文章编号: 1001-0920 (2014) 10-1816-07 DOI: 10.13195/j.kzyjc.2013.0687

基于对称交互熵的犹豫模糊信息相似度及聚类应用

刘小弟[1,2], 朱建军[1], 刘思峰[1]

(1. 南京航空航天大学 经济与管理学院，南京 211106；2. 安徽工业大学 数理学院，安徽 马鞍山 243002)

摘 要: 研究面向犹豫模糊信息的聚类方法. 首先, 定义犹豫模糊相对熵、对称交互熵, 并基于信息论的角度提出一个新的犹豫模糊相似度公式; 然后, 利用相似度公式构造相似系数矩阵, 基于编网聚类方法对犹豫模糊集进行聚类; 最后, 通过算例验证了所提出方法的有效性.

关键词: 犹豫模糊集；对称交互熵；聚类

中图分类号: C934 **文献标志码:** A

Similarity measure of hesitant fuzzy sets based on symmetric cross entropy and its application in clustering analysis

LIU Xiao-di[1,2], *ZHU Jian-jun*[1], *LIU Si-feng*[1]

(1. College of Economics and Management，Nanjing University of Aeronautics and Astronautics，Nanjing 211106, China；2. School of Mathematics and Physics，Anhui University of Technology，Ma'anshan 243002，China. Correspondent：LIU Xiao-di，E-mail：lxy1160@163.com)

Abstract: The clustering method is studied under the hesitant fuzzy environment. Firstly, the hesitant fuzzy relative entropy and symmetric cross entropy are defined. From the viewpoint of information theory, a new similarity measure between two hesitant fuzzy sets is proposed. Then, a similarity coefficient matrix is constructed, and a netting method is presented to make clustering analysis for hesitant fuzzy sets. Finally, a numerical example is given to verify the effectiveness of the proposed method.

Key words: hesitant fuzzy set；symmetric cross entropy；clustering

0 引 言

在经济管理决策问题中, 由于人类思维的复杂性和个人素质的差异, 对于同一问题往往有不同的看法, 并且常常难以达成一致. 例如一个决策小组在讨论某一元素属于某一集合的隶属度时, 有些人给出 0.6, 有些人给出 0.7, 还有一些人给出 0.8, 决策者之间各执己见, 难以说服彼此. 为了应对这种复杂情形, Torra 等[1-2]提出了犹豫模糊集, 并讨论了犹豫模糊集与直觉模糊集[3]、2 型模糊集[4]、模糊多集[5]之间的区别与联系. 作为模糊集[6]的一种拓展形式, 犹豫模糊集允许一个元素属于一个集合的隶属度可以是几个可能的值, 这样可以反映与兼顾决策者的不同偏好. 自从犹豫模糊集诞生以来, 已经引起许多学者的注意, 并被应用到各种不同的领域当中, 如多属性决策[7-8]、医疗诊断[9]、聚类分析[10]等. 在处理多属性决策问题时, 集结算子是一种简便快速的方法, 一些学者将其应用到犹豫模糊集中, 提出犹豫模糊加权算子[11]、犹豫模糊拟算子[12]、犹豫模糊优先集结算子[13]、诱导犹豫模糊集结算子[14-15]等. 上述算子在对犹豫模糊信息集结时, 只考虑了属性本身的重要性, 而忽视了属性之间的联系. 为了克服此缺陷, 文献 [16-17] 分别提出犹豫模糊与广义犹豫模糊 Bonferroni 算子, 并将其应用到多属性决策问题中. 距离和相似度在决策分析[18]、医疗诊断[19]、模式识别[20]等领域已经得到了广泛的应用, 其中应用最多的距离有海明距离、欧式距离、豪斯道夫距离等. 文献 [21] 将上述距离推广到犹豫模

收稿日期: 2013-05-26；**修回日期:** 2013-10-18.

基金项目: 国家自然科学基金项目(71171112)；江苏省高校哲学社会科学重点项目(2012ZDIXM007)；江苏省高校哲学社会科学重点研究基地重大项目(2012JDXM0030)；江苏省普通高校研究生科研创新计划项目(CXLX13_171)；中央高校基本科研业务费专项资金项目(NS2014086)；安徽工业大学青年教师科研基金项目(QZ201321).

作者简介: 刘小弟(1981−), 男, 讲师, 博士生, 从事多属性决策、复杂系统建模的研究；朱建军(1976−), 男, 教授, 博士生导师, 从事多属性决策理论与方法、智能优化算法、灰色系统理论等研究.

糊环境下, 给出了犹豫模糊集的距离测度公式; 另外, 为了减少过大或过小的偏差对决策结果的影响, 定义了有序距离公式, 并基于距离公式给出相应的犹豫模糊集的相似度公式. 文献 [8] 将模糊集的熵、交叉熵推广到犹豫模糊环境下, 定义了犹豫模糊集的熵、交叉熵公式, 同时讨论了犹豫模糊集的相似度、熵、交叉熵之间的关系, 并最终将其应用到多属性决策中.

聚类分析是按照数学的方法来解决给定样本的分类问题, 依靠样本的相似性作为类属划分原则, 而选择合适的样本相似性度量方法和聚类方法是聚类分析中需要解决的两个重要问题. 目前, 关于犹豫模糊信息的相似性度量方法和聚类方法研究较少. 文献 [21] 提出基于距离的犹豫模糊相似度公式, 但在度量样本的相似性时, 存在度量结果有时与事实相违背、分辨率不够高等缺点; 文献 [22] 基于传统的凝聚层次聚类法[23](agglomerative hierarchical clustering) 对犹豫模糊集进行聚类分析, 利用犹豫模糊平均算子[11]反复计算聚类中心, 计算量较大; 文献 [24] 提出一种犹豫模糊最小生成树 (minimal spanning tree) 聚类算法, 但基于文献 [21] 中的距离公式得到的犹豫模糊集距离同样存在分辨率不高, 甚至有时与实际不符等缺陷; 文献 [10] 定义了犹豫模糊集的关联系数公式, 利用关联系数公式获得犹豫模糊关联系数矩阵, 并基于等价关系对犹豫模糊集进行聚类分析, 但是采用其中的犹豫模糊关联公式度量样本的相似性时, 存在聚类结果不够精细、划分的类别有时与事实不符等缺点. 另外, 为了获得等价关联矩阵, 需要对关联系数矩阵不断迭代, 这样容易造成信息丢失, 且计算量大[22]. 针对现有方法存在的问题, 本文从信息论的角度提出犹豫模糊相对熵、对称交互熵, 结合 TOPSIS 等方法的思想提出新的犹豫模糊相似度来度量样本的相似性, 并基于传统的编网聚类法[25]给出更加快速有效的犹豫模糊信息聚类方法.

1 基本概念

定义 1[1-2] 设 $X=\{x_1,x_2,\cdots,x_m\}$ 为一个非空集合, 则称从 X 到 $[0,1]$ 的一个子集的函数为犹豫模糊集, 记作

$$h_A(X)=\{\langle x,h_A(x)\rangle|x\in X\}. \tag{1}$$

其中 $h_A(x)$ 是 $[0,1]$ 中几个可能的数的集合, 表示 $x\in X$ 属于集合 A 的可能的程度.

定义 2[1-2] 设 $X=\{x_1,x_2,\cdots,x_m\}$ 为一个非空集合, 则称

$$h_A^c(X)=\{\langle x,h_A^c(x)\rangle|x\in X\} \tag{2}$$

为犹豫模糊集 $h_A(X)$ 的补, 其中 $h_A^c(x)=\bigcup\limits_{\gamma\in h_A(x)}\{1-\gamma\}$.

定义在 $x\in X$ 上的犹豫模糊集 $h_A(x)$、$h_B(x)$, 其元素及元素的个数可能不同, 为了进行有效计算, 作如下规定[9,21]: 将 $h_A(x)$、$h_B(x)$ 中的元素按递增顺序排列, 即 $h_A^{\tau(j)}(x)$ 和 $h_B^{\tau(j)}(x)$ 分别表示 $h_A(x)$ 和 $h_B(x)$ 中第 j 小的元素. 当且仅当 $h_A^{\tau(j)}(x)=h_B^{\tau(j)}(x)(j=1,2,\cdots,n)$ 时, 有 $h_A(x)=h_B(x)$. 设 l_A 和 l_B 分别表示 $h_A(x)$ 和 $h_B(x)$ 中元素的个数, $l=\max\{l_A,l_B\}$, 在元素少的集合里添加元素使得该集合的元素个数达到 l, 添加的原则反应决策者风险偏好, 喜好风险的决策者会对预期结果有比较乐观的估计, 则添加集合中值较大的元素, 而厌恶风险的决策者正好相反. 不失一般性, 本文按照第 1 种形式添加元素. 例如已知两个犹豫模糊集 $h_A(x)=\{0.3,0.5,0.6\}$, $h_B(x)=\{0.2,0.4\}$, 为了便于运算, 将 $h_B(x)$ 延伸为 $h_B(x)=\{0.2,0.4,0.4\}$.

例 1 设 $X=\{x_1,x_2,x_3\}$, 犹豫模糊集 $h_A(X)=\{\langle x_1,\{0.1,0.3\}\rangle, \langle x_2,\{0.6,0.7\}\rangle, \langle x_3,\{0.4,0.5,0.8\}\rangle\}$, 其中犹豫模糊集 $\langle x_1,\{0.1,0.3\}\rangle$ 表示决策小组给出元素 x_1 属于集合 A 的可能的程度. 决策小组有两种观点, 即对于 x_1 属于集合 A 的可能的程度, 分别有 0.1、0.3 两种, 表明决策小组意见不一致, 则犹豫模糊集 $h_A(X)$ 的补为 $h_A^c(X)=\{\langle x_1,\{0.7,0.9\}\rangle,\langle x_2,\{0.3,0.4\}\rangle,\langle x_3,\{0.2,0.5,0.6\}\rangle\}$.

距离和相似度是模糊集理论研究中的重要内容, 在多属性决策、医疗诊断、模式识别等领域已经得到广泛的应用. 文献 [21] 给出了如下犹豫模糊信息下的距离与相似度公理化定义.

定义 3[21] 设 $h_A(X)$、$h_B(X)$ 为定义在 $X=\{x_1,x_2,\cdots,x_m\}$ 上的两个犹豫模糊集, 则距离测度 $d(h_A(X),h_B(X))$ 满足如下条件:

1) $0\leqslant d(h_A(X),h_B(X))\leqslant 1$;

2) 当且仅当 $h_A(X)=h_B(X)$ 时, 有 $d(h_A(X),h_B(X))=0$;

3) $d(h_A(X),h_B(X))=d(h_B(X),h_A(X))$.

定义 4[21] 设 $h_A(X)$、$h_B(X)$ 为定义在 $X=\{x_1,x_2,\cdots,x_m\}$ 上的两个犹豫模糊集, 则相似度 $s(h_A(X),h_B(X))$ 满足如下条件:

1) $0\leqslant s(h_A(X),h_B(X))\leqslant 1$;

2) 当且仅当 $h_A(X)=h_B(X)$ 时, 有 $s(h_A(X),h_B(X))=1$;

3) $s(h_A(X),h_B(X))=s(h_B(X),h_A(X))$.

基于定义 3、定义 4, 文献 [21] 给出犹豫模糊集的距离与相似度关系为

$$s(h_A(X),h_B(X))=1-d(h_A(X),h_B(X)). \tag{3}$$

以上是利用距离度量集合的相似程度, 本文将从信息论的角度, 定义一种新的度量集合相似程度的公式. 首先引入相对熵概念.

定义 5[26-27] 对于两个系统 $M=\{M_1,M_2,\cdots,$

$M_n\}, N=\{N_1, N_2, \cdots, N_n\}$, 其系统状态 M_i 与 N_i 之间的差异程度可表示为

$$D=\sum_{i=1}^{n}\Big[M_i\log\frac{M_i}{N_i}+(1-M_i)\log\frac{1-M_i}{1-N_i}\Big]. \tag{4}$$

则称 D 为系统 M 对 N 的相对熵, 且 D 越小, M 与 N 之间的差异程度越小, 当 $M=N$ 时, $D=0$.

由于相对熵 D 不满足对称性, 不是真正意义上的距离, 但是用相对熵表示差异程度比传统距离表示(如欧式距离等)的分辨率更高[27]. 目前, 相对熵在多属性决策[27-28]、信号处理[29]等领域已经得到广泛的应用.

2 主要结论及方法

2.1 犹豫模糊相对熵

基于定义5, 首先给出犹豫模糊环境下的相对熵.

定 义6 设 $h_A(x)=\{h_A^{\tau(1)}(x), h_A^{\tau(2)}(x),\cdots, h_A^{\tau(n)}(x)\}$、$h_B(x)=\{h_B^{\tau(1)}(x), h_B^{\tau(2)}(x),\cdots, h_B^{\tau(n)}(x)\}$ 为两个犹豫模糊集, 则称

$$R(h_A(x),h_B(x))=\\ \sum_{j=1}^{n}\Big[h_A^{\tau(j)}(x)\log\frac{h_A^{\tau(j)}(x)}{h_B^{\tau(j)}(x)}+\\ (1-h_A^{\tau(j)}(x))\log\frac{1-h_A^{\tau(j)}(x)}{1-h_B^{\tau(j)}(x)}\Big] \tag{5}$$

为定义在 $x\in X$ 上, 犹豫模糊集 $h_A(x)$ 对 $h_B(x)$ 的相对熵. 当 $h_A^{\tau(j)}(x)=h_B^{\tau(j)}(x)(j=1,2,\cdots,n)$, 即 $h_A(x)=h_B(x)$ 时, $R(h_A(x),h_B(x))=0$. 定义在整个集合 $X=\{x_1,x_2,\cdots,x_m\}$ 上的犹豫模糊集 $h_A(X)$ 对 $h_B(X)$ 的相对熵为

$$R(h_A(X),h_B(X))=\\ \frac{1}{m}\sum_{i=1}^{m}\Big(\sum_{j=1}^{n}\Big[h_A^{\tau(j)}(x_i)\log\frac{h_A^{\tau(j)}(x_i)}{h_B^{\tau(j)}(x_i)}+\\ (1-h_A^{\tau(j)}(x_i))\log\frac{1-h_A^{\tau(j)}(x_i)}{1-h_B^{\tau(j)}(x_i)}\Big]\Big). \tag{6}$$

显然, $R(h_A(X),h_B(X))$ 不满足对称性. 为此, 本文构造对称型的相对熵.

定 义7 设 $X=\{x_1,x_2,\cdots,x_m\}$, $h_A(X)$ 和 $h_B(X)$ 为定义在 X 上的两个犹豫模糊集, $R(h_A(X), h_B(X))$ 和 $R(h_B(X),h_A(X))$ 分别表示 $h_A(X)$ 对 $h_B(X)$ 和 $h_B(X)$ 对 $h_A(X)$ 的相对熵, 则称

$$C(h_A(X),h_B(X))=\\ R(h_A(X),h_B(X))+R(h_B(X),h_A(X))=\\ \frac{1}{m}\sum_{i=1}^{m}\Big(\sum_{j=1}^{n}\Big[h_A^{\tau(j)}(x_i)\log\frac{h_A^{\tau(j)}(x_i)}{h_B^{\tau(j)}(x_i)}+\\ (1-h_A^{\tau(j)}(x_i))\log\frac{1-h_A^{\tau(j)}(x_i)}{1-h_B^{\tau(j)}(x_i)}\Big]\Big)+\\ \frac{1}{m}\sum_{i=1}^{m}\Big(\sum_{j=1}^{n}\Big[h_B^{\tau(j)}(x_i)\log\frac{h_B^{\tau(j)}(x_i)}{h_A^{\tau(j)}(x_i)}+\\ (1-h_B^{\tau(j)}(x_i))\log\frac{1-h_B^{\tau(j)}(x_i)}{1-h_A^{\tau(j)}(x_i)}\Big]\Big) \tag{7}$$

为犹豫模糊集 $h_A(X)$ 和 $h_B(X)$ 的对称交互熵.

对于对称交互熵 $C(h_A(X),h_B(X))$, 容易验证其满足以下性质.

性 质1 设 $X=\{x_1,x_2,\cdots,x_m\}$, $h_A(X)$ 和 $h_B(X)$ 为定义在 X 上的两个犹豫模糊集, 则其对称交互熵 $C(h_A(X),h_B(X))$ 满足如下条件:

1) $C(h_A(X),h_B(X))\geqslant 0$;

2) 当且仅当 $h_A(X)=h_B(X)$ 时, 有 $C(h_A(X), h_B(X))=0$;

3) $C(h_A(X),h_B(X))=C(h_B(X),h_A(X))$.

证明 设 $f(t)=-\log t$, 则 $f''(t)=\dfrac{1}{t^2}>0$, $f(x)$ 为凹函数. 由凹函数的性质, 存在如下关系式: $f(\lambda_1t_1+\lambda_2t_2)\leqslant\lambda_1f(t_1)+\lambda_2f(t_2), \lambda_1+\lambda_2=1, \lambda_1,\lambda_2\in(0,1)$, 当且仅当 $t_1=t_2$ 时, 等号成立. 此时

$$-\log(\lambda_1t_1+\lambda_2t_2)\leqslant-\lambda_1\log t_1-\lambda_2\log t_2,$$

将 $t_1=\dfrac{h_B^{\tau(j)}(x_i)}{h_A^{\tau(j)}(x_i)}, t_2=\dfrac{1-h_B^{\tau(j)}(x_i)}{1-h_A^{\tau(j)}(x_i)}, \lambda_1=h_A^{\tau(j)}(x_i)$, $\lambda_2=1-h_A^{\tau(j)}(x_i)$ 代入上式, 可得

$$0\leqslant-h_A^{\tau(j)}(x_i)\log\frac{h_B^{\tau(j)}(x_i)}{h_A^{\tau(j)}(x_i)}-\\ (1-h_A^{\tau(j)}(x_i))\log\frac{1-h_B^{\tau(j)}(x_i)}{1-h_A^{\tau(j)}(x_i)},\\ i=1,2,\cdots,m,\ j=1,2,\cdots,n.$$

即当且仅当 $\dfrac{h_B^{\tau(j)}(x_i)}{h_A^{\tau(j)}(x_i)}=\dfrac{1-h_B^{\tau(j)}(x_i)}{1-h_A^{\tau(j)}(x_i)}$ 时, $R(h_A(X), h_B(X))\geqslant 0$, 当 $h_A^{\tau(j)}(x_i)=h_B^{\tau(j)}(x_i)(i=1,2,\cdots,m, j=1,2,\cdots,n)$ 时, $R(h_A(X),h_B(X))=0$, 此时 $h_A(X)=h_B(X)$.

同理可得 $R(h_B(X),h_A(X))\geqslant 0$, 当且仅当 $h_A(X)=h_B(X)$ 时, 有 $R(h_B(X),h_A(X))=0$. 综上可得1)、2)成立. 由 $C(h_A(X),h_B(X))$ 的表达式, 3)显然成立. □

2.2 基于对称交互熵的犹豫模糊相似度

文献[21]基于距离给出犹豫模糊相似度公式(3), 通过距离度量犹豫模糊集之间的相似度或差异度, 距离越小, 相似度越大. 下面基于对称交互熵给出犹豫模糊环境下的相似度公式.

定 义8 设 $X=\{x_1,x_2,\cdots,x_m\}$, $h_A(X)$ 和 $h_B(X)$ 为定义在 X 上的两个犹豫模糊集, 则称

$$S(h_A(X),h_B(X))=$$

$$\frac{C(h_A(X),h_B^c(X))}{C(h_A(X),h_B(X))+C(h_A(X),h_B^c(X))} \tag{8}$$

为 $h_A(X)$ 与 $h_B(X)$ 的相似度, $h_B^c(X)$ 为 $h_B(X)$ 的补.

容易验证, $S(h_A(X),h_B(X))$ 满足如下性质:

性质2 设 $X=\{x_1,x_2,\cdots,x_m\}$, $h_A(X)$ 和 $h_B(X)$ 为定义在 X 上的两个犹豫模糊集, 则相似度公式 $S(h_A(X),h_B(X))$ 满足如下条件:

1) $0\leqslant S(h_A(X),h_B(X))\leqslant 1$;

2) $S(h_A(X),h_B(X))=S(h_B(X),h_A(X))$;

3) 当且仅当 $h_A(X)=h_B(X)$ 时, 有 $S(h_A(X),h_B(X))=1$;

4) 当且仅当 $C(h_A(X),h_B(X))=C(h_A(X),h_B^c(X))$ 时, 有 $S(h_A(X),h_B(X))=1/2$;

5) 当且仅当 $h_A(X)=h_B^c(X)$ 时, 有 $S(h_A(X),h_B(X))=0$.

事实上, 式(8)是基于TOPSIS思想提出的, $S(h_A(X),h_B(X))$ 越大, $h_A(X)$ 与 $h_B(X)$ 的相似程度越高. 与文献[21]中基于距离的相似度公式相比, 不仅可以检验 $h_A(X)$ 与 $h_B(X)$ 的相似度或差异度, 还可以检验 $h_A(X)$ 与 $h_B^c(X)$ 的差异度, 并且在某些情形下, $S(h_A(X),h_B(X))$ 所得的结果比基于距离的相似度更加合理, 也更容易分辨.

例2 设 $X=\{x_1,x_2\}$, 定义在 X 上的3个犹豫模糊集 $h_A(X),h_B(X),h_C(X)$ 分别为

$$h_A(X)=\{\langle x_1,\{0.2,0.3,0.5\}\rangle,\langle x_2,\{0.4,0.6\}\rangle\},$$

$$h_B(X)=\{\langle x_1,\{0.2,0.4,0.5\}\rangle,\langle x_2,\{0.4,0.7\}\rangle\},$$

$$h_C(X)=\{\langle x_1,\{0.3,0.4,0.6\}\rangle,\langle x_2,\{0.5,0.7\}\rangle\}.$$

x_1 属于 A,B,C 的可能的程度分别为 $h_A(x_1)=\{0.2,0.3,0.5\}$, $h_B(x_1)=\{0.2,0.4,0.5\}$, $h_C(x_1)=\{0.3,0.4,0.6\}$. 通过比较可知, $h_A(x_1)$ 与 $h_B(x_1)$ 中的隶属度更接近, 有对应的两个甚至是相同的, 相似程度也更高, 对于 x_2 也一样. 所以直观上 $h_A(X)$ 与 $h_B(X)$ 的相似度比与 $h_C(X)$ 高, 即 $S(h_A(X),h_B(X))>S(h_A(X),h_C(X))$.

若用相似度公式(3), 基于文献[21]中广义犹豫规范化豪斯道夫距离

$$d_1(h_A(X),h_B(X))=\Big[\frac{1}{m}\sum_{i=1}^{m}\max_j|h_A^{\tau(j)}(x_i)-h_B^{\tau(j)}(x_i)|^\lambda\Big]^{\frac{1}{\lambda}},\ \lambda>0, \tag{9}$$

则可得 $s_1(h_A(X),h_B(X))=s_1(h_A(X),h_C(X))$, 与直观认识不符; 若用本文方法, 则可得 $S(h_A(X),h_B(X))=0.934>S(h_A(X),h_C(X))=0.787$, 与直观认识一致; 若基于文献[21]中广义混合犹豫规范化距离, 则不失一般性, 取 $\lambda=2$.

$$d_2(h_A(X),h_B(X))=\Big[\frac{1}{2m}\sum_{i=1}^{m}\Big(\frac{1}{l_{x_i}}\sum_{j=1}^{l_{x_i}}|h_A^{\tau(j)}(x_i)-h_B^{\tau(j)}(x_i)|^\lambda+\max_j|h_A^{\tau(j)}(x_i)-h_B^{\tau(j)}(x_i)|^\lambda\Big)\Big]^{\frac{1}{\lambda}},\ \lambda>0, \tag{10}$$

则 $s_2(h_A(X),h_B(X))=0.916>s_2(h_A(X),h_C(X))=0.9$, 与本文方法结果一致. 另外, 从上述结果可以发现, 基于距离求得的相似度分辨率也不高, 而利用本文方法求得的相似度则更容易分辨. 事实上, 这两种方法在求解相似度时侧重点不同, 前者基于几何角度侧重决策信息的差异, 后者基于信息论的角度侧重决策信息的模糊程度. 当然, 在实际问题中没有哪一种方法可以适用所有情形, 因此可以将不同的方法结合起来使用, 取长补短, 快速、有效地解决问题.

实际应用中, 不同的元素 $x_i\in X$ 占有不同的地位, 应赋予不同的权重. 设元素 $x_i(i=1,2,\cdots,m)$ 对应的权重为 w_i, 满足 $\sum_{i=1}^{m}w_i=1, w_i\in[0,1]$, 依据式(8), 考虑权重的相似度公式为

$$S_w(h_A(X),h_B(X))=\frac{C_w(h_A(X),h_B^c(X))}{C_w(h_A(X),h_B(X))+C_w(h_A(X),h_B^c(X))}, \tag{11}$$

其中

$$\begin{aligned}&C_w(h_A(X),h_B(X))=\\&R_w(h_A(X),h_B(X))+R_w(h_B(X),h_A(X))=\\&\sum_{i=1}^{m}\Big(w_i\sum_{j=1}^{n}\Big[h_A^{\tau(j)}(x_i)\log\frac{h_A^{\tau(j)}(x_i)}{h_B^{\tau(j)}(x_i)}+\\&(1-h_A^{\tau(j)}(x_i))\log\frac{1-h_A^{\tau(j)}(x_i)}{1-h_B^{\tau(j)}(x_i)}\Big]\Big)+\\&\sum_{i=1}^{m}\Big(w_i\sum_{j=1}^{n}\Big[h_B^{\tau(j)}(x_i)\log\frac{h_B^{\tau(j)}(x_i)}{h_A^{\tau(j)}(x_i)}+\\&(1-h_B^{\tau(j)}(x_i))\log\frac{1-h_B^{\tau(j)}(x_i)}{1-h_A^{\tau(j)}(x_i)}\Big]\Big).\end{aligned} \tag{12}$$

当 $w_1=w_2=\cdots=w_m=1/m$ 时, 式(11)和(12)分别退化为(8)和(7). 容易验证相似度(11)满足性质2.

2.3 基于犹豫模糊相似度的聚类应用

编网聚类方法首先在模糊数学中得到了应用, 其具体过程如下[25]: 对于相似系数矩阵 P, 选择截割水平 $\lambda\in[0,1]$, 构建 λ 截矩阵 P_λ, 主对角线上填入方案符号. 在主对角线左下方用符号"*"取代"1", 将"0"去掉. 符号"*"所在的位置称为结点, 过结点作经纬线, 所谓编网就是将过结点处的经纬线对应的方案捆绑起来, 实现分类, 通过打结能互相联结的点属于一类. 编网聚类法通过对相应的截矩阵"编网"可以快速有

效地获得聚类结果, 下面给出犹豫模糊环境下的编网聚类分析方法, 具体计算过程如下:

1) 设 $Y=\{Y_1,Y_2,\cdots,Y_m\}$、$G=\{G_1,G_2,\cdots,G_n\}$、$W=(w_1,w_2,\cdots,w_n)^{\mathrm{T}}$ 分别为方案集、属性集、属性权重向量. 决策小组对所有方案按各属性进行测度, 得到犹豫模糊决策矩阵 $D=(h_{ij})_{m\times n}$, 利用式 (11) 计算犹豫模糊相似系数矩阵 $P=(S_w(h_A,h_B))_{m\times n}$.

2) 去掉相似系数矩阵 P 主对角线以上元素, 将主对角线上的元素用方案集 $Y=\{Y_1,Y_2,\cdots,Y_m\}$ 表示.

3) 选定截割水平 $\lambda\in[0,1]$, 构建 λ 截矩阵 P_λ, 在主对角线下方以符号“*”取代“1”, 将“0”去掉. 以“*”所在的位置为结点, 过结点作经纬线, 得到编网图, 则过结点处对应的经纬线上的方案归为一类.

3 算例分析

例 3 某手机大卖场欲对 7 种不同的手机 $Y_i(i=1,2,\cdots,7)$ 进行分类, 需要考虑 4 个因素: 价格 G_1, 功能 G_2, 售后服务 G_3, 质量 G_4. 决策小组给出方案 Y_i 满足属性 G_j 的值用犹豫模糊集 h_{ij} 表示, 构成决策矩阵 $D=(h_{ij})_{7\times 4}$, 如表 1 所示. $h_{11}=\{0.4,0.6\}$ 表示在价格方面, 决策小组有两种观点, 即对于方案 Y_1 满足该属性的程度分别有 0.4 和 0.6 两种, 表明决策小组意见不一致. 设属性权重为 $(0.3,0.25,0.2,0.25)^{\mathrm{T}}$.

表 1　犹豫模糊决策矩阵

	G_1	G_2	G_3	G_4
Y_1	{0.4,0.6}	{0.3,0.5,0.6}	{0.2,0.3}	{0.4}
Y_2	{0.2,0.3,0.5}	{0.2,0.4}	{0.2,0.4,0.5}	{0.3,0.4}
Y_3	{0.2,0.3,0.4,}	{0.2,0.4}	{0.2,0.3,0.5}	{0.3}
Y_4	{0.3,0.45,0.75}	{0.3,0.6}	{0.3,0.6,0.75}	{0.45,0.6}
Y_5	{0.8}	{0.4,0.6,0.7}	{0.3,0.4}	{0.6,0.8}
Y_6	{0.4,0.5,0.7}	{0.4,0.6}	{0.4,0.6,0.7}	{0.5,0.6}
Y_7	{0.6,0.7,0.9}	{0.6,0.8}	{0.6,0.8,0.9}	{0.7,0.8}

1) 利用式 (11) 计算犹豫模糊相似系数矩阵

$$P=\begin{bmatrix} 1 & 0.7141 & 0.7329 & 0.4221 & 0.5287 & 0.3038 & 0.2465 \\ 0.7141 & 1 & 0.9777 & 0.5019 & 0.2432 & 0.3226 & 0.0630 \\ 0.7329 & 0.9777 & 1 & 0.4716 & 0.2179 & 0.3243 & 0.0327 \\ 0.4221 & 0.5019 & 0.4716 & 1 & 0.5064 & 0.7121 & 0.5511 \\ 0.5287 & 0.2432 & 0.2179 & 0.5064 & 1 & 0.5657 & 0.7464 \\ 0.3038 & 0.3226 & 0.3243 & 0.7121 & 0.5657 & 1 & 0.6583 \\ 0.2465 & 0.0630 & 0.0327 & 0.5511 & 0.7464 & 0.6583 & 1 \end{bmatrix}.$$

2) 去掉相似系数矩阵 P 主对角线以上的元素, 将主对角线上的元素用方案集 $Y=\{Y_1,Y_2,\cdots,Y_7\}$ 表示, 即

$$P=\begin{bmatrix} Y_1 & & & & & & \\ 0.7141 & Y_2 & & & & & \\ 0.7329 & 0.9777 & Y_3 & & & & \\ 0.4221 & 0.5019 & 0.4716 & Y_4 & & & \\ 0.5287 & 0.2432 & 0.2179 & 0.5064 & Y_5 & & \\ 0.3038 & 0.3226 & 0.3243 & 0.7121 & 0.5657 & Y_6 & \\ 0.2465 & 0.0630 & 0.0327 & 0.5511 & 0.7464 & 0.6583 & Y_7 \end{bmatrix}.$$

3) 选定截割水平 $\lambda\in[0,1]$, 构建 λ 截矩阵 P_λ, 作编网图, 则:

① 当 $0.9777<\lambda\leqslant 1$ 时, 分为 7 类: $\{Y_1\},\{Y_2\},\{Y_3\},\{Y_4\},\{Y_5\},\{Y_6\},\{Y_7\}$;

② 当 $0.7464<\lambda\leqslant 0.9777$ 时, 分为 6 类: $\{Y_1\},\{Y_2,Y_3\},\{Y_4\},\{Y_5\},\{Y_6\},\{Y_7\}$;

③ 当 $0.7329<\lambda\leqslant 0.7464$ 时, 分为 5 类: $\{Y_1\},\{Y_2,Y_3\},\{Y_4\},\{Y_5,Y_7\},\{Y_6\}$;

④ 当 $0.7121<\lambda\leqslant 0.7329$ 时, 分为 4 类: $\{Y_1,Y_2,Y_3\},\{Y_4\},\{Y_5,Y_7\},\{Y_6\}$;

⑤ 当 $0.6583<\lambda\leqslant 0.7121$ 时, 分为 3 类: $\{Y_1,Y_2,Y_3\},\{Y_4,Y_6\},\{Y_5,Y_7\}$;

⑥ 当 $0.5287<\lambda\leqslant 0.6583$ 时, 分为 2 类: $\{Y_1,Y_2,Y_3\},\{Y_4,Y_5,Y_6,Y_7\}$;

⑦ 当 $0\leqslant\lambda\leqslant 0.5287$ 时, 归为 1 类: $\{Y_1,Y_2,Y_3,Y_4,Y_5,Y_6,Y_7\}$.

如果利用文献 [10] 中基于等价关系的聚类方法, 则基于犹豫模糊集的关联系数公式[10]为

$$\rho_{AB}= \tag{13}$$

$$\sum_{i=1}^{n}w_i\Big(\frac{1}{l_i}\sum_{j=1}^{l_i}h_A^{\tau(j)}(x_i)h_B^{\tau(j)}(x_i)\Big)\Big/ \Big\{\Big[\sum_{i=1}^{n}w_i\Big(\frac{1}{l_i}\sum_{j=1}^{l_i}(h_A^{\tau(j)}(x_i))^2\Big)\Big]^{1/2}\times \Big[\sum_{i=1}^{n}w_i\Big(\frac{1}{l_i}\sum_{j=1}^{l_i}(h_B^{\tau(j)}(x_i))^2\Big)\Big]^{1/2}\Big\}. \tag{14}$$

计算关联系数矩阵 C, 得

$$C=\begin{bmatrix} 1 & 0.9513 & 0.9577 & 0.9513 & 0.9787 & 0.9636 & 0.9657 \\ 0.9513 & 1 & 0.9920 & 1 & 0.9370 & 0.9938 & 0.9865 \\ 0.9577 & 0.9920 & 1 & 0.9920 & 0.9340 & 0.9917 & 0.9872 \\ 0.9513 & 1 & 0.9920 & 1 & 0.9370 & 0.9938 & 0.9865 \\ 0.9787 & 0.9370 & 0.9340 & 0.9370 & 1 & 0.9554 & 0.9606 \\ 0.9636 & 0.9938 & 0.9917 & 0.9938 & 0.9554 & 1 & 0.9986 \\ 0.9657 & 0.9865 & 0.9872 & 0.9865 & 0.9606 & 0.9986 & 1 \end{bmatrix}.$$

构造等价关联矩阵, 计算 $C^2, C^4, \cdots, C^{2n}, \cdots$, 此时 $C^8 = C^4$, C^4 为等价关联矩阵, 即

$$C^4=\begin{bmatrix} 1 & 0.9657 & 0.9657 & 0.9657 & 0.9787 & 0.9657 & 0.9657 \\ 0.9657 & 1 & 0.9920 & 1 & 0.9657 & 0.9938 & 0.9938 \\ 0.9657 & 0.9920 & 1 & 0.9920 & 0.9657 & 0.9920 & 0.9920 \\ 0.9657 & 1 & 0.9920 & 1 & 0.9657 & 0.9938 & 0.9938 \\ 0.9787 & 0.9657 & 0.9657 & 0.9657 & 1 & 0.9657 & 0.9657 \\ 0.9657 & 0.9938 & 0.9920 & 0.9938 & 0.9657 & 1 & 0.9986 \\ 0.9657 & 0.9938 & 0.9920 & 0.9938 & 0.9657 & 0.9986 & 1 \end{bmatrix}.$$

选择不同的截割水平 $\lambda \in [0,1]$, 构造 λ 截矩阵, 利用文献 [10]、文献 [22] 以及文献 [24] 中的聚类方法得到的聚类结果如表 2 所示.

表 2　聚类结果的比较

分类	文献 [10] 的方法	文献 [22] 的方法	文献 [24] 的方法
1	$\{Y_1,Y_2,Y_3,Y_4,Y_5,Y_6,Y_7\}$	$\{Y_1,Y_2,Y_3,Y_4,Y_5,Y_6,Y_7\}$	$\{Y_1,Y_2,Y_3,Y_4,Y_5,Y_6,Y_7\}$
2		$\{Y_1,Y_2,Y_3,Y_4,Y_5,Y_6\},\{Y_7\}$	$\{Y_1,Y_2,Y_3,Y_4,Y_6,Y_7\},\{Y_5\}$
3	$\{Y_1,Y_5\},\{Y_2,Y_3,Y_4\},\{Y_6,Y_7\}$	$\{Y_1,Y_2,Y_3,Y_4,Y_6\},\{Y_5\},\{Y_7\}$	$\{Y_1,Y_2,Y_3,Y_4,Y_6\},\{Y_5\},\{Y_7\}$
4	$\{Y_1\},\{Y_2,Y_3,Y_4\},\{Y_5\},\{Y_6,Y_7\}$	$\{Y_1,Y_4,Y_6\},\{Y_2,Y_3\},\{Y_5\},\{Y_7\}$	$\{Y_1,Y_2,Y_3\},\{Y_4,Y_6\},\{Y_5\},\{Y_7\}$
5	$\{Y_1\},\{Y_2,Y_4\},\{Y_3\},\{Y_5\},\{Y_6,Y_7\}$	$\{Y_1\},\{Y_2,Y_3\},\{Y_4,Y_6\},\{Y_5\},\{Y_7\}$	$\{Y_1\},\{Y_2,Y_3\},\{Y_4,Y_6\},\{Y_5\},\{Y_7\}$
6	$\{Y_1\},\{Y_2,Y_4\},\{Y_3\},\{Y_5\},\{Y_6\},\{Y_7\}$	$\{Y_1\},\{Y_2,Y_3\},\{Y_4\},\{Y_5\},\{Y_6\},\{Y_7\}$	$\{Y_1\},\{Y_2,Y_3\},\{Y_4\},\{Y_5\},\{Y_6\},\{Y_7\}$
7		$\{Y_1\},\{Y_2\},\{Y_3\},\{Y_4\},\{Y_5\},\{Y_6\},\{Y_7\}$	$\{Y_1\},\{Y_2\},\{Y_3\},\{Y_4\},\{Y_5\},\{Y_6\},\{Y_7\}$

通过比较可以发现, 利用文献 [10] 中的方法和本文方法所得到的结果主要存在以下差别: 1) 用文献 [10] 的方法最多可将样本分为 6 类, 而利用本文方法所得的结果更加精细, 可以满足不同需求. 另外, 文献 [10] 中基于等价关系的划分方法, 为了获得等价关联矩阵, 首先需要对关联系数矩阵不断迭代, 直到 $C^{2^n} = C^{2^{n+1}}$, 这样容易造成信息丢失, 并且计算量大[22]. 2) 在聚类结果上存在显著差异, 如在聚为 6 类时, 利用本文的方法是 $\{Y_1\},\{Y_2,Y_3\},\{Y_4\},\{Y_5\},\{Y_6\},\{Y_7\}$, Y_2、Y_3 归为一类, 而用文献 [10] 中的方法是 $\{Y_1\}$, $\{Y_2,Y_4\},\{Y_3\},\{Y_5\},\{Y_6\},\{Y_7\}$, Y_2、Y_4 归为一类. 事实上, 若将方案 Y_2、Y_3、Y_4 用犹豫模糊集表示, 则

$$\begin{aligned} Y_2 = \{&\langle G_1,\{0.2,0.3,0.5\}\rangle,\langle G_2,\{0.2,0.4\}\rangle,\\ &\langle G_3,\{0.2,0.4,0.5\}\rangle,\langle G_4,\{0.3,0.4\}\rangle\},\\ Y_3 = \{&\langle G_1,\{0.2,0.3,0.4\}\rangle,\langle G_2,\{0.2,0.4\}\rangle,\\ &\langle G_3,\{0.2,0.3,0.5\}\rangle,\langle G_4,\{0.3\}\rangle\},\\ Y_4 = \{&\langle G_1,\{0.3,0.45,0.75\}\rangle,\langle G_2,\{0.3,0.6\}\rangle,\\ &\langle G_3,\{0.3,0.6,0.75\}\rangle,\langle G_4,\{0.45,0.6\}\rangle\}. \end{aligned}$$

同 Y_4 相比, Y_2 和 Y_3 对应的各属性值更加接近, 有的甚至相等 (如在 G_2 下), 相似程度也更高, 因此将 Y_2 和 Y_3 归为一类更加合理 (这也与文献 [22] 和文献 [24] 的方法所获得的结果一致). 造成这些差异的原因除了聚类的方法不同之外, 更重要的是在于样本相似性度量方法的选择上, 文献 [10] 的关联系数公式 (13) 事实上是基于向量的观点, 利用向量的夹角余弦公式度量样本相似性, 只考虑了向量的方向, 忽视了向量本身的模长, 这样必然会出现即使两个样本不同时其相关系数也为 1 的情况, 如 $\rho_{Y_2Y_4} = 1$ (但 $Y_2 \neq Y_4$), 使得 Y_2、Y_4 始终归为一类. 另外, 文献 [22] 和文献 [24] 的方法所得的聚类结果与本文结果也有所不同: 文献 [22] 利用最小距离作为聚类依据, 一旦一组样本被合并时, 需要利用犹豫模糊平均算子[11]在新生成的类上重新计算聚类中心, 并且随着类中样本数的增加, 计算量也越来越大; 文献 [24] 利用最小生成树 (MST) 聚类算法得到聚类结果, 首先基于文献 [21] 中的距离公式得到犹豫模糊距离及模糊图, 但其中的距离公式分辨率不够高, 且结果有时与事实不符, 这必然会影响到聚类的结果. 本文利用对称交互熵获得犹豫模糊信息相似度, 通过编网聚类方法, 在相似系数矩阵上直接进行表上作业, 可以更加快速有效地获得聚类结果, 相比其他方法, 计算简单, 易于操作.

4　结　论

本文对属性值是犹豫模糊信息形式的聚类问题进行了研究. 定义了犹豫模糊相对熵、对称交互熵; 针

对基于距离的相似度公式的缺陷, 结合TOPSIS等方法的思想, 利用对称交互熵定义了一种新的犹豫模糊相似度公式; 根据编网聚类法原理, 对犹豫模糊集进行聚类分析, 从而获得了不同的聚类结果.

参考文献(References)

[1] Torra V, Narukawa Y. On hesitant fuzzy sets and decision[C]. The 18th IEEE Int Conf on Fuzzy Systems. Jeju Island, 2009: 1378-1382.

[2] Torra V. Hesitant fuzzy sets[J]. Int J of Intelligent Systems, 2010, 25(6): 529-539.

[3] Atanassov K. Intuitionistic fuzzy sets[J]. Fuzzy Sets and Systems, 1986, 20(1): 87-96

[4] Dubois D, Prade H. Fuzzy sets and systems: Theory and applications[M]. New York: Academic Press, 1980: 30-32.

[5] Yager R R. On the theory of bags[J]. Int J of General Systems, 1986, 13(1): 23-37.

[6] Zadeh L A. Fuzzy sets[J]. Information and Control, 1965, 8(3): 338-353.

[7] Zhang N, Wei G W. Extension of VIKOR method for decision making problem based on hesitant fuzzy set[J]. Applied Mathematical Modelling, 2013, 37(7): 4938-4947.

[8] Xu Z S, Xia M M. Hesitant fuzzy entropy and cross-entropy and their use in multiattribute decision-making[J] Int J of Intelligent Systems, 2012, 27(9): 799-822.

[9] Xu Z S, Xia M M. On distance and correlation measures of hesitant fuzzy information[J]. Int J of Intelligent Systems, 2011, 26(5): 410-425.

[10] 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.

[11] Xia M M, Xu Z S. Hesitant fuzzy information aggregation in decision making[J]. Int J of Approximate Reasoning, 2011, 52(3): 395-407.

[12] Xia M M, Xu Z S, Chen N. Some hesitant fuzzy aggregation operators with their application in group decision making[J]. Group Decision and Negotiation, 2013, 22(2): 259-279.

[13] Wei G W. Hesitant fuzzy prioritized operators and their application to multiple attribute decision making[J]. Knowledge-Based Systems, 2012, 31(1): 176-182.

[14] Xia M M, Xu Z S, Chen N. Induced aggregation under confidence levels[J]. Int J of Uncertainty, Fuzziness & Knowledge-Based Systems, 2011, 19(2): 201-227.

[15] 夏梅梅. 模糊决策信息集成方式及测度研究[D]. 南京: 东南大学经济管理学院, 2012: 66-67.
(Xia M M. Research on fuzzy decision information aggregation techniques and measures[D]. Nanjing: School of Economics & Management, Southeast University, 2012: 66-67.)

[16] Zhu B, Xu Z S, Xia M M. Hesitant fuzzy geometric Bonferroni means[J]. Information Sciences, 2012, 205(1): 72-85.

[17] Yu D J, Wu Y Y, Zhou W. Generalized hesitant fuzzy Bonferroni mean and its application in multi-criteria group decision making[J]. J of Information & Computational Science, 2012, 9(2): 267-274.

[18] Xu Z S, Yager R R. Intuitionistic and interval-valued intuitionistic fuzzy preference relations and their measures of similarity for the evaluation of agreement within a group[J]. Fuzzy Optimization and Decision Making, 2009, 8(2): 123-139.

[19] Szmidt E, Kacprzyk J. A similarity measure for intuitionistic fuzzy sets and its application in supporting medical diagnostic reasoning[C]. The 7th Int Conf on Artificial Intelligence and Soft Computing. Zakopane, 2004: 388-393.

[20] Li D F, Cheng C T. New similarity measures of intuitionistic fuzzy sets and application to pattern recognitions[J]. Pattern Recognition Letters, 2002, 23(1): 221-225.

[21] Xu Z S, Xia M M. Distance and similarity measures for hesitant fuzzy sets[J]. Information Sciences, 2011, 181(11): 2218-2138.

[22] Zhang X L, Xu Z S. Hesitant fuzzy agglomerative hierarchical clustering algorithms[J]. Int J of Systems Science, 2013, DOI: 10.1080/00207721.2013.797037.

[23] Miyamoto, S. Fuzzy sets in information retrieval and cluster analysis[M]. Dordrecht: Kluwer, 1990: 181-192.

[24] Zhang X L, Xu Z S. A MST clustering analysis method under hesitant fuzzy environment[J]. Control and Cybernetics, 2012, 41(3): 645-666.

[25] 贺仲雄. 模糊数学及其应用[M]. 天津: 天津科学技术出版社, 1983: 165-169.
(He Z X. Fuzzy mathematics and its application[M]. Tianjin: Tianjin Science and Technology Press, 1983: 165-169.)

[26] Cover T M, Thomas Joy A. Elements of information theory[M]. New York: John Wiley and Sons, 2006: 19-20.

[27] 赵萌, 邱菀华, 刘北上. 基于相对熵的多属性决策排序方法[J]. 控制与决策, 2010, 25(7): 1098-1100.
(Zhao M, Qiu W H, Liu B S. Relative entropy evaluation method for multiple attribute decision making[J]. Control and Decision, 2010, 25(7): 1098-1100.)

[28] Chen H Y, Zhou L G. A relative entropy approach to group decision making with interval reciprocal relations based on COWA operator[J]. Group Decision and Negotiation, 2012, 21(4): 585-599.

[29] Lai J, Ford J. Relative entropy rate based multiple hidden markov model approximation[J]. IEEE Trans on Signal Process, 2010, 58(1): 165-174.

(责任编辑: 齐 霁)