计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210700191

带关注度模糊序决策数据集的分布约简

徐伟华　张俊杰　陈修伟
西南大学人工智能学院　重庆 400715

摘　要　随着大数据时代的到来,数据的结构变得越来越复杂,数据集的维度变得越来越高,这极大地影响了数据挖掘的效率。因此,很有必要进行数据压缩或对信息系统进行属性约简,即去掉不必要的冗余属性,降低数据维度,提高数据挖掘效率。在现实生活中,人们对数据集中每个条件属性的关注度往往是不一样的。首先,在经典模糊决策数据集的基础上,对不同的条件属性进行加权,定义加权得分函数,进一步建立带关注度的模糊序决策信息系统。然后在该系统中引入分布函数,并通过分布可辨识矩阵建立求分布约简的方法。最后,通过案例分析验证了该方法的可行性。

关键词:模糊集;序决策数据集;关注度;分布约简;分布可辨识矩阵
中图法分类号　TP18

Distribution Reduction in Fuzzy Order Decision Data Sets with Attention Degree

XU Wei-hua,ZHANG Jun-jie and CHEN Xiu-wei
College of Artificial Intelligence,Southwest University,Chongqing 400715,China

Abstract　With the advent of the era of big data,the structure of data becomes more and more complex,and the dimensions of data set become higher and higher,which will affect the efficiency of data mining greatly. Therefore,it is necessary to perform data compression or attribute reduction to information systems,that is,to remove unnecessary redundant attributes,reduce data dimensions,and improve the efficiency of data mining. The reduction methods proposed by many scholars in the past regard each attribute as equally important. But in real life,people's attention to each conditional attribute in the data set is often different. Aiming at this phenomenon,based on the classical fuzzy decision data set,this paper weights different conditional attributes,defines the weighted score function,and further establishes the fuzzy order decision information system with attention degree. Then the distribution function is introduced into the system and the distribution reduction method is established by the distribution discernible matrix. Finally,the feasibility of the method is verified by a case study.

Keywords　Fuzzy set,Order decision data set,Attention degree,Distribution reduction,Distribution discernible matrix

1　引言

模糊现象在现实世界中大量存在,如颜色的深浅、身高的高矮程度以及人们的喜怒哀乐的程度。在 20 世纪 60 年代,美国的控制论专家 Zadeh 开始使用数学方法来研究模糊性并根据元素的隶属度提出了模糊集[1-2]的概念。该理论利用数学方法去定量化研究模糊性,从而为数学在模糊控制领域的应用打开了一片全新的天地。模糊集理论不同于概率论,它利用样本的隶属度描述不确定性,可以反映一个样本属于一个集合的程度。而传统的经典集合论考虑一个元素要么属于一个集合,要么不属于一个集合,条件过于严苛。正因如此,模糊集理论被广泛应用于数据挖掘和机器学习中,如分类问题[3-5]、回归问题[6-8]以及聚类问题[9-11]等等。

另一方面,粗糙集理论[12]是由波兰学者 Pawlak 在 1982 年提出的一种处理不确定性、模糊知识问题的工具。它广泛应用于数据挖掘、模式识别等领域。属性约简是粗糙集理论研究的一个重要内容。随着大数据时代的到来,数据的维度变得越来越大,因此属性约简作为一种数据预处理的手段在数据挖掘[13-17]方面被广泛使用。研究实际问题时,对于一个知识库来说,并不是所有的属性都是必要的,因此我们可以通过属性约简[18-20],在保证知识库决策能力不变的前提下,将冗余的属性去掉,使知识得到简化并且不丢失必要信息,从而提高分类准确度。通过属性约简减少了信息系统中的冗余属性,降低了高维数据的维度,从而在一定程度上可以节省计算的时间成本和空间成本。随着属性约简的发展,很多约简方法和特征选取方法被提出,以针对不同类型的复杂数据进行建模,如区间值型数据[21]、集值型数据[22]、模糊值型数据[23]等。本文研究的正是如何利用粗糙集理论建立针对模糊型数据的约简方法[24-25]。

虽然很多学者研究了基于优势关系下的信息系统[26-28],但在这些工作中,每个属性被看作是同等重要的。然而在现实生活中,人们对于一个信息系统中的每个属性的关注程度是不一样的,出于不同的目的,有些属性应该被重点关注,而有些属性则不那么重要,甚至可以忽略。因此,本文考虑对条件属性进行加权,并且建立加权得分函数,从而建立带关注度的模糊序决策信息系统。在基于带关注度的模糊序决策信息

基金项目:国家自然科学基金面上项目(61976245)
This work was supported by the National Natural Science Foundation of China(61976245).
通信作者:徐伟华(chxuwh@gmail.com)

系统中引入分布约简的概念，得到分布约简的判定定理以及辨识矩阵，建立带关注度的模糊序决策信息系统下的分布约简的具体方法，并通过案例分析验证此方法的有效性。

2 带关注度的模糊序决策数据集

决策数据集是不仅有条件属性而且有决策属性的一种特殊信息系统，研究的主要问题是条件属性和决策属性之间的关系。下面给出一些相关概念。

定义1 称(U,AT,F)为一个信息系统，其中U是有限对象集，记作$U=\{x_1,x_2,\cdots,x_n\}$；AT是有限条件属性集，记作$AT=\{a_1,a_2,\cdots,a_p\}$；$F$是$U$与$AT$的关系集，记作$F=\{f_i:U\rightarrow V_i,i\leqslant n\}$，其中$V_i$为$a_i$的有限值域。更进一步，称五元组$I=(U,AT,F,D,G)$为一个决策信息系统，其中$(U,AT,F)$为一个信息系统；$D$是有限目标属性集，记作$D=\{d_1,d_2,\cdots,d_q\}$；$G$是$U$与$D$的关系集，记做$G=\{g_i:U\rightarrow V'_j,j\leqslant q\}$，其中$V'_j$为$d_j$的有限值域。

定义2 已知$I=(U,AT,F,D,G)$是一个决策信息系统，R为实数，对$\forall f\in F,g\in G,a\in AT$和$x\in U$都有：

$$f(x,a,\omega)=(\theta_a(x),\omega\theta_a(x)),g(x,d)\in R$$

其中，函数$\theta_a:U\rightarrow[0,1]$和函数$\omega\theta_a:U\rightarrow[0,1]$分别表示$U$中对象$x$在条件属性$a$下的隶属度和加权隶属度，并且始终满足$\sum_{i=1}^{n}\omega_i=1,\omega_n\in\Omega$，称$f(x,a,\omega)$为加权模糊数，$\omega$为关注度，$\Omega$为关注度的集合。进一步，记$f(a)=\{(\theta_a(x),\omega\theta_a(x))\mid a\in AT\}$，称$f(a)$为$U$上的加权模糊集。如果满足上述性质，则称$I^*=(U,AT,F,\Omega)$是带关注度的模糊信息系统，$I^*=(U,AT,D,F,G,\Omega)$是带关注度的模糊决策信息系统。

下面给出带关注度的模糊决策信息系统的得分函数，以及在得分函数下的序关系。

定义3 设$I^*=(U,AT,F,\Omega)$是带关注度的模糊信息系统，$\forall x\in U$，定义对象x在条件属性集AT下的得分函数：

$$S_a(x)=\omega_1\theta_{a_1}(x)+\omega_2\theta_{a_2}(x)+\cdots+\omega_n\theta_{a_n}(x)$$

其中ω和$\theta_a(x)$分别表示对象x在条件属性a下的权重和隶属度，且满足$\sum_{i=1}^{n}\omega_i=1$。需要注意的是，ω_i为条件属性的权重，权重越大，说明得分评价时越看重此属性。因此，在实际应用中要根据具体需求给出相应的权重，从而确定得分函数。

定义4 设$I^*=(U,AT,D,F,G,\Omega)$是带关注度的模糊决策信息系统，对$\forall f\in F,g\in G,a\in AT$和$x_i,x_j\in U$，都有：

$$f(x_i,a,\omega)\geqslant f(x_j,a,\omega)\Leftrightarrow S_a(x_i)\geqslant S_a(x_j)$$

$$f(x_i,a,\omega)\leqslant f(x_j,a,\omega)\Leftrightarrow S_a(x_i)\leqslant S_a(x_j)$$

$$g(x_i,d)\geqslant g(x_j,d),g(x_i,d)\leqslant g(x_j,d)$$

则称$I_*^{\leqslant}=(U,AT,D,F,G,\Omega)$是带关注度的模糊决策序信息系统。

在带关注度的模糊决策序信息系统中，根据构造的得分函数，可以得到在条件属性a下的递增或递减的偏序关系。若在某一条件属性的值域下存在递增或递减的偏序关系，则称该条件属性是带关注度的模糊决策信息系统的一个准则。本文主要讨论由递增偏序关系获得的优势关系，则由递减偏序关系获得的优势关系的情况可以得到类似结论。

在带关注度的模糊决策序信息系统中，$a\in AT$，对于$\forall x,y\in U$，存在优势关系“$\geqslant$”，“$x\geqslant y$”表示x关于准则a优于y。

定义5 设$I_*^{\leqslant}=(U,AT,D,F,G,\Omega)$是带关注度的模糊序决策信息系统，对于$A\subseteq AT$，令：

$$\begin{aligned}R_A^{\leqslant}&=\{(x_i,x_j)\in U\times U\mid x_i\leqslant x_j\}\\&=\{(x_i,x_j)\in U\times U\mid f(x_i,a,\omega)\leqslant f(x_j,a,\omega),\forall a\in AT,f\in F,\omega\in\Omega\}\\&=\{(x_i,x_j)\in U\times U\mid S_a(x_i)\leqslant S_a(x_j),\forall a\in AT,f\in F,\omega\in\Omega\}\end{aligned}$$

$$R_d^{\leqslant}=\{(x_i,x_j)\in U\times U\mid g(x_i,d)\leqslant g(x_j,d),g\in G\}$$

则称$R_A^{\leqslant}$，$R_d^{\leqslant}$为带关注度的模糊决策序信息系统的优势关系。在一个带关注度的模糊决策序信息系统中，若$R_{AT}^{\leqslant}\subseteq R_d^{\leqslant}$，我们称该带关注度的模糊决策序信息系统是协调的，若$R_{AT}^{\leqslant}\not\subset R_d^{\leqslant}$，则称这个系统是不协调的。但是在日常生活中，由于各种原因我们得到的信息系统常常是不协调的。

由$R_A^{\leqslant}$，$R_d^{\leqslant}$可以得出优势类的表达式如下。

记：

$$\begin{aligned}[x_i]_A^{\leqslant}&=\{x_j\in U\mid(x_i,x_j)\in R_A^{\leqslant}\}\\&=\{x_j\in U\mid S_a(x_i)\leqslant S_a(x_j),\forall a\in A,f\in F,\omega\in\Omega\}\end{aligned}$$

$$[x_i]_d^{\leqslant}=\{x_j\in U\mid g(x_i,d)\geqslant g(x_j,d),\forall g\in G\}$$

性质1 设$I^*=(U,AT,D,F,G,\Omega)$是带关注度的模糊序决策信息系统，$R_A^{\leqslant}$，$R_d^{\leqslant}$为带关注度的模糊决策序信息系统的优势关系，则满足以下性质：

(1) $R_A^{\leqslant}$满足自反性和传递性，但不一定满足对称性，一般不是等价关系；

(2) 当$A_1\subseteq A_2\subseteq AT$时有$R_{AT}^{\leqslant}\subseteq R_{A_2}^{\leqslant}\subseteq R_{A_1}^{\leqslant}$；

(3) 当$A_1\subseteq A_2\subseteq AT$时有$[x_i]_{AT}^{\leqslant}\subseteq[x_i]_{A_2}^{\leqslant}\subseteq[x_i]_{A_1}^{\leqslant}$；

(4) 当$x_j\in[x_i]_A^{\leqslant}$时有$[x_j]_A^{\leqslant}\subseteq[x_i]_A^{\leqslant}$。

3 带关注度的模糊序决策数据集的分布约简理论

本节给出带关注度的模糊序决策信息系统的分布函数、最大分布函数、分布约简，及最大分布约简的定义。

定义6 已知$I^*=(U,AT,D,F,G,\Omega)$是一个带关注度的模糊序决策信息系统，对于$A\subseteq AT,x\in U$记：

$$U/R_A^{\leqslant}=\{[x_i]_A^{\leqslant}:x_i\in U\}$$

$$U/R_d^{\leqslant}=\{D_1,D_2,\cdots,D_r\}$$

$$\mu_A(x)=\left(\frac{|D_1\cap[x]_A^{\leqslant}|}{|U|},\frac{|D_2\cap[x]_A^{\leqslant}|}{|U|},\cdots,\frac{|D_r\cap[x]_A^{\leqslant}|}{|U|}\right)$$

$$\gamma_A(x)=\max\left\{\frac{|D_1\cap[x]_A^{\leqslant}|}{|U|},\frac{|D_2\cap[x]_A^{\leqslant}|}{|U|},\cdots,\frac{|D_r\cap[x]_A^{\leqslant}|}{|U|}\right\}$$

由于带关注度的模糊序决策信息系统是基于优势关系的，因此在对象集上形成的是一个覆盖，则$U/R_A^{\leqslant}$表示所有优势类，$\mu_A(x)$，$\gamma_A(x)$分别称为论域U上对象x关于条件属性子集A和决策属性d的分布函数和最大分布函数。

定义7 设$\boldsymbol{\alpha}=(a_1,a_2,\cdots,a_n)^{\mathrm{T}}$和$\boldsymbol{\beta}=(b_1,b_2,\cdots,b_n)^{\mathrm{T}}$是两个$n$维向量，若$a_i=b_i(i=1,2,\cdots,n)$，则称向量$\boldsymbol{\alpha}$等于向量$\boldsymbol{\beta}$，记作$\boldsymbol{\alpha}=\boldsymbol{\beta}$；若$a_i\leqslant b_i(i=1,2,\cdots,n)$，则称向量$\boldsymbol{\alpha}$小于等于向量$\boldsymbol{\beta}$，记作$\alpha\leqslant\beta$；否则若$\exists i_0,i_0\in\{1,2,\cdots,n\}$，使得$a_{i_0}>b_{i_0}$，称向量$\boldsymbol{\alpha}$不小于等于$\boldsymbol{\beta}$，记作$\boldsymbol{\alpha}\not\leqslant\boldsymbol{\beta}$。

由以上定理可得以下相关性质。

性质2 设$I_*^{\leqslant}=(U,AT,D,F,G,\Omega)$是一个模糊序决策信息系统，对于$\forall A\in AT$，以下性质成立：

(1)当 $B\subseteq A$ 时，对 $\forall x\in U$ 有$\mu_A(x)\leqslant\mu_B(x)$；

(2)当 $B\subseteq A$ 时，对 $\forall x\in U$ 有$\gamma_A(x)\leqslant\gamma_B(x)$；

(3)当 $\forall x,y\in U$，$[y]_A^{\leqslant}\subseteq[x]_A^{\leqslant}$ 时，有$\mu_A(x)\leqslant\mu_B(x)$；

(4)当 $\forall x,y\in U$，$[y]_A^{\leqslant}\subseteq[x]_A^{\leqslant}$ 时，有$\gamma_A(x)\leqslant\gamma_B(x)$。

证明：(1)－(4)由定义 5 可得。

定义 8　设$I_*^{\leqslant}=(U,AT,D,F,G,\Omega)$是不协调模糊序决策信息系统。若 $\forall x\in U$，$A\subseteq AT$ 有$\mu_A(x)=\mu_{AT}(x)$，则称 A 是$I_*^{\leqslant}$中关于序关系$R_{AT}^{\leqslant}$的分布相对协调集，简称分布协调集。进而，若 A 的任何真子集不是分布协调集，则称 A 是$I_*^{\leqslant}$中关于序关系$R_{AT}^{\leqslant}$的分布相对约简，简称分布约简。

定义 9　设$I_*^{\leqslant}=(U,AT,D,F,G,\Omega)$是不协调模糊序决策信息系统。若 $\forall x\in U$，$A\subseteq AT$ 有$\gamma_A(x)=\gamma_{AT}(x)$，则称 A 是$I_*^{\leqslant}$中关于序关系$R_{AT}^{\leqslant}$的最大分布相对协调集，简称最大分布协调集。进而，若 A 的任何真子集不是最大分布协调集，则称 A 是$I_*^{\leqslant}$中关于序关系$R_{AT}^{\leqslant}$的最大分布相对约简，简称最大分布约简。

命题 1　设$I_*^{\leqslant}=(U,AT,D,F,G,\Omega)$是一个不协调模糊序决策信息系统，$A\subseteq AT$，则 A 是分布协调集，当且仅当 A 是最大分布协调集。

证明：根据分布协调集、最大分布协调集的定义和序关系的定义可得，在模糊序决策信息系统中，当 A 为分布协调集时，对 $\forall x\in U$ 有$[x]_A^{\leqslant}\subseteq[x]_{AT}^{\leqslant}$；反之，当 $\forall x\in U$ 满足$[x]_A^{\leqslant}\subseteq[x]_{AT}^{\leqslant}$时，可得到 A 为分布协调集。最大分布协调集同理，因此可得：

A 是分布协调集$\Leftrightarrow\forall x\in U,[x]_A^{\leqslant}\subseteq[x]_{AT}^{\leqslant}\Rightarrow A$ 是最大分布协调集。

推论 1　设$I_*^{\leqslant}=(U,AT,D,F,G,\Omega)$是不协调模糊序决策信息系统，$A\subseteq AT$，则 A 是分布协约简，当且仅当 A 一定是最大分布约简。

命题 2　设$I_*^{\leqslant}=(U,AT,D,F,G,\Omega)$是不协调模糊序决策信息系统，$A\subseteq AT$，则 A 是分布协调集，当且仅当 $\forall x,y\in U$，$\mu_{AT}(x)\not<\mu_A(x)$时有$[y]_A^{\leqslant}\not\subset[x]_A^{\leqslant}$。

证明：“$\Rightarrow$”反证。

假设当$\mu_{AT}(x)\not<\mu_A(x)$时$[y]_A^{\leqslant}\not\subset[x]_A^{\leqslant}$不成立。故此时有$[y]_A^{\leqslant}\subseteq[x]_A^{\leqslant}$，由性质 2(3)知$\mu_A(x)\leqslant\mu_A(x)$，而 A 又是分布协调集，于是有$\mu_A(x)=\mu_{AT}(x)$，$\mu_A(y)=\mu_{AT}(y)$，故有$\mu_{AT}(y)\leqslant\mu_{AT}(x)$，矛盾。

因此，当$\mu_{AT}(x)\not<\mu_A(x)$时有$[y]_A^{\leqslant}\not\subset[x]_A^{\leqslant}$成立。

“$\Leftarrow$”要证 A 是分布协调集须证明对任意的 $x\in U$ 有$\mu_A(x)=\mu_{AT}(x)$。而由性质 2(1)知，只须证明$\mu_A(x)\leqslant\mu_{AT}(x)$即可。

当$\mu_A(x)=0$ 时显然成立。下面证明$\mu_A(x)\neq 0$ 时同样成立。

由条件知，对 $\forall x,y\in U$，若$\mu_{AT}(y)\not<\mu_{AT}(x)$时有$[y]_A^{\leqslant}\not\subset[x]_A^{\leqslant}$成立。因此对 $\forall x,y\in U$，若$[y]_A^{\leqslant}\subseteq[x]_A^{\leqslant}$成立，则$\mu_{AT}(y)\leqslant\mu_A(x)$成立。

另外，当$\frac{|D_i\cap[x]_A^{\leqslant}|}{|U|}\neq 0$时，便有$|D_i\cap[x]_A^{\leqslant}|\neq 0$，不妨设$y_i\in D_i\cap[x]_A^{\leqslant}$，则$y_i\in D_i$且$y_i\in[x]_A^{\leqslant}$，故由性质 2(4)知$[y_i]_A^{\leqslant}\subseteq[x]_A^{\leqslant}$成立。因此可得$\mu_A(y_i)\leqslant\mu_{AT}(x)$。又因为$y_i\in[y_i]_{AT}^{\leqslant}$，所以$y_i\in D_i\cap[y_i]_{AT}^{\leqslant}$，即有$|D_i\cap[x]_A^{\leqslant}|\leqslant|D_i\cap[y_i]_{AT}^{\leqslant}|$。

因此$\mu_A(x)\leqslant\mu_{AT}(y_i)$，故$\mu_A(x)\leqslant\mu_{AT}(x)$成立。

同理可得如下最大分布协调集的充要条件。

命题 3　设$I_*^{\leqslant}=(U,AT,D,F,G,\Omega)$是不协调带关注度的模糊序决策信息系统，$A\subseteq AT$，则 A 是最大分布协调集，当且仅当 $\forall x,y\in U$，当$\gamma_A(y)>\gamma_B(x)$时有$[y]_A^{\leqslant}\not\subset[x]_{AT}^{\leqslant}$。

证明：由命题 2 可得。

4　带关注度的模糊序决策数据集的分布约简方法

第 3 节中给出了带关注度的模糊序决策信息系统的分布函数以及最大分布函数的定义和判定定理，这是判断属性子集是否协调的理论基础。本节将通过定义分布辨识矩阵的方式给出获得分布约简的方法。

定义 10　已知$I_*^{\leqslant}=(U,AT,D,F,G,\Omega)$是一个带关注度的模糊序决策信息系统，记：

$$Dis^{\mu}_{\leqslant AT}(x_i,x_j)=\{a\in AT\mid(x_i,x_j)\notin R_a^{\leqslant}\}$$

则称$Dis^{\mu}_{\leqslant AT}(x_i,x_j)$是$I_*^{\leqslant}$中$x_i,x_j$关于带关注度的模糊优势关系$R_{AT}^{\leqslant}$的分布可辨识属性集。

记：

$$\boldsymbol{Dis}^{\mu}_{\leqslant AT}=(Dis^{\mu}_{\leqslant AT}(x_i,x_j))_{|U|\times|U|}$$

则称$\boldsymbol{Dis}^{\mu}_{\leqslant AT}$是$I_*^{\leqslant}$中$x_i,x_j$关于带关注度的模糊优势关系$R_{AT}^{\leqslant}$的分布可辨识矩阵。特别地，对于 $\forall x_i,y_j\in U$，有：

$$Dis^{\mu}_{\leqslant AT}(x_i,x_j)=\emptyset$$

$$Dis^{\mu}_{\leqslant AT}(x_i,x_j)\cap Dis^{\mu}_{\leqslant AT}(x_j,x_i)=\emptyset$$

命题 4　设$I_*^{\leqslant}=(U,AT,D,F,G,\Omega)$是一个带关注度的模糊序决策信息系统，$A\subseteq AT$，对象关于$R_{AT}^{\leqslant}$的分布可辨识矩阵属性集$Dis^{\mu}_{\leqslant AT}(x_i,x_j)$，则：

$$R_A^{\leqslant}=R_{AT}^{\leqslant}\Leftrightarrow(Dis^{\mu}_{\leqslant AT}(x_i,x_j)\neq\emptyset)[A\cap Dis^{\mu}_{\leqslant AT}(x_i,x_j)\neq\emptyset]$$

证明：(1)“$\Rightarrow$”由$R_A^{\leqslant}=\bigcap_{a\in A}R_a^{\leqslant}$，对于 $\forall a\in A$，如果$(x_i,x_j)\notin R_a^{\leqslant}$，则有$(x_i,x_j)\notin R_A^{\leqslant}$。设，由模糊序关系的定义，对 $\forall x_i\in U$，有$[x_i]_A^{\leqslant}=[x_i]_{AT}^{\leqslant}$。当$Dis^{\mu}_{\leqslant AT}(x_i,x_j)\neq\emptyset$ 时，$\exists a\in AT$ 使得$(x_i,x_j)\notin R_a^{\leqslant}$，由此可以得到$(x_i,x_j)\notin R_{AT}^{\leqslant}$，即有$(x_i,x_j)\notin R_A^{\leqslant}$。因而 $\exists a\in A$ 使得$(x_i,x_j)\notin R_a^{\leqslant}$，于是 $a\in Dis^{\mu}_{\leqslant A}(x_i,x_j)$，由 $A\subseteq AT$ 及分布可辨识公式定义得$Dis^{\mu}_{\leqslant A}(x_i,x_j)\subseteq Dis^{\mu}_{\leqslant AT}(x_i,x_j)$，因此 $a\in Dis^{\mu}_{\leqslant AT}(x_i,x_j)$。

(2)“$\Leftarrow$”对 $\forall(x_i,x_j)\notin R_{AT}^{\leqslant}$，$\exists a\in AT$ 满足$(x_i,x_j)\notin R_a^{\leqslant}$，那么$Dis^{\mu}_{\leqslant AT}(x_i,x_j)\neq\emptyset$。于是 $\exists a\in A$ 使得 $a\in Dis^{\mu}_{\leqslant A}(x_i,x_j)$，即 $a\in A$ 且$(x_i,x_j)\notin R_a^{\leqslant}$，得到$(x_i,x_j)\notin R_{AT}^{\leqslant}$，即有：

$$(x_i,x_j)\notin R_{AT}^{\leqslant}\Rightarrow(x_i,x_j)\notin R_A^{\leqslant}$$

其逆否命题为：

$$(x_i,x_j)\in R_A^{\leqslant}\Rightarrow(x_i,x_j)\in R_{AT}^{\leqslant}$$

由此得到$R_A^{\leqslant}\subseteq R_{AT}^{\leqslant}$。另一方面由 $A\subseteq AT$ 显然可得$R_A^{\leqslant}\supseteq R_{AT}^{\leqslant}$，即证得$R_A^{\leqslant}=R_{AT}^{\leqslant}$。

定义 11　设$I_*^{\leqslant}=(U,AT,D,F,G,\Omega)$是不协调带关注度的模糊序决策信息系统，辨识矩阵为$\boldsymbol{Dis}^{\mu}_{\leqslant AT}$。则称$M^{\mu}_{\leqslant AT}=\wedge\{\vee\{a\mid a\in Dis^{\mu}_{\leqslant AT}(x_i,x_j)\}\mid\forall x_i,x_j\in U\}$为该不协调带关注度的模糊序决策信息系统的分布可辨识公式。

命题 5　设$I_*^{\leqslant}=(U,AT,D,F,G,\Omega)$是不协调带关注度的模糊序决策信息系统，分布辨识公式$M^{\mu}_{\leqslant AT}$的极小析取范式为$M^{\mu}_{\leqslant\min}=\bigvee_{k=1}^{p}(\bigwedge_{s=1}^{q_k}a_s)$，若记$B_{\mu}^{k}=\{a_s,s=1,2,\cdots,q_k\}$，则$\{B_{\mu}^{k}$，

$k=1,2,\cdots p\}$是所有分布约简形式的集合。

证明:由命题4可得。

5 案例分析

表1列出了某学校的6名学生3门课程的成绩,成绩用优、良、中、差等不确定性的词语来评价。每门课程占总分的权重不相同,老师需要根据成绩权重对学生进行综合排名。学生集合为$U=\{x_1,x_2,\cdots,x_6\}$,课程成绩集合为$AT=\{a_1,a_2,a_3\}$,决策集合为$D=\{d\}$,关注度集合为$\Omega=\{\omega_1,\omega_2,\omega_3\}$。

表1 带关注度的模糊序决策信息系统

Table 1 Weighted fuzzy order decision information system

$U\times(AT,D)$	a_1 ($\omega_1=0.1$)	a_2 ($\omega_2=0.4$)	a_3 ($\omega_3=0.5$)	d
x_1	0.7	0.5	0.6	1
x_2	0.6	0.6	0.7	3
x_3	0.7	0.6	0.8	2
x_4	0.4	0.6	0.6	2
x_5	0.7	0.4	0.7	3
x_6	0.6	0.7	0.5	1

根据定义5定义的带关注度的模糊序决策信息系统的优势关系可得:

$[x_1]_{AT}^{\leqslant}=\{x_1,x_2,x_3,x_4,x_5,x_6\}$

$[x_2]_{AT}^{\leqslant}=\{x_2,x_3\}$

$[x_3]_{AT}^{\leqslant}=\{x_3\}$

$[x_4]_{AT}^{\leqslant}=\{x_2,x_3,x_4,x_5,x_6\}$

$[x_5]_{AT}^{\leqslant}=\{x_2,x_3,x_4,x_5,x_6\}$

$[x_6]_{AT}^{\leqslant}=\{x_2,x_3,x_6\}$

$[x_1]_d^{\leqslant}=[x_6]_d^{\leqslant}=\{x_1,x_6\}$

$[x_3]_d^{\leqslant}=[x_4]_d^{\leqslant}=\{x_1,x_3,x_4,x_6\}$

$[x_2]_d^{\leqslant}=[x_5]_d^{\leqslant}=\{x_1,x_2,x_3,x_4,x_5,x_6\}$

显然$R_A^{\leqslant}\not\subset R_d^{\leqslant}$,则该带关注度的模糊序决策信息系统是不协调的。

方法1 利用定义8和定义9求解。

记:

$D_1=[x_1]_d^{\leqslant}=[x_6]_d^{\leqslant}$

$D_2=[x_3]_d^{\leqslant}=[x_4]_d^{\leqslant}$

$D_3=[x_2]_d^{\leqslant}=[x_5]_d^{\leqslant}$

则有:

$$\mu_{AT}(x_1)=\left(\frac{1}{3},\frac{2}{3},1\right),\mu_{AT}(x_2)=\left(0,\frac{1}{6},\frac{1}{3}\right)$$

$$\mu_{AT}(x_3)=\left(0,\frac{1}{6},\frac{1}{6}\right),\mu_{AT}(x_4)=\left(\frac{1}{6},\frac{1}{2},\frac{5}{6}\right)$$

$$\mu_{AT}(x_5)=\left(\frac{1}{6},\frac{1}{2},\frac{5}{6}\right),\mu_{AT}(x_6)=\left(\frac{1}{6},\frac{1}{3},\frac{1}{2}\right)$$

$$\gamma_{AT}(x_1)=1,\gamma_{AT}(x_2)=\frac{1}{3},\gamma_{AT}(x_3)=\frac{1}{6}$$

$$\gamma_{AT}(x_4)=\frac{5}{6},\gamma_{AT}(x_5)=\frac{5}{6},\gamma_{AT}(x_6)=\frac{1}{2}$$

易验证不存在AT的任意真子集A,使得$\forall x\in U$有$[x]_A^{\leqslant}=[x]_{AT}^{\leqslant}$,因此$AT=\{a_1,a_2,a_3\}$是该带关注度的模糊序决策信息系统的分布协调集,也是最大分布协调集。

方法2 利用命题5求解。

根据表1的不协调带关注度的模糊序决策信息系统,可得到该系统的分布约简辨识矩阵,如表2所列。

表2 不协调带关注度的模糊序决策信息系统的分布辨识矩阵

Table 2 Distributed discernible matrix of inconsistent and weighted fuzzy order decision information system

$Dis_{\leqslant A}^{\mu}$	x_1	x_2	x_3	x_4	x_5	x_6
x_1	$\emptyset$	a_1	$a_1a_2a_3$	a_1	a_2	a_1a_3
x_2	a_2a_3	$\emptyset$	$\emptyset$	a_1a_3	a_2	a_3
x_3	a_2	a_1	$\emptyset$	a_1	a_2	a_1a_3
x_4	a_2	$\emptyset$	$\emptyset$	$\emptyset$	a_2	a_3
x_5	a_3	a_1	$\emptyset$	a_1a_3	$\emptyset$	a_1a_3
x_6	a_2	a_2	a_2	a_1a_2	a_2	$\emptyset$

故可得:

$$M_{\leqslant AT}^{\mu}=(a_1\vee a_2\vee a_3)\wedge(a_1\vee a_3)\wedge(a_2\vee a_3)\wedge(a_1\vee a_2)\wedge a_1\wedge a_2\wedge a_3$$

因此,$\{a_1,a_2,a_3\}$是该带关注度的模糊序决策信息系统的分布约简,也是其最大分布约简。

可见方法2求得的分布约简与方法1的结果一致。

综上所述,在对学生成绩进行综合评价时,每门课程的成绩都不可或缺。两种方法的结果相同,但是方法1相比方法2计算过程更为复杂,计算时间更长,因此分布辨识矩阵在计算分布约简时更有优势。

结束语 本文在模糊序决策信息系统中,通过对每个条件属性进行加权,建立了带关注度的模糊序决策信息系统,并定义了加权得分函数,进而提出带关注度的模糊序决策信息系统的两种分布约简方法。最后通过实例证明了两种方法的可行性,比较了两种方法的优劣。

参考文献

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

[2] ZADEH L A, GUPTA M M, RAGADE R K, et al. Fuzzy sets and information granularity[M]. Amsterdam, 1979.

[3] TABAKOV M, CHLOPOWIEC A, DLUBAK A. Classification with Fuzzification Optimization Combining Fuzzy Information Systems and Type-2 Fuzzy Inference[J]. Applied Sciences, 2021, 11(8): 3484.

[4] KANZAWA Y, MIYAMOTO S. Generalized Fuzzy c-Means Clustering and its Property of Fuzzy Classification Function[J]. Journal of Advanced Computational Intelligence and Intelligent Informatics, 2021, 25(1): 73-82.

[5] SANZ J A, BUSTINCE H. A wrapper methodology to learn interval-valued fuzzy rule-based classification systems[J]. Applied Soft Computing, 2021, 104(3): 107249.

[6] SHEN S, CUI J. Estimation and inference of the fuzzy linear regression model with L fuzzy observations[C]// Fifth International Joint Conference on Computational Sciences & Optimization. IEEE, 2012: 354-358.

[7] KHAMMAR A H, AREFI M, AKBARI M G. A general approach to fuzzy regression models based on different loss functions[J]. Soft Computing, 2021, 25: 1-15.

[8] SPILIOTISM, GARROTE L. Unit hydrograph identification based on fuzzy regression analysis[J]. Evolving Systems, 2021 (8): 1-22.

[9] CHEN Z, BAGHERINIA A, MINAEI-BIDGOLI B, et al. Fuzzy Clustering Ensemble Considering Cluster Dependability[J]. International Journal of Artificial Intelligence Tools, 2021, 30(2): 2150007.

[10] FERRAROM B. A class of two-mode clustering algorithms in a fuzzy setting[J]. Econometrics and Statistics,2020(18):63-78.

[11] GUPTA A,DATTA S,DAS S. Fuzzy Clustering to Identify Clusters at Different Levels of Fuzziness:An Evolutionary Multiobjective Optimization Approach[J]. IEEE Transactions on Cybernetics,2019,51(5):1-11.

[12] PAWLAK Z. Rough sets[J]. International Journal of Computer & Information Sciences,1982,11(5):341-356.

[13] XU W H,ZHANG W X. Distribution Reduction in Inconsistent Information Systems Based on Dominance Relations[J]. Fuzzy Systems and Mathematics,2007,21(4):124-131.

[14] MI J S,WU W Z,ZHANG W X. Comparative Studies of Knowledge Reductions in Inconsistent Systems[J]. Fuzzy Systems and Mathematics,2003,17(3):54-60.

[15] ZHANG W X,MI J S,WU W W. Knowledge Reductions in Inconsistent Information Systems[J]. Chinese Journal of Computers,2003,26(1):12-18.

[16] XU W H,ZHANG W X. Knowledge Reductions in Inconsistent Information Systems Based on Dominance Relations[J]. Computer Science,2006,33(2):182-184.

[17] CHEN D,DONG L,MI J. Incremental mechanism of attribute reduction based on discernible relations for dynamically increasing attribute[J]. Soft Computing,2020,24(1):321-332.

[18] DONG L,CHEN D. Incremental attribute reduction with rough set for dynamic datasets with simultaneously increasing samples and attributes[J]. International Journal of Machine Learning and Cybernetics,2020,11(5):1339-1355.

[19] DING W,LIN C T,WITOLD P. Multiple Relevant Feature Ensemble Selection Based on Multilayer Co-Evolutionary Consensus MapReduce[J]. IEEE Transactions on Cybernetics,2018,50(2):425-439.

[20] DING W,LIN C T,CAO Z. Shared Nearest-Neighbor Quantum Game-Based Attribute Reduction With Hierarchical Coevolutionary Spark and Its Application in Consistent Segmentation of Neonatal Cerebral Cortical Surfaces[J]. IEEE Transactions on Neural Networks and Learning Systems,2019,30(7):2013-2027.

[21] SANG B,CHEN H,YANG L,et al. Feature selection for dynamic interval-valued ordered data based on fuzzy dominance neighborhood rough set[J],Knowledge-Based System,2021,227,107223.

[22] SINGH S,SHREEVASTAVA S,SOM T,et al. A fuzzy similarity-based rough set approach for attribute selection in set-valued information systems[J]. Soft Computing,2020,24(6):4675-4691.

[23] SANG B,CHEN H,YANG L,et al. Incremental Feature Selection Using a Conditional Entropy Based on Fuzzy Dominance Neighborhood Rough Sets[J]. IEEE Transactions on Fuzzy Systems,2022,30(6):1683-1697.

[24] WEN S,BAO Q. Dominance-based rough set approach to incomplete ordered information systems[J]. Information Sciences:An International Journal,2016,346:106-129.

[25] DAI J H,ZHENG G J,HAN H F,et al. Probability Approach for Interval-valued Ordered Decision Systems in Dominance-based Fuzzy Rough Set Theory[J]. Journal of Intelligent & Fuzzy Systems:Applications in Engineering and Technology,2017,32(1):703-710.

[26] GUAN L. A heuristic algorithm of attribute reduction in incomplete ordered decision systems[J]. Journal of Intelligent & Fuzzy Systems,2019,36(4):3891-3901.

[27] DENG S,GUAN S,MIN L I,et al. Decomposition for a new kind of imprecise information system[J]. Frontiers of Computer Science(print),2018,12(2):376-395.

[28] SUN B,MA W,GONG Z. Dominance-based rough set theory over interval-valued information systems[J]. Expert Systems,2014,31:185-197.

XU Wei-hua,born in 1979,Ph.D,professor,Ph.D supervisor,is a member of CAAI. His main research interests include cognitive computing,data mining,machine learning and so on.