控制与决策
Control and Decision

基于指数熵加权的降维犹豫模糊兰氏距离测度及应用

沙秀艳，尹传存，徐泽水

引用本文:
沙秀艳, 尹传存, 徐泽水. 基于指数熵加权的降维犹豫模糊兰氏距离测度及应用[J]. 控制与决策, 2020, 35(3): 728-734.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2018.0910

文章编号: 1001-0920(2020)03-0728-07 DOI: 10.13195/j.kzyjc.2018.0910

基于指数熵加权的降维犹豫模糊兰氏距离测度及应用

沙秀艳[1], 尹传存[1†], 徐泽水[2]

(1. 曲阜师范大学 统计学院，山东 曲阜 273165；2. 四川大学 商学院，成都 610064)

摘 要: 首先提出几种基于兰氏距离的犹豫模糊集距离测度. 然后针对两个犹豫模糊数中的隶属度个数不相等问题,提出新的犹豫模糊数降维方案. 该方案不需要反复添加最大最小隶属度数值到犹豫模糊数中,不仅很好地保留了数据的原始信息,而且减少了计算距离时的计算量. 针对属性权重信息完全未知的情况,采用实际数据信息构造犹豫模糊指数熵,并利用信息熵最小化原则计算得到属性权重. 最后利用指数熵加权的降维犹豫模糊兰氏距离测度,结合实际的医疗诊断数据进行实例分析. 结果表明,所提出的基于指数熵加权的降维犹豫模糊兰氏距离测度不仅在 λ 取不同值时诊断结果一致,而且减少了计算量,提高了诊断效率,对实时、有效的医疗诊断具有一定的应用价值.

关键词: 犹豫模糊集；兰氏距离；降维；指数熵

中图分类号: O21 **文献标志码:** A

Weighted hesitant fuzzy Lance distance measure of dimension reduction based on exponential entropy and its application

SHA Xiu-yan[1], *YIN Chuan-cun*[1†], *XU Ze-shui*[2]

(1. School of Statistics，Qufu Normal University，Qufu 273165，China；2. Business School，Sichuan University，Chengdu 610064，China)

Abstract: This paper proposes several distance measures about hesitant fuzzy sets based on the Lance distance at first. Then, aiming at the unequal number of membership numbers in two hesitant fuzzy elements, a new dimension reduction scheme of hesitant fuzzy elements is introduced. The scheme does not need to add the maximum or minimum membership value to the hesitant fuzzy element, which not only preserves the original information of the data well, but also reduces the computation of the calculated distance. In the case that the information of attribute weights are completely unknown, the hesitant fuzzy exponential entropy is constructed by using the actual data information. This paper uses the principle of entropy minimization to calculate attribute weights. Finally, a medical diagnosis example is provided to illustrate the weighted hesitant fuzzy Lance distance measure of dimension reduction based on the exponential entropy. The results show that, the proposed distance measure not only achieves the consistent diagnosis results when λ takes different values, but also reduces the computation and improves the diagnostic efficiency, which has certain application value for real-time and effective medical diagnosis.

Keywords: hesitant fuzzy set；Lance distance；dimension reduction；exponential entropy

0 引 言

1965 年,著名控制论专家 Zadeh 发表了开创性的论文《模糊集合》,提出了模糊集的概念[1],提供了一套严格的数学方法来描述带有模糊不确定性的现象和事物. 经过几十年的发展,模糊集理论已经成功地应用于数据挖掘、决策分析、聚类分析、图像处理、模式识别和人工智能等领域. 但是人们在对事物进行决策时,常常在多个决策信息之间犹豫,而且不同的决策者之间难以达成一致的最终决策结果. 2010 年, Torra[2] 提出了犹豫模糊集的概念. 当决策者在确定隶属度值时,若在多个数值之间难以取舍、犹豫不定,则可以选定多个数值作为最终的隶属度值.

此后,国内外学者对犹豫模糊集及其扩充形式的距离测度、相似度、熵测度、相关系数等问题进行了深入的研究. Xu 等[3] 首次提出了犹豫模糊集的几种距离测度,并得到了相应的相似度; Zhu 等[4] 研究了犹豫模糊几何 Bonferroni 平均算子; 陈树伟等[5] 提出了区间值犹豫模糊集的定义,给出了区间值犹豫模糊

收稿日期: 2018-07-04；**修回日期:** 2018-09-27.

基金项目: 国家自然科学基金项目 (11171179, 11571198, 61273209).

责任编委: 王光臣.

†通讯作者. E-mail: ccyin@qfnu.edu.cn.

元素的一些基本运算和性质;吴婉莹等[6]提出了直觉对偶犹豫模糊集的概念和基本运算,构造了直觉对偶犹豫模糊集的算术加权平均集结算子和几何加权平均算子,并考虑有序和广义的情况分别给出相对应的公式;Li等[7]研究了一种基于犹豫度的犹豫模糊集合的距离和贴近度;谢海[8]给出了犹豫模糊集截集的概念,并研究截集的性质,建立了沟通犹豫模糊集合与经典集合之间的桥梁;关欣等[9]在犹豫模糊框架内,提出犹豫模糊集相关系数计算方法,并将其应用到多源异类数据的融合识别中. 在这些研究热点中,当两个犹豫模糊数中隶属度个数不相等时,如何进行距离度量是一个较为关键的问题. Xu等[3]提出了主观添加一些隶属度数值使得两个犹豫模糊数长度相同的方法. 这种方法操作简单,但是反复添加相同的隶属度会使得原始数据信息失真,最终可能导致错误的决策. 为了克服这个缺陷, Hu等[10]提出了一种新的犹豫模糊距离测度,该方法不需要反复添加隶属度,但是计算量较大;林松等[11]利用犹豫模糊数中元素个数和偏差构造一种新的犹豫度,并提出一种基于符号距离的犹豫模糊多属性决策方法;同年,彭定洪等[12]提出了一种犹豫模糊排序加权的Hausdorff距离,并将其应用到完全犹豫模糊环境下的Topsis方法中.

犹豫模糊集允许多个隶属度数值的存在,能够较好地体现决策者的犹豫心理,因此已经广泛地应用于决策分析、图像处理、模式识别等领域. 本文首先将传统的兰氏距离推广到犹豫模糊领域,提出几种基于兰氏距离的犹豫模糊集距离测度. 然后为了较好地保留数据的原始信息,提出新的犹豫模糊数降维方案. 针对属性权重需要人为给出,或者采用均等权重的问题,本文采用实际数据信息构造犹豫模糊指数熵,并利用信息熵最小化原则计算得到属性权重. 最后,通过一个医疗诊断实例来说明本文所提出方法的有效性和实时性.

1　几种经典的模糊集距离

设A和B为$X=\{x_1,x_2,\ldots,x_n\}$上的两个模糊集,则几种经典的模糊集距离表示如下:

1) 规范化的海明距离

$$d_{\mathrm{nh}}(A,B)=\frac{1}{n}\sum_{i=1}^{n}|\mu_A(x_i)-\mu_B(x_i)|;$$

2) 豪斯道夫距离

$$d_{\mathrm{hd}}(A,B)=\max|\mu_A(x_i)-\mu_B(x_i)|;$$

3) 规范化的欧几里德距离

$$d_{\mathrm{ne}}(A,B)=\frac{1}{n}\Big(\sum_{i=1}^{n}|\mu_A(x_i)-\mu_B(x_i)|^2\Big)^{\frac{1}{2}};$$

4) 规范化的兰氏距离

$$d_{nl}(A,B)=\frac{1}{n}\sum_{i=1}^{n}\frac{|\mu_A(x_i)-\mu_B(x_i)|}{\mu_A(x_i)+\mu_B(x_i)}.$$

以上各式中的$\mu_A(x_i)$和$\mu_B(x_i)$分别表示A和B的隶属度函数,并且满足$0\leqslant\mu_A(x_i),\mu_B(x_i)\leqslant 1, x_i\in X, i=1,2,\ldots,n$.

2　犹豫模糊集概述

2.1　犹豫模糊集的定义

设论域$X=\{x_1,x_2,\ldots,x_n\}$,则X上的犹豫模糊集(hesitant fuzzy set, HFS)[2,13-14]定义为

$$M=\{\langle x,h_M(x)\rangle|x\in X\},$$

其中$h_M(x)$为犹豫模糊数,是[0,1]中一些数值的集合,表示集合X中的任一元素x对集合M的隶属度. 这里,如果每个犹豫模糊数中的隶属度有且仅有一个,则犹豫模糊数退化为普通的模糊集.

2.2　犹豫模糊数的元素降维方案

设M和N是两个定义在$X=\{x_1,x_2,\ldots,x_n\}$上的犹豫模糊集,任意选取两个犹豫模糊数$h_M(x_i)$和$h_N(x_i)$,一般情况下$l_{h_{M(x_i)}}\neq l_{h_{N(x_i)}}$,其中$l_{h_{M(x_i)}}$和$l_{h_{N(x_i)}}$分别表示$h_M(x_i)$和$h_N(x_i)$中隶属度的个数. 由于犹豫模糊数中的隶属度个数不相等,导致无法进行距离度量. 为了方便计算,目前较多的研究是将长度短的犹豫模糊数补齐至与长度长的犹豫模糊数长度一样,具体选取的数值取决于决策者的风险程度. 喜好风险的决策者采用乐观补齐方案,即选取犹豫模糊数中最大的隶属度数值进行补齐;讨厌风险的决策者采用悲观补齐方案,即选取犹豫模糊数中最小隶属度数值进行补齐[3].

例如,犹豫模糊数$h_M(x_i)=\{0.3,0.4,0.6,0.8\}$和$h_N(x_i)=\{0.6,0.8\}$,由于犹豫模糊数中的隶属度个数不相等,则需对长度短的犹豫模糊数$h_N(x_i)$进行补齐至与长度长的犹豫模糊数$h_M(x_i)$相等. 悲观补齐后$h'_N(x_i)=\{0.6,0.6,0.6,0.8\}$,乐观补齐后$h'_N(x_i)=\{0.6,0.8,0.8,0.8\}$. 这种通过反复添加最大最小隶属度数值的方法使得$h_N(x_i)$中的原始信息发生了很大的改变. 采用这种偏离原始信息的数据进行进一步的研究,在一定程度上可能会产生完全不同于决策者初始意愿的结果. 因此,该方法有待于进一步改进.

针对两个犹豫模糊数中的隶属度个数不相等问题,为了能够很好地保留数据的原始信息,并减少计算量,提出一种新的犹豫模糊数降维方案. 这里需对长度长的犹豫模糊数$h_M(x_i)$中的隶属度个数降维至数量与长度短的犹豫模糊数$h_N(x_i)$相等. 悲观

降维方案是保留$h_M(x_i)$中的最小隶属度0.3, $h_M(x_i)$中的另一个隶属度是其余3个隶属度数值的平均,即$(0.4+0.6+0.8)/3 = 0.6$,则悲观降维方案处理后$h'_M(x_i) = \{0.3, 0.6\}$. 乐观降维方案是保留$h_M(x_i)$中的最大隶属度0.8, $h_M(x_i)$中的另一个隶属度是其余3个隶属度数值的平均,即$(0.3+0.4+0.6)/3 = 0.4333$,则乐观降维方案处理后$h'_M(x_i) = \{0.4333, 0.8\}$. 因为该降维方案是采用已有原始数据的平均,所以这种方法处理后的$h'_M(x_i)$既较大程度地利用了数据的原始信息,又减少了计算距离时的计算量. 本文在接下来的讨论中,均设决策者为喜好风险的,即采用乐观降维方案.

2.3 犹豫模糊集的距离度量

定义1 对于任意集合$X = \{x_1, x_2, \ldots, x_n\}$上的犹豫模糊集$M$和$N$, M和N的距离$d(M,N)$需满足下面的条件:

1) $0 \leqslant d(M,N) \leqslant 1$;

2) $d(M,N) = 0$,当且仅当$M = N$;

3) $d(M,N) = d(N,M)$.

3 基于兰氏距离的犹豫模糊集距离公式

兰氏距离是由Lance和Williams最早提出[15],因其与各变量的单位无关,且是通过比值的形式来度量数据之间的距离,而由于比值受极端值的影响较小,因而它对偏倚数据敏感性较低. 对于存在属性值偏倚较大的数据,兰氏距离公式计算得到的距离优于其他距离,所以它是数据分析中一种度量距离的常用方法. 本文将提出几种基于兰氏距离的犹豫模糊集距离公式,并将其应用于医疗诊断领域.

3.1 几种基于兰氏距离的犹豫模糊集距离定义

1) 标准化的犹豫模糊兰氏距离

$$d_{\mathrm{nhl}}(M,N) = \frac{1}{n}\sum_{i=1}^{n}\Big[\frac{1}{l_{x_i}}\sum_{j=1}^{l_{x_i}}\frac{|h_M^{\sigma(j)}(x_i) - h_N^{\sigma(j)}(x_i)|}{h_M^{\sigma(j)}(x_i) + h_N^{\sigma(j)}(x_i)}\Big], \tag{1}$$

这里$h_M^{\sigma(j)}(x_i)$表示犹豫模糊数$h_M(x_i)$中第j大的元素.

2) 广义犹豫模糊兰氏距离

$$d_{\mathrm{ghl}}(M,N) = \Big\{\frac{1}{n}\sum_{i=1}^{n}\Big[\frac{1}{l_{x_i}}\sum_{j=1}^{l_{x_i}}\Big(\frac{|h_M^{\sigma(j)}(x_i) - h_N^{\sigma(j)}(x_i)|}{h_M^{\sigma(j)}(x_i) + h_N^{\sigma(j)}(x_i)}\Big)^{\lambda}\Big]\Big\}^{\frac{1}{\lambda}}, \tag{2}$$

这里$\lambda > 0$. 若$\lambda = 1$,则广义犹豫模糊兰氏距离退化为标准化的犹豫模糊兰氏距离.

实际应用中常遇到这样的情形,某中学要对两名学生甲、乙进行考察,学校邀请10位专家根据两名学生的综合能力进行评估. 设10位专家给甲打分时,7位给出0.9, 3位给出0.4;专家给乙打分时,8位给出0.4, 2位给出0.9. 学校领导很可能在0.4和0.9这两个数值中犹豫,这时采用将这2个数值都保留,用犹豫模糊数$\{0.4, 0.9\}$来描述评价结果较为合理. 但是按照常规的犹豫模糊数的处理方法,甲的评价结果为$\{0.4, 0.9\}$,乙的评价结果也为$\{0.4, 0.9\}$,甲和乙的综合评价结果就完全相同. 显然,这个结论与实际情况不相符合. 因此,为了更准确地反映决策者的犹豫心理,引入加权犹豫模糊集的概念,将犹豫模糊集中每个隶属度值都赋予了一定的权重,用以强调每个值被选为隶属度值的可能性大小. 由此,给出下面的加权犹豫模糊兰氏距离公式.

假设任意$x_i \in X$的权重为w_i,其中$i = 1, 2, \ldots, n$,并满足$w_i \in [0,1]$, $\sum_{i=1}^{n} w_i = 1$,则有如下的定义.

3) 加权犹豫模糊兰氏距离

$$d_{\mathrm{whl}}(M,N) = \sum_{i=1}^{n} w_i\Big[\frac{1}{l_{x_i}}\sum_{j=1}^{l_{x_i}}\frac{|h_M^{\sigma(j)}(x_i) - h_N^{\sigma(j)}(x_i)|}{h_M^{\sigma(j)}(x_i) + h_N^{\sigma(j)}(x_i)}\Big]. \tag{3}$$

4) 广义加权犹豫模糊兰氏距离

$$d_{\mathrm{gwhl}}(M,N) = \Big\{\sum_{i=1}^{n} w_i\Big[\frac{1}{l_{x_i}}\sum_{j=1}^{l_{x_i}}\Big(\frac{|h_M^{\sigma(j)}(x_i) - h_N^{\sigma(j)}(x_i)|}{h_M^{\sigma(j)}(x_i) + h_N^{\sigma(j)}(x_i)}\Big)^{\lambda}\Big]\Big\}^{\frac{1}{\lambda}}, \tag{4}$$

这里$\lambda > 0$. 若$\lambda = 1$,则广义加权犹豫模糊兰氏距离退化为加权犹豫模糊兰氏距离.

上面的距离公式可以推广到连续情况下,设对于任意的$x \in X = [a,b]$,其权重为$w(x)$,并满足$w(x) \in [0,1]$, $\int_a^b w(x)\mathrm{d}x = 1$.

1) 连续的加权犹豫模糊兰氏距离

$$d_{\mathrm{cwhl}} = \int_a^b w(x)\Big[\frac{1}{l_x}\sum_{j=1}^{l_x}\frac{|h_M^{\sigma(j)}(x) - h_N^{\sigma(j)}(x)|}{h_M^{\sigma(j)}(x) + h_N^{\sigma(j)}(x)}\Big]\mathrm{d}x. \tag{5}$$

2) 广义连续的加权犹豫模糊兰氏距离

$$d_{\mathrm{gcwhl}} = \Big\{\int_a^b w(x)\Big[\frac{1}{l_x}\sum_{j=1}^{l_x}\Big(\frac{|h_M^{\sigma(j)}(x) - h_N^{\sigma(j)}(x)|}{h_M^{\sigma(j)}(x) + h_N^{\sigma(j)}(x)}\Big)^{\lambda}\Big]\mathrm{d}x\Big\}^{\frac{1}{\lambda}}, \tag{6}$$

这里$\lambda > 0$.

3.2 属性权重的计算方法

在属性权重信息完全未知的情况下,属性权重需要主观地人为给出,或者采用均等权重的方法计算,即各属性均采用常权,这两种处理方式有时会导致不科学的决策结果. 这里利用实际数据信息构造犹豫模糊指数熵,然后利用信息熵最小化原则计算得到属性权重[16].

设犹豫模糊数为

$$h_M(x_i)=\{h_M^{\sigma(1)}(x_i),h_M^{\sigma(2)}(x_i),\ldots,h_M^{\sigma(l_{x_i})}(x_i)\},$$

这里l_{x_i}为$h_M(x_i)$中隶属度的个数,则$h_M(x_i)$的犹豫模糊指数熵定义为

$$E(h_M(x_i))=\frac{1}{l_{x_i}(\sqrt{\mathrm{e}}-1)}\sum_{j=1}^{l_{x_i}}[h_M^{\sigma(j)}(x_i)\mathrm{e}^{1-h_M^{\sigma(j)}(x_i)}+(1-h_M^{\sigma(j)}(x_i))\mathrm{e}^{h_M^{\sigma(j)}(x_i)}-1], \tag{7}$$

这里$j=1,2,\ldots,l_{x_i}$.

对于某一多属性决策问题,假设$A=\{A_1,A_2,\ldots,A_m\}$为一组方案集,$G=\{G_1,G_2,\ldots,G_n\}$为方案的属性集合,属性的权重为w_q,其中$q=1,2,\ldots,n$,并满足$w_q\in[0,1]$,$\sum\limits_{q=1}^{n}w_q=1$. 邀请多位专家组成一个决策小组来评价每种方案在各个属性下的属性值. 由于各专家之间不能相互说明彼此,用犹豫模糊数h_{pq}来表示每种方案$A_p(p=1,2,\ldots,m)$对各属性$G_q(q=1,2,\ldots,n)$的隶属度. 专家组提供的所有决策信息可构成犹豫模糊决策矩阵$D=(h_{pq})_{m\times n}$.

根据式(7)计算犹豫模糊指数熵,然后利用信息熵最小化原则确定属性权重

$$w_q=\frac{1-E_q}{n-\sum\limits_{q=1}^{n}E_q},\ q=1,2,\ldots,n. \tag{8}$$

这里$E_q=\dfrac{1}{m}\sum\limits_{p=1}^{m}E(h_{pq}),\ p=1,2,\ldots,m,\ q=1,2,\ldots,n.$

4 实例分析

为了更好地说明新提出的基于指数熵加权的降维犹豫模糊兰氏距离的有效性和实时性,结合医生给病人医疗诊断的实例数据,首先分别采用新提出的基于指数熵加权的降维犹豫模糊兰氏距离公式和文献[3]中Xu等提出的广义加权豪斯道夫距离从医疗诊断效果和诊断时间两个方面进行对比分析;然后,利用该实例对新提出的几种基于兰氏距离的犹豫模糊集距离公式进行对比分析.

例1 病毒性发热、痢疾、伤寒、胃病、心脏病是5种常见的疾病. 医生可以通过3位病人A、B、C的体温、头疼、咳嗽、胃疼、胸痛的临床表现为每个病人确诊. 这里5种疾病各个特征的标准信息以及3个病人的临床症状信息均用犹豫模糊集表示,详见表1和表2所示.

表1 5种疾病各个特征的标准信息

疾病	体温	头疼	咳嗽	胃疼	胸痛
发热	{0.3, 0.4, 0.6, 0.8}	{0.2, 0.3, 0.5, 0.7}	{0.3, 0.5}	{0.1, 0.2, 0.3, 0.4, 0.5}	{0.2, 0.4, 0.5}
痢疾	{0.6, 0.7, 0.8, 0.9}	{0.1, 0.2, 0.3, 0.5}	{0.1, 0.2}	{0.1, 0.2, 0.3, 0.5, 0.6}	{0.1, 0.2, 0.4}
伤寒	{0.1, 0.3, 0.6, 0.7}	{0.6, 0.7, 0.8, 0.9}	{0.3, 0.5}	{0.1, 0.2, 0.3, 0.4, 0.5}	{0.1, 0.4, 0.6}
胃病	{0.2, 0.4, 0.5, 0.6}	{0.1, 0.2, 0.3, 0.4}	{0.3, 0.4}	{0.5, 0.6, 0.7, 0.8, 0.9}	{0.1, 0.2, 0.4}
心脏病	{0.1, 0.2, 0.3, 0.4}	{0.1, 0.2, 0.3, 0.5}	{0.2, 0.3}	{0.2, 0.3, 0.5, 0.6, 0.7}	{0.7, 0.8, 0.9}

表2 3个病人的临床症状信息

病人	体温	头疼	咳嗽	胃疼	胸痛
A	{0.6, 0.8}	{0.1, 0.3, 0.4}	{0.25, 0.3, 0.4, 0.45}	{0.2, 0.3, 0.4 }	{0.1, 0.4 }
B	{0.7, 0.9}	{0.3, 0.4, 0.6}	{0.15, 0.2, 0.25, 0.3}	{0.2, 0.3, 0.4}	{0.1, 0.6}
C	{0.6, 0.9}	{0.1, 0.4, 0.65}	{0.25, 0.3, 0.35, 0.4}	{0.2, 0.3, 0.4}	{0.2, 0.4}

基于指数熵加权的降维犹豫模糊兰氏距离具体步骤如下.

1) 结合3个实际病人的临床症状信息(见表2),利用式(7)计算犹豫模糊指数熵,以$h_{11}=\{0.6,0.8\}$为例.

$$E(h_{11})=\frac{1}{2(\sqrt{\mathrm{e}}-1)}\times(0.6\mathrm{e}^{0.4}+0.4\mathrm{e}^{0.6}-1+0.8\mathrm{e}^{0.2}+0.2\mathrm{e}^{0.8}-1)=0.806\,3.$$

计算所有的犹豫模糊指数熵,得到的结果如表3所示.

表3 犹豫模糊指数熵矩阵

病人	体温	头疼	咳嗽	胃疼	胸痛
A	0.806 3	0.726 3	0.889 4	0.819 7	0.666 4
B	0.608 6	0.923 3	0.694 5	0.819 7	0.666 4
C	0.666 4	0.748 9	0.870 2	0.819 7	0.806 3

计算各临床症状信息下的平均熵,结果如下:

$E_1 = \dfrac{1}{3} \times (0.8063 + 0.6086 + 0.6664) = 0.6938.$

同理可以计算得到: $E_2 = 0.7995, E_3 = 0.8180, E_4 = 0.8197, E_5 = 0.7130$.

2) 利用式(8)计算得到属性权重 $w_1 = 0.2649, w_2 = 0.1734, w_3 = 0.1574, w_4 = 0.1560, w_5 = 0.2483$.

3) 利用2.2节提出的犹豫模糊数乐观降维方案将5种疾病各个特征的标准信息和3个病人的临床症状信息数据进行预处理,结果见表4和表5.

表 4 降维后的5种疾病各个特征的标准信息

疾病	体温	头疼	咳嗽	胃疼	胸痛
发热	{0.4333, 0.8}	{0.25, 0.5, 0.7}	{0.3, 0.5}	{0.2, 0.4, 0.5}	{ 0.3, 0.5 }
痢疾	{0.7, 0.9}	{0.15, 0.3, 0.5}	{0.1, 0.2}	{0.2, 0.5, 0.6}	{0.15, 0.4}
伤寒	{0.3333, 0.7}	{0.65, 0.8, 0.9}	{0.3, 0.5}	{0.2, 0.4, 0.5}	{0.25, 0.6}
胃病	{0.3667, 0.6}	{0.15, 0.3, 0.4}	{0.3, 0.4}	{0.6, 0.8, 0.9}	{0.15, 0.4}
心脏病	{0.2, 0.4}	{0.15, 0.3, 0.5}	{0.2, 0.3}	{0.3333, 0.6, 0.7}	{0.75, 0.9}

表 5 降维后的3个病人的临床症状信息

病人	体温	头疼	咳嗽	胃疼	胸痛
A	{0.6, 0.8}	{0.1, 0.3, 0.4 }	{0.3167, 0.45}	{0.2, 0.3, 0.4}	{0.1, 0.4}
B	{0.7, 0.9}	{0.3, 0.4, 0.6 }	{0.2, 0.3}	{0.2, 0.3, 0.4}	{0.1, 0.6}
C	{0.6, 0.9}	{0.1, 0.4, 0.65}	{0.3, 0.4}	{0.2, 0.3, 0.4}	{0.2, 0.4}

4) 利用3.1节提出的广义加权犹豫模糊兰氏距离(式(4))计算每个病人与每种疾病之间的距离测度,结果见表6~表8. 表中$\lambda = 1$时,是采用加权犹豫模糊兰氏距离(式(3))计算每个病人与每种疾病之间的距离测度.

表 6 采用新方法计算病人A与各种疾病间的距离测度

疾病	$\lambda=1$	$\lambda=2$	$\lambda=4$	$\lambda=6$	$\lambda=8$	$\lambda=10$
发热	0.1717	0.2377	0.3193	0.3643	0.3916	0.4098
痢疾	0.1748	0.2407	0.3208	0.3674	0.3970	0.4174
伤寒	0.2351	0.3045	0.4006	0.4682	0.5171	0.5525
胃病	0.1637	0.2224	0.2933	0.3375	0.3664	0.3864
心脏病	0.3491	0.4023	0.4879	0.5497	0.5922	0.6217

表 7 采用新方法计算病人B与各种疾病间的距离测度

疾病	$\lambda=1$	$\lambda=2$	$\lambda=4$	$\lambda=6$	$\lambda=8$	$\lambda=10$
发热	0.1771	0.2262	0.3050	0.3546	0.3855	0.4059
痢疾	0.1478	0.1877	0.2241	0.2466	0.2626	0.2742
伤寒	0.2175	0.2606	0.3050	0.3298	0.3458	0.3573
胃病	0.2532	0.2725	0.3119	0.3447	0.3692	0.3875
心脏病	0.3216	0.3988	0.4950	0.5554	0.5954	0.6235

表 8 采用新方法计算病人C与各种疾病间的距离测度

疾病	$\lambda=1$	$\lambda=2$	$\lambda=4$	$\lambda=6$	$\lambda=8$	$\lambda=10$
发热	0.1231	0.1565	0.2192	0.2679	0.3003	0.3223
痢疾	0.1443	0.2038	0.2825	0.3328	0.3658	0.3885
伤寒	0.1860	0.2471	0.3695	0.4573	0.5137	0.5515
胃病	0.1794	0.2318	0.2955	0.3380	0.3665	0.3864
心脏病	0.3356	0.3672	0.4136	0.4443	0.4658	0.4818

5) 根据距离大小对3个病人进行诊断. 如果某种疾病和病人之间的距离测度越小,则该病人得此病的可能性就越大.

从表6~表8可以看出,3个病人与各种疾病间的距离测度的数值都随着λ的增加而增加,而且不论λ取何值,诊断结果均为病人A患胃病,B患痢疾,C患病毒性发热. 本文提出的指数熵加权的降维犹豫模糊兰氏距离测度,不需要反复添加最大最小隶属度数值到犹豫模糊数中,尽可能地保留了数据的原始信息,在λ取不同值时,取得了一致的决策结果.

采用文献[3]中Xu等提出的广义加权豪斯道夫距离测度,并采用该文中的乐观补齐方法计算每个病人与每种疾病之间的距离. 为了更好地比较两种距离,这里权值仍取$w_1 = 0.2649, w_2 = 0.1734, w_3 = 0.1574, w_4 = 0.1560, w_5 = 0.2483$. 具体实验结果见表9~表11.

表 9 采用Xu等方法计算病人A与各种疾病间的距离测度

疾病	$\lambda=1$	$\lambda=2$	$\lambda=4$	$\lambda=6$	$\lambda=8$	$\lambda=10$
发热	0.2299	0.2614	0.3034	0.3270	0.3416	0.3515
痢疾	0.2531	0.2101	0.2050	0.2097	0.2143	0.2182
伤寒	0.3159	0.3569	0.4093	0.4361	0.4510	0.4604
胃病	0.2667	0.3076	0.3609	0.3904	0.4090	0.4222
心脏病	0.3957	0.4517	0.5104	0.5373	0.5521	0.5613

表 10 采用Xu等方法计算病人B与各种疾病间的距离测度

疾病	$\lambda=1$	$\lambda=2$	$\lambda=4$	$\lambda=6$	$\lambda=8$	$\lambda=10$
发热	0.2623	0.3060	0.3676	0.4028	0.4241	0.4380
痢疾	0.2513	0.2716	0.3032	0.3246	0.3392	0.3494
伤寒	0.3234	0.3696	0.4398	0.4825	0.5085	0.5254
胃病	0.3933	0.4085	0.4304	0.4447	0.4544	0.4615
心脏病	0.4490	0.5004	0.5602	0.5919	0.6112	0.6242

表 11　采用Xu等方法计算病人C与各种疾病间的距离测度

疾病	$\lambda=1$	$\lambda=2$	$\lambda=4$	$\lambda=6$	$\lambda=8$	$\lambda=10$
发热	0.2304	0.2837	0.3607	0.4009	0.4235	0.4378
痢疾	0.2260	0.2330	0.2503	0.2681	0.2830	0.2943
伤寒	0.3424	0.3956	0.4627	0.4971	0.5173	0.5307
胃病	0.3365	0.3714	0.4138	0.4364	0.4501	0.4591
心脏病	0.4328	0.4783	0.5368	0.5732	0.5977	0.6149

从表9～表11可以看出,诊断结果为病人B和病人C患痢疾,病人A在$\lambda=1$时患病毒性发热,在$\lambda=2$、4、6、8、10时患痢疾,出现了诊断结果不一致的情况. Xu等提出的距离测度需要通过反复添加最大最小隶属度数值到犹豫模糊数中的方法会一定程度上改变原始的数据信息,具有一定的局限性. 利用这样的数据进行距离测度的计算有时会导致计算结果不准确,进而可能会改变最终的决策结果.

为了更好地验证新方法采用降维方案可以减少计算距离时的计算量, 在Matlab R2014a, Windows 7, 1 TB硬盘, 16 GB内存, Intel(R) Core(TM) i7-6700CPU的实验环境下对新提出的基于指数熵加权的降维犹豫模糊兰氏距离公式和文献[3]中Xu等提出的广义加权豪斯道夫距离两种方法的程序运行时间进行对比实验. 这里以病人A为例,利用其实际数据采用两种方法分别计算该病人与病毒性发热、痢疾、伤寒、胃病、心脏病5种常见疾病之间的距离. 为了保证数据的准确性,这里计算病人A与每一种疾病距离间的程序运行时间,采用程序运行30次的平均时间. 具体实验结果见表12,单位为s.

表 12　两种方法的运行时间比较

	新方法	Xu等方法
发热	0.0242	0.0326
痢疾	0.0264	0.0349
伤寒	0.0240	0.0307
胃病	0.0251	0.0319
心脏病	0.0249	0.0308
总计	0.1246	0.1609

从表12的实验结果可以看出,Xu等的方法对病人A做出诊断需要程序运行时间为0.1609 s,而本文提出的新方法需要0.1246 s. 因此,本文提出的新方法不仅很好地保留了数据的原始信息,取得了一致的决策结果,并且在一定程度上减少了计算量,提高了诊断效率.

为了更好地对比分析本文提出的几种基于兰氏距离的犹豫模糊集距离测度公式,结合上面的医疗诊断实例,利用广义犹豫模糊兰氏距离(式(2))计算每个病人与每种疾病之间的距离测度,具体实验结果见表13～表15. 表中$\lambda=1$时,是采用标准化的犹豫模糊兰氏距离(式(1))计算每个病人与每种疾病之间的距离测度.

表 13　采用广义犹豫模糊兰氏距离公式计算病人A与各种疾病间的距离

疾病	$\lambda=1$	$\lambda=2$	$\lambda=4$	$\lambda=6$	$\lambda=8$	$\lambda=10$
发热	0.1656	0.2297	0.3101	0.3556	0.3837	0.4026
痢疾	0.1748	0.2390	0.3174	0.3642	0.3943	0.4151
伤寒	0.2278	0.3050	0.4084	0.4775	0.5259	0.5603
胃病	0.1696	0.2351	0.3090	0.3512	0.3778	0.3961
心脏病	0.3187	0.3722	0.4628	0.5301	0.5763	0.6084

表 14　采用广义犹豫模糊兰氏距离公式计算病人B与各种疾病间的距离

疾病	$\lambda=1$	$\lambda=2$	$\lambda=4$	$\lambda=6$	$\lambda=8$	$\lambda=10$
发热	0.1690	0.2143	0.2910	0.3424	0.3753	0.3972
痢疾	0.1611	0.1981	0.2328	0.2541	0.2689	0.2796
伤寒	0.2129	0.2531	0.2965	0.3220	0.3388	0.3510
胃病	0.2599	0.2820	0.3243	0.3567	0.3799	0.3969
心脏病	0.2854	0.3660	0.4690	0.5352	0.5792	0.6100

表 15　采用广义犹豫模糊兰氏距离公式计算病人C与各种疾病间的距离

疾病	$\lambda=1$	$\lambda=2$	$\lambda=4$	$\lambda=6$	$\lambda=8$	$\lambda=10$
发热	0.1196	0.1574	0.2249	0.2739	0.3057	0.3269
痢疾	0.1669	0.2253	0.2995	0.3464	0.3769	0.3979
伤寒	0.1821	0.2517	0.3810	0.4681	0.5229	0.5594
胃病	0.1864	0.2449	0.3111	0.3516	0.3779	0.3961
心脏病	0.3077	0.3403	0.3922	0.4277	0.4526	0.4709

从表13～表15可以看出,诊断结果为病人B患痢疾,病人C患病毒性发热. 但是病人A在$\lambda=1$、2时,患病毒性发热,在$\lambda=4$、6、8、10时,患胃病,出现了诊断结果不一致的情况. 因为标准化的犹豫模糊兰氏距离和广义犹豫模糊兰氏距离没有充分考虑不同特征信息赋予不同的属性权重问题,所以导致了不一致的决策结果. 而在属性权重信息完全未知的情况下,采用广义加权犹豫模糊兰氏距离(式(4))计算每个病人与每种疾病之间的距离测度,取得了一致合理的决策结果(见表6～表8). 表中$\lambda=1$时,是采用加权犹豫模糊兰氏距离(式(3))计算每个病人与每种疾病之间的距离测度. 这里利用每个病人的5种临床症状数据信息构造犹豫模糊指数熵,进而得到各自特征不同的属性权重.

5　结　论

犹豫模糊集允许多个隶属度数值的存在,能够较好地体现决策者的犹豫心理,因此非常便于处理复杂的不确定性决策问题. 本文首先将传统的兰氏距

离推广到犹豫模糊领域,提出了几种基于兰氏距离的犹豫模糊集距离公式. 然后针对传统数据补齐方案需要人为添加最大最小隶属度数值到犹豫模糊数中的缺点,提出新的犹豫模糊数降维方案. 该方案不仅很好地保留了数据的原始信息,而且减少了计算距离时的计算量. 针对属性权重信息完全未知的情况下,本文采用实际问题中5种疾病各个特征的标准信息以及3个病人的临床症状数据信息构造犹豫模糊指数熵,并利用信息熵最小化原则计算得到属性权重. 最后,结合一个医疗诊断实例将新提出的基于指数熵加权的降维犹豫模糊兰氏距离测度和Xu等提出的距离测度从诊断效果和诊断时间两个方面进行对比. 结果表明,本文提出的基于指数熵加权的降维犹豫模糊兰氏距离测度不仅在λ取不同值时诊断结果一致,而且减少了计算量,提高了诊断效率. 面对医疗大数据到来的今天,新方法对实时、有效的医疗诊断具有一定的应用价值.

参考文献(References)

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

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

[3] Xu Zeshui, Xia Meimei. Distance and similarity measures for hesitant fuzzy sets[J]. Information Sciences, 2011, 181(11): 2128-2138.

[4] Zhu Bin, Xu Zeshui, Xia Meimei. Hesitant fuzzy geometric bonferroni means[J]. Information Sciences, 2012, 205: 72-85.

[5] 陈树伟, 蔡丽娜. 区间值犹豫模糊集[J]. 模糊系统与数学, 2013, 27(6): 38-44.
(Chen S W, Cai L N. Interval-valued hesitant fuzzy sets[J]. Fuzzy Systems and Mathematics, 2013, 27(6): 38-44.)

[6] 吴婉莹, 何迎东, 郭甦, 等. 直觉对偶犹豫模糊集的集结算子及其应用[J]. 武汉理工大学学报:信息与管理工程版, 2014, 36(2): 225-228.
(Wu W Y, He Y D, Guo S, et al. Intuitionistic dual hesitant fuzzy set operators and its application[J]. Journal of Wuhan University of Technology: Information and Management Engineering, 2014, 36(2): 225-228.)

[7] Li Deqing, Zeng Wenyi, Li Junhong. New distance and similarity measures on hesitant fuzzy sets and their applications in multiple criteria decision making[J]. Engineering Applications of Artificial Intelligence, 2015, 40(5): 11-16.

[8] 谢海. 犹豫模糊集的截集及其性质[J]. 数学的实践与认识, 2017, 47(3): 251-257.
(Xie H. On properties of level sets of hesitant fuzzy sets[J]. Journal of Mathematics in Practice and Theory, 2017, 47(3): 251-257.)

[9] 关欣, 孙贵东, 衣晓, 等. 基于犹豫模糊集统计相关系数的多源异类数据融合识别[J]. 系统工程与电子技术, 2018, 40(3): 509-517.
(Guan X, Sun G D, Yi X, et al. Multi-source heterogeneous data fusion recognition based on statistical correlation coefficients between hesitant fuzzy sets[J]. Journal of Systems Engineering and Electronics, 2018, 40(3): 509-517.)

[10] Hu Junhua, Zhang Xiaolong. Hesitant fuzzy information measures and their applications in multi-criteria decision making[J]. International Journal of Systems Science, 2016, 47(1): 62-76.

[11] 林松, 刘小弟, 朱建军, 等. 基于改进符号距离的权重未知犹豫模糊决策方法[J]. 控制与决策, 2018, 33(1): 186-192.
(Lin S, Liu X D, Zhu J J, et al. Hesitant fuzzy decision making method with unknown weight information based on an improved signed distance[J]. Control and Decision, 2018, 33(1): 186-192.)

[12] 彭定洪, 杨东可. 一种完全犹豫模糊环境下的Topsis方法[J]. 模糊系统与数学, 2018, 32(1): 39-50.
(Peng D H, Yang D K. A Topsis method in completely hesitant fuzzy envionment[J]. Fuzzy Systems and Mathematics, 2018, 32(1): 39-50.)

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

[14] Xia Meimei, Xu Zeshui. Hesitant fuzzy information aggregation in decision making[J]. International Journal of Approximate Reasoning, 2011, 52(3): 395-407.

[15] 高惠璇. 应用多元统计分析[M]. 第1版. 北京: 北京大学出版社, 2005: 218-228.
(Gao H X. Application of multivariate statistical analysis[M]. 1st edition. Beijing: Beijing University Press, 2005: 218-228.)

[16] 谭吉玉, 刘高常. 犹豫模糊集指数熵及其应用[J]. 统计与决策, 2017, 492(24): 66-69.
(Tan J Y, Liu G C. Hesitant fuzzy exponential entropy and its application[J]. Statistics and Decision, 2017, 492(24): 66-69.)

作者简介

沙秀艳(1977−), 女, 讲师, 博士生, 从事决策分析、数据预测的研究, E-mail: shaxiuyan@sina.com;

尹传存(1963−), 男, 教授, 博士生导师, 从事决策分析和概率统计等研究, E-mail: ccyin@qfnu.edu.cn;

徐泽水(1968−), 男, 教授, 博士生导师, 从事决策分析、信息融合等研究, E-mail: xuzeshui@263.net.

(责任编辑: 齐 霁)