\left[\sum_{i=1}^{n}\left[\frac{\sqrt{\omega_i}}{l_{A_i}}\sum_{j=1}^{l_{A_i}}h_{A_j(x_i)}\cdot\frac{\sqrt{\omega_i}}{l_{B_i}}\sum_{k=1}^{l_{B_i}}h_{B_k(x_i)}\right]\right]^2\leqslant$$

$$\left[\sum_{i=1}^{n}\frac{\omega_i}{l^2_{A_i}}\sum_{j_1=1}^{l_{A_i}}\sum_{j_2=1}^{l_{A_i}}h_{A_{j_1}}(x_i)h_{A_{j_2}}(x_i)\right]\cdot\left[\sum_{i=1}^{n}\frac{\omega_i}{l^2_{B_i}}\sum_{k_1=1}^{l_{B_i}}\sum_{k_2=1}^{l_{B_i}}h_{B_{k_1}}(x_i)h_{B_{k_2}}(x_i)\right]=$$

$$\left[\sum_{i=1}^{n}\omega_i\left(\frac{1}{l^2_{A_i}}\sum_{j_1=1}^{l_{A_i}}\sum_{j_2=1}^{l_{A_i}}h_{A_{j_1}}(x_i)h_{A_{j_2}}(x_i)\right)\right]\cdot\left[\sum_{i=1}^{n}\omega_i\left(\frac{1}{l^2_{B_i}}\sum_{k_1=1}^{l_{B_i}}\sum_{k_2=1}^{l_{B_i}}h_{B_{k_1}}(x_i)h_{B_{k_2}}(x_i)\right)\right]=$$

$$C_{\mathrm{WHFS}X1}(A,A)\cdot C_{\mathrm{WHFS}X1}(B,B)。$$

$$C^2_{\mathrm{WHFS}X2}(A,B)=\left[\sum_{i=1}^{n}\omega_i\left[S(A(x_i))S(B(x_i))\right]\right]^2=$$

$$\left[\sqrt{\omega_1}\,S\left(A\left(x_1\right)\right)\cdot\sqrt{\omega_1}\,S\left(B\left(x_1\right)\right)+\cdots+\sqrt{\omega_n}\,S\left(A\left(x_n\right)\right)\cdot\sqrt{\omega_n}\,S\left(B\left(x_n\right)\right)\right]^2\leqslant$$

$$\left[\omega_1S^2\left(A\left(x_1\right)\right)+\omega_2S^2\left(A\left(x_2\right)\right)+\cdots+\omega_nS^2\left(A\left(x_n\right)\right)\right]\cdot$$

$$\left[\omega_1S^2\left(B\left(x_1\right)\right)+\omega_2S^2\left(B\left(x_2\right)\right)+\cdots+\omega_nS^2\left(B\left(x_n\right)\right)\right]=$$

$$\left[\sum_{i=1}^{n}\omega_iS^2\left(A\left(x_i\right)\right)\right]\cdot\left[\sum_{i=1}^{n}\omega_iS^2\left(B\left(x_i\right)\right)\right]=$$

$$C_{\mathrm{WHFS}X2}(A,A)\cdot C_{\mathrm{WHFS}X2}(B,B)。$$

由于给予两个部分偏好程度之和为1，所以综合可得

$$0\leqslant\rho_{\mathrm{WHFS}X}(A,B)\leqslant 1。$$

2.2 数值算例

应用文献[2]中的数值算例，为了更好地评估不同类型的软件$A_i(i=1,2,\cdots,7)$，根据功能性(x_1)、可用性(x_2)、可移植性(x_3)、成熟度(x_4)这4个属性对软件进行聚类。分别对4个属性赋予不同的权重$\omega=\{0.35,0.3,0.15,0.2\}^{\mathrm{T}}$，考虑到不同的评价专家具有不同的水平、背景、经验，会导致评估信息具有差异，为了准确反映专家意见，其评估信息用犹豫模糊集的形式表示在表1中。针对不同的偏好情况，将改进的新型相关系数计算公式(1)应用于HFSC算法[2]，得到的聚类结果如表2所示。

通过对比可以发现，在使用本文改进的相关系数公式进行聚类时，当赋予犹豫模糊集隶属度差异部分和犹豫度差异部分的权重比大于等于1时，聚类效果与文献[2]的聚类效果相近，当赋予两部分的权重比小于1时，聚类效果差异较大。造成差异的原因一方面在于计算相关系数时，文献[2]中的相关系数公式需要人为

表1 文献[2]算例的决策矩阵

Table 1 The decision matrix of the example in reference [2]

	x_1	x_2	x_3	x_4
A_1	$\{0.9,0.85,0.8\}$	$\{0.8,0.75,0.7\}$	$\{0.8,0.65\}$	$\{0.35,0.3\}$
A_2	$\{0.9,0.85\}$	$\{0.8,0.7,0.6\}$	$\{0.2\}$	$\{0.15\}$
A_3	$\{0.4,0.3,0.2\}$	$\{0.5,0.4\}$	$\{1.0,0.9\}$	$\{0.65,0.5,0.45\}$
A_4	$\{1.0,0.95,0.8\}$	$\{0.2,0.15,0.1\}$	$\{0.3,0.2\}$	$\{0.8,0.7,0.6\}$
A_5	$\{0.5,0.4,0.35\}$	$\{1.0,0.9,0.7\}$	$\{0.4\}$	$\{0.35,0.3,0.2\}$
A_6	$\{0.7,0.6,0.5\}$	$\{0.9,0.8\}$	$\{0.6,0.4\}$	$\{0.2,0.1\}$
A_7	$\{1,0.8\}$	$\{0.35,0.2,0.15\}$	$\{0.2,0.1\}$	$\{0.85,0.7\}$

表2 文献[2]算例的各偏好对应聚类情况

Table 2 The clustering of each preference of the example in reference [2]

偏好情况	第一次分类		第二次分类	
	λ的取值范围	分类情况	λ的取值范围	分类情况
0.1/0.9	[0.839 7, 0.924 4]	$\{A_1,A_3,A_4,A_5,A_6,A_7\},\{A_2\}$	[0.924 4, 0.926 6]	$\{A_1,A_6\},\{A_2\},\{A_3,A_4,A_5,A_7\}$
0.2/0.8	[0.844 9, 0.896 9]	$\{A_1,A_3,A_4,A_5,A_6A_7\},\{A_2\}$	[0.896 9, 0.910 7]	$\{A_1,A_3,A_4,A_6,A_7\},\{A_2\},\{A_5\}$
0.3/0.7	[0.850 1, 0.876 9]	$\{A_1,A_3,A_4,A_5,A_6,A_7\},\{A_2\}$	[0.878 6, 0.896 9]	$\{A_1,A_3,A_4,A_6,A_7\},\{A_2\},\{A_5\}$
0.5/0.5	[0.856 0, 0.860 4]	$\{A_1,A_2,A_3,A_5,A_6\},\{A_4,A_7\}$	[0.860 4, 0.869 5]	$\{A_1,A_3,A_5,A_6\},\{A_2\},\{A_4,A_7\}$
0.7/0.3	[0.843 0, 0.861 3]	$\{A_1,A_2,A_3,A_5,A_6\},\{A_4,A_7\}$	[0.861 3, 0.870 7]	$\{A_1,A_2,A_5,A_6\},\{A_3\},\{A_4,A_7\}$
0.8/0.2	[0.836 4, 0.859 7]	$\{A_1,A_2,A_3,A_5,A_6\},\{A_4,A_7\}$	[0.859 7, 0.890 2]	$\{A_1,A_2,A_5,A_6\},\{A_3\},\{A_4,A_7\}$
0.9/0.1	[0.829 9, 0.858 1]	$\{A_1,A_2,A_3,A_5,A_6\},\{A_4,A_7\}$	[0.858 1, 0.922 0]	$\{A_1,A_2,A_5,A_6\},\{A_3\},\{A_4,A_7\}$
文献[2]	[0.829 2, 0.846 1]	$\{A_1,A_2,A_3,A_5,A_6\},\{A_4,A_7\}$	[0.846 1, 0.953 1]	$\{A_1,A_2,A_5,A_6\},\{A_3\},\{A_4,A_7\}$

偏好情况	第三次分类		第四次分类	
	λ的取值范围	分类情况	λ的取值范围	分类情况
0.1/0.9	[0.926 6, 0.963 8]	$\{A_1,A_6\},\{A_2\},\{A_3,A_4,A_7\},\{A_5\}$	[0.963 8, 0.966 8]	$\{A_1,A_6\},\{A_2\},\{A_3\}\{A_4,A_7\},\{A_5\}$
0.2/0.8	[0.910 7, 0.930 6]	$\{A_1,A_6\},\{A_2\},\{A_3,A_4,A_7\},\{A_5\}$	[0.930 6, 0.969 9]	$\{A_1,A_6\},\{A_2\},\{A_3\}\{A_4,A_7\},\{A_5\}$
0.3/0.7	[0.896 9, 0.897 5]	$\{A_1,A_6\},\{A_2\},\{A_3,A_4,A_7\},\{A_5\}$	[0.897 5, 0.973 0]	$\{A_1,A_6\},\{A_2\},\{A_3\}\{A_4,A_7\},\{A_5\}$
0.5/0.5	[0.869 5, 0.886 7]	$\{A_1,A_5,A_6\},\{A_2\},\{A_3\},\{A_4,A_7\}$	[0.886 7, 0.972 5]	$\{A_1,A_6\},\{A_2\},\{A_3\}\{A_4,A_7\},\{A_5\}$
0.7/0.3	[0.870 7, 0.923 9]	$\{A_1,A_5,A_6\},\{A_2\},\{A_3\},\{A_4,A_7\}$	[0.923 9, 0.970 9]	$\{A_1,A_6\},\{A_2\},\{A_3\}\{A_4,A_7\},\{A_5\}$
0.8/0.2	[0.890 2, 0.942 4]	$\{A_1,A_5,A_6\},\{A_2\},\{A_3\},\{A_4,A_7\}$	[0.942 4, 0.970 1]	$\{A_1,A_6\},\{A_2\},\{A_3\}\{A_4,A_7\},\{A_5\}$
0.9/0.1	[0.922 0, 0.961 0]	$\{A_1,A_5,A_6\},\{A_2\},\{A_3\},\{A_4,A_7\}$	[0.961 0, 0.969 3]	$\{A_1,A_6\},\{A_2\},\{A_3\}\{A_4,A_7\},\{A_5\}$
文献[2]	[0.953 1, 0.968 6]	$\{A_1,A_5,A_6\},\{A_2\},\{A_3\},\{A_4,A_7\}$	[0.968 6, 0.970 2]	$\{A_1\},\{A_2\},\{A_3\}\{A_4,A_7\},\{A_5,A_6\}$

偏好情况	第五次分类	
	λ的取值范围	分类情况
0.1/0.9	[0.966 8, 0.975 7]	$\{A_1,A_6\},\{A_2\},\{A_3\},\{A_4\},\{A_5\},\{A_7\}$
0.2/0.8	[0.969 9, 0.974 9]	$\{A_1,A_6\},\{A_2\},\{A_3\},\{A_4\},\{A_5\},\{A_7\}$
0.3/0.7	[0.973 0, 0.974 1]	$\{A_1,A_6\},\{A_2\},\{A_3\},\{A_4\},\{A_5\},\{A_7\}$
0.5/0.5	[0.972 5, 0.979 1]	$\{A_1\},\{A_2\},\{A_3\}\{A_4,A_7\},\{A_5\},\{A_6\}$
0.7/0.3	[0.970 9, 0.985 2]	$\{A_1\},\{A_2\},\{A_3\}\{A_4,A_7\},\{A_5\},\{A_6\}$
0.8/0.2	[0.970 1, 0.988 3]	$\{A_1\},\{A_2\},\{A_3\}\{A_4,A_7\},\{A_5\},\{A_6\}$
0.9/0.1	[0.969 3, 0.991 3]	$\{A_1\},\{A_2\},\{A_3\}\{A_4,A_7\},\{A_5\},\{A_6\}$
文献[2]	[0.970 2, 0.990 6]	$\{A_1\},\{A_2\},\{A_3\}\{A_4,A_7\},\{A_5\},\{A_6\}$

主观填充数据，而改进的相关系数公式则避免了这一过程，保证了数据的真实准确性；另一方面，文献[2]中的相关系数只考虑一个影响相关系数的特征，如果只强调犹豫模糊元中数据的个体影响，其在一定程度上是正确的，当既考虑数据个体影响又考虑总体影响时，本文的聚类方法已涵盖了文献[2]中的情况，因此本文所提出的相关系数公式更为全面。

3 改进的区间值犹豫模糊集相关系数及算例分析

3.1 改进的区间值犹豫模糊集相关系数

定义8　将改进的犹豫模糊集相关系数的计算公式推广到区间值犹豫模糊集上，定义区间值犹豫模糊集相关系数计算公式为

$$\rho_{\mathrm{WIVHFS}X}(A,B)=\alpha\rho_{\mathrm{WIVHFS}X_1}(A,B)+\beta\rho_{\mathrm{WIVHFS}X_2}(A,B),\tag{2}$$

其中，

$$\rho_{\mathrm{WIVHFS}X_1}(A,B)=\frac{C_{\mathrm{WIVHFS}X_1}(A,B)}{\sqrt{C_{\mathrm{WIVHFS}X_1}(A,A)}\sqrt{C_{\mathrm{WIVHFS}X_1}(B,B)}},$$

$$C_{\mathrm{WIVHFS}X_1}(A,B)=\sum_{i=1}^{n}\omega_i\left[\frac{1}{l_{A_i}l_{B_i}}\sum_{j=1}^{l_{A_i}}\sum_{k=1}^{l_{B_i}}\left(\tilde{h}_{A_j}^{L}(x_i)\tilde{h}_{B_k}^{L}(x_i)+(\tilde{h}_{A_j}^{U}(x_i)\tilde{h}_{B_k}^{U}(x_i)\right)\right],$$

$$C_{\mathrm{WIVHFS}X_1}(A,A)=\sum_{i=1}^{n}\omega_i\left[\frac{1}{l_{A_i}^2}\sum_{j_1=1}^{l_{A_i}}\sum_{j_2=1}^{l_{A_i}}\left(\tilde{h}_{A_{j_1}}^{L}(x_i)\tilde{h}_{A_{j_2}}^{L}(x_i)+(\tilde{h}_{A_{j_1}}^{U}(x_i)\tilde{h}_{A_{j_2}}^{U}(x_i)\right)\right],$$

$$C_{\mathrm{WIVHFS}X_1}(B,B)=\sum_{i=1}^{n}\omega_i\left[\frac{1}{l_{B_k}^2}\sum_{k_1=1}^{l_{B_k}}\sum_{k_2=1}^{l_{B_k}}\left(\tilde{h}_{B_{k_1}}^{L}(x_i)\tilde{h}_{B_{k_2}}^{L}(x_i)+(\tilde{h}_{B_{k_1}}^{U}(x_i)\tilde{h}_{B_{k_2}}^{U}(x_i)\right)\right];$$

$$\rho_{\mathrm{WIVHFS}X_2}(A,B)=\frac{\sum_{i=1}^{n}\omega_i\left[S\left(A\left(x_i\right)\right)S\left(B\left(x_i\right)\right)\right]}{\sqrt{\sum_{i=1}^{n}\omega_i\left(S^2(A(x_i))\right)}\sqrt{\sum_{i=1}^{n}\omega_i\left(S^2(B(x_i))\right)}},$$

$\alpha,\beta\geqslant 0$且$\alpha+\beta=1$。$\rho_{\mathrm{WIVHFS}X_1}(A,B)$和$\rho_{\mathrm{WIVHFS}X_2}(A,B)$分别代表区间值犹豫模糊集隶属度部分的相关性。

对于区间部分的减法，采用左右端点部分等偏好的方法，具体形式如下：

$$\left|\left[a^L,a^U\right]-\left[b^L,b^U\right]\right|=0.5\left|a^L-b^L\right|+0.5\left|a^U-b^U\right|。$$

根据改进相关系数公式的定义，为了保证其普适性，需要考虑分母为0的特殊情况．即2个区间值犹豫模糊集中各属性对应元中的区间个数都为1或其中一个为1这两种情况。当属性对应的区间值犹豫模糊元中区间个数都为1时，则规定在犹豫度部分其相关性为1；当各个属性对应元$\tilde{h}_A(x_i)$中的区间个数都为1，在$\tilde{h}_B(x_i)$中其数值个数不都为1时，其犹豫度部分的相关系数定义如下：

$$\rho_{\mathrm{WIVHFS}X_2}(A,B)=\sum_{i=1}^{n}\omega_i\frac{1}{l_{B_i}}=\rho_{\mathrm{WIVHFS}X_2}(B,A),$$

当各个属性对应元$\tilde{h}_B(x_i)$中的区间个数都为1，在$\tilde{h}_A(x_i)$中其数值个数不都为1时，其犹豫度部分的相关系数定义如下：

$$\rho_{\mathrm{WIVHFS}X_2}(A,B)=\sum_{i=1}^{n}\omega_i\frac{1}{l_{A_i}}=\rho_{\mathrm{WIVHFS}X_2}(B,A),$$

其中l_{A_i}代表$\tilde{h}_A(x_i)$中区间的个数，l_{B_i}代表$\tilde{h}_B(x_i)$中区间的个数。

证明　改进的区间值犹豫模糊集相关系数计算公式显然满足定理1中性质的后两条，性质(1)证明如下：

根据柯西施瓦茨不等式可以得到：

$$C_{\mathrm{WIVHFS}X_1}(A,B)=\sum_{i=1}^{n}\omega_i\left[\frac{1}{l_{A_i}l_{B_i}}\sum_{j=1}^{l_{A_i}}\sum_{k=1}^{l_{B_i}}\left(\tilde{h}_{A_j}^{L}(x_i)\tilde{h}_{B_k}^{L}(x_i)+(\tilde{h}_{A_j}^{U}(x_i)\tilde{h}_{B_k}^{U}(x_i)\right)\right]=$$

$$\sum_{i=1}^{n}\omega_i\frac{1}{l_{A_i}l_{B_i}}\sum_{j=1}^{l_{A_i}}\sum_{k=1}^{l_{B_i}}\tilde{h}_{A_j}^{L}(x_i)\tilde{h}_{B_k}^{L}(x_i)+\sum_{i=1}^{n}\omega_i\frac{1}{l_{A_i}l_{B_i}}\sum_{j=1}^{l_{A_i}}\sum_{k=1}^{l_{B_i}}\tilde{h}_{A_j}^{U}(x_i)\tilde{h}_{B_k}^{U}(x_i)=$$

$$\sum_{i=1}^{n}\left[\frac{\sqrt{\omega_i}}{l_{A_i}}\sum_{j=1}^{l_{A_i}}\tilde{h}_{A_j}^{L}(x_i)\cdot\frac{\sqrt{\omega_i}}{l_{B_i}}\sum_{k=1}^{l_{B_i}}\tilde{h}_{B_k}^{L}(x_i)\right]+\sum_{i=1}^{n}\left[\frac{\sqrt{\omega_i}}{l_{A_i}}\sum_{j=1}^{l_{A_i}}\tilde{h}_{A_j}^{U}(x_i)\cdot\frac{\sqrt{\omega_i}}{l_{\tilde{B}_i}}\sum_{k=1}^{l_{B_i}}\tilde{h}_{B_k}^{U}(x_i)\right]\leqslant$$

$$\sqrt{\sum_{i=1}^{n}\left[\frac{\sqrt{\omega_i}}{l_{A_i}}\sum_{j=1}^{l_{A_i}}\tilde{h}_{A_j}^{L}(x_i)\right]^2}\cdot\sqrt{\sum_{i=1}^{n}\left[\frac{\sqrt{\omega_i}}{l_{B_i}}\sum_{k=1}^{l_{B_i}}\tilde{h}_{B_k}^{L}(x_i)\right]^2}+\sqrt{\sum_{i=1}^{n}\left[\frac{\sqrt{\omega_i}}{l_{A_i}}\sum_{j=1}^{l_{A_i}}\tilde{h}_{A_j}^{U}(x_i)\right]^2}\cdot\sqrt{\sum_{i=1}^{n}\left[\frac{\sqrt{\omega_i}}{l_{B_i}}\sum_{k=1}^{l_{B_i}}\tilde{h}_{B_k}^{U}(x_i)\right]^2}\leqslant$$

$$\sqrt{C_{\mathrm{WIVHFS}X_1}(A,A)}\cdot\sqrt{C_{\mathrm{WIVHFS}X_1}(B,B)}。$$

对$0\leqslant\rho_{\mathrm{WIVHFS}X_2}(A,B)\leqslant 1$的证明与对$0\leqslant\rho_{\mathrm{WHFS}X_2}(A,B)\leqslant 1$的证明类似,综合可得$0\leqslant\rho_{\mathrm{WIVHFS}X}(A,B)\leqslant 1$。

3.2 数值算例

一个汽车市场想要将6辆不同的汽车$A_i(i=1,2,\cdots,6)$进行分类,以下面4个属性为衡量指标:安全性(x_1)、燃油经济性(x_2)、设计(x_3)、价格(x_4),分别对4个属性赋予不同的权重,权重向量为$\omega=\{0.35,0.2,0.2,0.25\}^{\mathrm{T}}$,每辆车的评估数据用区间数的形式表达,如表3所示,表中数据代表专家对各个备选方案中各属性的满意程度,因此记为[0,1]内的区间值数字。

针对不同的偏好情况,将改进的新型相关系数计算公式(2)应用于HFSC算法[2],得到的聚类结果如表4所示。

表3 本文算例的决策矩阵

Table 3 The decision matrix of the example in this paper

	x_1	x_2	x_3	x_4
A_1	{[0.2,0.4],[0.5,0.7]}	{[0.1,0.3],[0.4,0.5]}	{[0.5,0.7],[0.8,0.9]}	{[0.6,0.8]}
A_2	{[0.3,0.5],[0.6,0.7]}	{[0.2,0.4]}	{[0.6,0.8]}	{[0.5,0.7],[0.7,0.8]}
A_3	{[0.4,0.6]}	{[0.3,0.4],[0.5,0.6]}	{[0.4,0.6],[0.7,0.8]}	{[0.4,0.6]}
A_4	{[0.3,0.6],[0.7,0.8]}	{[0.2,0.4],[0.5,0.7]}	{[0.5,0.6],[0.8,0.9]}	{[0.4,0.5],[0.7,0.8]}
A_5	{[0.2,0.4],[0.6,0.7]}	{[0.2,0.3],[0.4,0.6],[0.7,0.8]}	{[0.4,0.7]}	{[0.5,0.7]}
A_6	{[0.5,0.7]}	{[0.5,0.7]}	{[0.5,0.7],[0.8,0.9]}	{[0.4,0.7]}

表4 本文算例的各偏好对应聚类情况

Table 4 The clustering of each preference of the example in this paper

偏好情况	第一次分类		第二次分类	
	λ的取值范围	分类情况	λ的取值范围	分类情况
0.1/0.9	[0.696 6, 0.736 0]	$\{A_1,A_2,A_4,A_5\},\{A_3,A_6\}$	[0.736 0, 0.777 2]	$\{A_1,A_2,A_4,A_5\},\{A_3\},\{A_6\}$
0.2/0.8	[0.728 0, 0.765 0]	$\{A_1,A_2,A_4,A_5\},\{A_3,A_6\}$	[0.765 0, 0.800 7]	$\{A_1,A_2,A_4,A_5\},\{A_3\},\{A_6\}$
0.3/0.7	[0.759 5, 0.793 9]	$\{A_1,A_2,A_4,A_5\},\{A_3,A_6\}$	[0.793 9, 0.824 1]	$\{A_1,A_2,A_4,A_5\},\{A_3\},\{A_6\}$
0.5/0.5	[0.822 3, 0.851 8]	$\{A_1,A_2,A_4,A_5\},\{A_3,A_6\}$	[0.851 8, 0.871 0]	$\{A_1,A_2,A_4,A_5\},\{A_3\},\{A_6\}$
0.7/0.3	[0.887 6, 0.909 6]	$\{A_1,A_2,A_4,A_5\},\{A_3,A_6\}$	[0.909 6, 0.917 9]	$\{A_1,A_2,A_4,A_5\},\{A_3\},\{A_6\}$
0.8/0.2	[0.924 1, 0.938 6]	$\{A_1,A_2,A_4,A_5\},\{A_3,A_6\}$	[0.938 6, 0.941 3]	$\{A_1,A_2,A_4,A_5\},\{A_3\},\{A_6\}$
0.9/0.1	[0.960 5, 0.964 7]	$\{A_1,A_2,A_4,A_5\},\{A_3,A_6\}$	[0.964 7, 0.965 0]	$\{A_1,A_4,A_5\},\{A_2\},\{A_3,A_6\}$

偏好情况	第三次分类		第四次分类	
	λ的取值范围	分类情况	λ的取值范围	分类情况
0.1/0.9	[0.777 2, 0.874 2]	$\{A_1,A_4,A_5\},\{A_2\},\{A_3\},\{A_6\}$	[0.874 2, 0.881 3]	$\{A_1,A_5\},\{A_2\},\{A_3\},\{A_4\},\{A_6\}$
0.2/0.8	[0.800 7, 0.886 0]	$\{A_1,A_4,A_5\},\{A_2\},\{A_3\},\{A_6\}$	[0.886 0, 0.891 7]	$\{A_1,A_5\},\{A_2\},\{A_3\},\{A_4\},\{A_6\}$
0.3/0.7	[0.824 1, 0.897 7]	$\{A_1,A_4,A_5\},\{A_2\},\{A_3\},\{A_6\}$	[0.897 7, 0.902 2]	$\{A_1,A_5\},\{A_2\},\{A_3\},\{A_4\},\{A_6\}$
0.5/0.5	[0.871 0, 0.921 3]	$\{A_1,A_4,A_5\},\{A_2\},\{A_3\},\{A_6\}$	[0.921 3, 0.923 1]	$\{A_1,A_5\},\{A_2\},\{A_3\},\{A_4\},\{A_6\}$
0.7/0.3	[0.917 9, 0.944 1]	$\{A_1,A_4,A_5\},\{A_2\},\{A_3\},\{A_6\}$	[0.944 1, 0.944 8]	$\{A_1,A_4\},\{A_2\},\{A_3\},\{A_5\},\{A_6\}$
0.8/0.2	[0.941 3, 0.954 5]	$\{A_1,A_4,A_5\},\{A_2\},\{A_3\},\{A_6\}$	[0.954 5, 0.956 6]	$\{A_1,A_4\},\{A_2\},\{A_3\},\{A_5\},\{A_6\}$
0.9/0.1	[0.965 0, 0.967 5]	$\{A_1,A_4\},\{A_2\},\{A_3,A_6\},\{A_5\}$	[0.967 5, 0.968 3]	$\{A_1,A_4\},\{A_2\},\{A_3\},\{A_5\},\{A_6\}$

将得到的犹豫模糊集之间的相关系数计算公式推广到区间上,通过数值计算可以得到,在区间值犹豫模糊集中当赋予$\rho_{WIVHFSX1}$和$\rho_{WIVHFSX2}$ 2个部分不同的偏好时得到的聚类情况也有所不同。对比已有研究中只考虑了对应区间值犹豫模糊元之间的隶属度值对总体相关性带来的影响,本文还将对应区间值犹豫模糊元的内部差异作为衡量相关性的指标,考虑了2个影响相关系数的特征,比传统相关系数只考虑单一特征更为全面。

4 结论

改进的相关系数计算公式弥补了现有公式中需要进行主观填充的不足,另外,将犹豫度作为衡量相关性的另一指标加入公式中,把内部偏差考虑在内,用于衡量内部总体差异。无论是犹豫模糊集还是区间值犹豫模糊集,通过数值计算可以发现,当赋予隶属度差异部分和犹豫度差异部分不同的权重偏好时得到的聚类结果不同。本文的聚类方法已涵盖了只考虑单一特征的情况,另外,通过理论证明了新公式的科学性,因此本文改进的相关系数公式是合理有效的。

参考文献:

[1] TORRA V. Hesitant fuzzy sets[J]. International Journal of Intelligent Systems,2010,25(6):529-539.

[2] CHEN N,XU Z S,XIA M M. Correlation coefficients of hesitant fuzzy sets and their applications to clustering analysis[J]. Applied Mathematical Modelling,2013,37(4):2197-2211.

[3] 蔡丽娜.区间值犹豫模糊集及其在决策中的应用研究[D].郑州:郑州大学,2013.

[4] XU Z S,XIA M M. On distance and correlation measures of hesitant fuzzy information[J]. International Journal of Intelligent Systems,2011,26(5):410-425.

[5] LIAO H C,XU Z S,ZENG X J. Novel correlation coefficients between hesitant fuzzy sets and their application in decision making [J]. Knowledge-Based Systems,2015,82:115-127.

[6] GUAN X,SUN G D,YI X,et al. Synthetic correlation coefficient between hesitant fuzzy sets with applications[J]. International Journal of Fuzzy Systems,2018,20(6):1968-1985.

[7] YANG J X,TANG X A,YANG S L. Novel correlation coefficients for hesitant fuzzy sets and their applications to supplier selection and medical diagnosis[J]. Journal of Intelligent & Fuzzy Systems,2018,35(6):6427-6441.

[8] 陈娜.犹豫模糊环境下的决策方法及聚类算法研究[D].南京:东南大学,2015.

[9] MENG F Y,CHEN X H. Correlation coefficients of hesitant fuzzy sets and their application based on fuzzy measures[J]. Cognitive Computation,2015,7(4):445-463.

[10] YANG X,XU Z S,LIAO H C. Correlation coefficients of hesitant multiplicative sets and their applications in decision making and clustering analysis[J]. Applied Soft Computing,2017,61:935-946.

[11] 汪峰,毛军军,祖璇,等.犹豫模糊信息下的协相关度与聚类分析[J].计算机科学与探索,2018,12(5):828-838

[12] ASIM A,NASAR R,RASHID T. Correlation coefficient of intuitionistic hesitant fuzzy sets based on informational energy and their applications to clustering analysis[J]. Soft Computing,2019,23(20):10393-10406.

[13] PARK D G,KWUN Y C,PARK J H,et al. Correlation coefficient of interval-valued intuitionistic fuzzy sets and its application to multiple attribute group decision making problems[J]. Mathematical and Computer Modelling,2009,50(9/10):1279-1293.

[14] MENG F Y,WANG C,CHEN X H,et al. Correlation coefficients of interval-valued hesitant fuzzy sets and their application based on the shapley function[J]. International Journal of Intelligent Systems,2016,31(1):17-43.

(责任编辑:刘 炜)