第 58 卷 第 2 期
2022 年 3 月

南京大学学报(自然科学)
JOURNAL OF NANJING UNIVERSITY
(NATURAL SCIENCE)

Vol. 58, No. 2
Mar.,2022

DOI:10.13232/j.cnki.jnju.2022.02.009

含属性加权模糊序决策信息系统的近似约简

徐伟华*,孔子默,陈曜琦
(西南大学人工智能学院,重庆,400715)

摘　要:为保证关键属性在属性约简时能够被保留,可对信息系统的属性进行加权,从而提高关键属性的影响力.基于此,在属性加权的模糊序决策信息系统中建立了上、下近似约简的模型,得到两种约简的判定定理,并且给出求解上、下近似约简的辨识矩阵以及约简方法.最后,通过实例验证了该约简方法的有效性.

关键词:辨识矩阵,模糊序决策信息系统,属性加权,近似约简

中图分类号:TP18　　　**文献标志码**:A

Approximate reduction of fuzzy ordered decision information system with attribute weighting

Xu Weihua*,Kong Zimo,Chen Yaoqi
(College of Artificial Intelligent,Southwest University,Chongqing,400715,China)

Abstract: In order to ensure that the key attributes can be retained during attribute reduction process, attributes of the information system can be weighted, so as to improve the influence of the key attributes. Based on this, this paper establishes the models of upper and lower approximate reduction in fuzzy order decision information system with attribute weighting, obtains the judgment theorems of two kinds of reduction, and gives the identification matrix and reduction method for solving the upper and lower approximate reduction. Finally, an example is given to verify the effectiveness of the reduction method.

Key words: identification matrix, fuzzy ordered decision information system, attribute weighting, approximate reduction

粗糙集理论是1982年由波兰科学院院士、华沙理工大学教授Pawlak提出的,用于处理不确定、不精确和模糊知识的软计算,其主要是利用已知的信息来近似描述不确定的、模糊的信息.在分类问题中,粗糙集理论利用特征来归纳二元关系,从而将样本区分开.研究知识约简正是利用不同二元关系下形成的类以对特征进行提取.目前,在数据挖掘[1-4]、模式识别[5-6]、人工智能以及智能信息处理[7-10]等领域有广泛的应用.

粗糙集理论经过多年的发展,已经形成了比较完善的理论体系.属性约简是粗糙集理论体系中非常重要的一个部分,要求在决策能力和知识库分类不变的情况下,尽可能地删除其中不相关或者不重要的属性,从而消除冗余的属性,在降低空间复杂度和时间复杂度的情况下没有丢失重要的信息[10].目前,已经有很多学者对属性约简进行了深入研究[11-17].

模糊集是1965年由美国学者Zadeh提出的,它将经典集合进行扩充、推广,引入元素隶属度这一概念,对生活中那些不确定的问题量化、建模以及研究.模糊集是研究不确定性问题的一个重要工具.在模糊集中,元素与集合的关系不再是简

基金项目:国家自然科学基金(61976245)
收稿日期:2021－10－22
*通讯联系人,E-mail:chxuwh@gmail.com

单的属于或不属于,而是利用一个在0到1之间的隶属度来表示和刻画的.

属性加权是根据个人的喜好以及对事物的认知,人为地给不同的事物添加权重,事物的重要程度是根据个人的认知与习惯而改变的.属性加权在生活中有重要的作用,据此可以对不同的事物进行分类,而将其加入模糊序信息系统则可以更好地量化生活中不确定的事物,并按照个人的偏好对其进行重要程度的排序.

现实生活中,很多不确定性问题的信息系统属于不协调的基于属性加权的模糊序决策信息系统.因此,本文首先定义模糊序关系,根据个人喜好对每个属性加权,并根据加权得分函数重新定义模糊序信息系统,引入决策属性,从而建立不协调属性加权的模糊序决策信息系统.在此基础上引入上、下近似函数,上、下近似协调集以及上、下近似约简,上、下近似约简辨识矩阵,得到近似约简的方法,并通过案例分析比较了两种约简方法的优劣性.

1 基于属性加权的模糊序决策信息系统

决策信息系统是既有条件属性也有决策属性的一种特殊的信息系统,其研究的主要问题是这两种属性之间的关系.为了方便理解,下面给出一些相关的基本概念.

定义 1[10] $I=\left(U, AT\cup DT, F, G\right)$是一个五元组,称为决策信息系统;$\left(U, AT, F\right)$是一个三元组,称为信息系统;$AT$被称为条件属性集,$DT$被称为目标属性集.其中,

U为对象集,$U=\left\{x_1, x_2, \cdots, x_n\right\}$;

AT为条件属性集,$AT=\left\{a_1, a_2, \cdots, a_m\right\}$;

DT为决策属性集,$DT=\left\{d_1, d_2, \cdots, d_p\right\}$;

F是U与AT的关系集,$F=\left\{f_k: U\rightarrow V_k, k\leqslant p\right\}$,其中,$V_k$是$a_k$的值域;

G是U与DT的关系集,$G=\left\{g_{k'}: U\rightarrow V_{k'}, k'\leqslant q\right\}$,其中,$V_{k'}$是$d_{k'}$的值域.

定义 2[10] 设$I=\left(U, AT\cup\{d\}, F, G\right)$为决策信息系统,对任意的$f\in F, g\in G, a\in AT, x\in U$都有:

$$f\left(x_i, a\right)=\mu_a\left(x_i\right), g\left(x, d\right)\in R\ (R\text{为实数集})$$

其中,$\mu_a: U\rightarrow[0,1]$并满足$0\leqslant\mu_a(x)\leqslant 1$. $\mu_a(x)$称为$x\in U$在条件属性a下的隶属度,记$f(a)=\left\{f(x,a)\middle|a\in AT\right\}$,称$f(a)$为$U$上的模糊集,$I_*=\left(U, AT\cup\{d\}, F, G\right)$为模糊决策信息系统.

定义 3 设$I=\left(U, AT\cup DT, F, G\right)$为一个模糊决策信息系统,$\forall x\in U, \forall a_i\in AT$,定义对象$x$对属性$\left(a_1, a_2, \cdots, a_n\right)$的加权等分函数为:

$$S(x)=\sum_{i=1}^{n}\omega_i\mu_{a_i}(x), i=1,2,\cdots,n$$

其中,$\mu_{a_i}(x)$表示对象x在条件属性a_i下的隶属度,ω_i表示在属性a_i下的权重,对于所有属性权重满足$\omega_1+\omega_2+\omega_3+\cdots+\omega_n=1\ (0\leqslant\omega_i\leqslant 1, i=1,2,\cdots,n)$.

ω_i为a_i的权重,在得分评价时,越看重某个属性,对应的ω_i越大.因此,需要根据实际需求,给出相应的权重.

以上可以得出在模糊序信息系统中对象集在条件属性下的得分情况,并根据得分的大小得出之间的大小关系.下面考虑这样的优势关系.

定义 4 称一个四元数组$I_*^{\leqslant}=\left(U, AT\cup\{d\}, F, G\right)$为模糊决策信息系统,对于$A\subseteq AT$,令:

$$R_A^{\leqslant}=\left\{\left(x_i,x_j\right)\in U\times U\middle|x_i\leqslant x_j\right\}=$$
$$\left\{\left(x_i,x_j\right)\in U\times U\middle|S\left(x_i\right)\leqslant S\left(x_j\right), \forall a\in A, f\in F\right\}$$
$$R_d^{\leqslant}=\left\{\left(x_i,x_j\right)\in U\times U\middle|g\left(x_i,d\right)\leqslant g\left(x_j,d\right), g\in G\right\}$$

$R_A^{\leqslant}, R_d^{\leqslant}$称为模糊序决策信息系统的优势关系,此时该决策信息系统称为基于优势关系下的模糊序决策信息系统.

记:

$$\left[x_i\right]_A^{\leqslant}=\left\{x_j\in U\middle|\left(x_i,x_j\right)\in R_A^{\leqslant}\right\}=$$
$$\left\{x_j\in U\middle|S\left(x_i\right)\leqslant S\left(x_j\right), \forall a\in A, f\in F\right\}$$
$$\left[x_i\right]_d^{\leqslant}=\left\{x_j\in U\middle|g\left(x_i,d\right)\leqslant g\left(x_j,d\right), \forall g\in G\right\}$$

定理 1 设$I_*^{\leqslant}=\left(U, AT\cup\{d\}, F, G\right)$为模糊序决策信息系统,$R_A^{\leqslant}, R_d^{\leqslant}$称为模糊序决策信息系统的优势关系,则以下命题成立:

(1)$R_A^{\leqslant}$满足自反性和传递性,未必满足对称性,因而不一定是等价关系.

(2)若$B_1 \subseteq B_2 \subseteq AT$,则$R_{AT}^{\leqslant} \subseteq R_{B_2}^{\leqslant} \subseteq R_{B_1}^{\leqslant}$.

(3)若 $B_1 \subseteq B_2 \subseteq AT$,则 $\left[x_i\right]_{AT}^{\leqslant} \subseteq \left[x_i\right]_{B_2}^{\leqslant} \subseteq \left[x_i\right]_{B_1}^{\leqslant}$.

(4)若$x_j \in \left[x_i\right]_A^{\leqslant}$,则$\left[x_j\right]_A^{\leqslant} \subseteq \left[x_i\right]_A^{\leqslant}$.

对于任意$X \subseteq U$,定义X在优势关系$R_A^{\leqslant}$下的下近似和上近似分别为:

$$\underline{R_A^{\leqslant}}(X) = \left\{x_i \in U : \left[x_i\right]_A^{\leqslant} \subseteq X\right\}$$

$$\overline{R_A^{\leqslant}}(X) = \left\{x_i \in U : \left[x_i\right]_A^{\leqslant} \cap X \neq \varnothing\right\}$$

根据上述定理,即可求出决策信息系统的上、下近似.后续章节将研究不协调模糊序信息系统的近似约简的具体基于辨识矩阵的求解方法.

2 不协调模糊序信息系统的近似约简

由于优势关系不同于等价关系,在对象集上不能形成划分,而是形成一个覆盖,因此,对于优势关系下的模糊序信息系统,不能采用Pawlak的近似空间下决策信息系统中的方法定义分布函数和最大分布函数.下面给出模糊序决策信息系统的分布函数和最大分布函数的定义.

定义5 设$I = \left(U, AT \cup \{d\}, F, G\right)$是模糊序决策信息系统,$R_A^{\leqslant}$,$R_d^{\leqslant}$分别为条件属性集$AT$和决策属性$d$生成的$U$上的优势关系,对于$A \subseteq AT$,$x \in U$,记:

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

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

$$\sigma_A^{\leqslant} = \left(\underline{R_A^{\leqslant}}\left(D_1\right), \underline{R_A^{\leqslant}}\left(D_2\right), \cdots, \underline{R_A^{\leqslant}}\left(D_r\right)\right)$$

$$\lambda_A^{\leqslant} = \left(\overline{R_A^{\leqslant}}\left(D_1\right), \overline{R_A^{\leqslant}}\left(D_2\right), \cdots, \overline{R_A^{\leqslant}}\left(D_r\right)\right)$$

称$\sigma_A^{\leqslant}$与$\lambda_A^{\leqslant}$分别为U上关于属性子集A的下近似函数与上近似函数.

定义6 设$I = \left(U, AT \cup \{d\}, F, G\right)$是模糊序决策信息系统,$A \subseteq AT$:

若$\sigma_A^{\leqslant} = \sigma_{AT}^{\leqslant}$,则称$A$是下近似协调集;进一步,若$A$是下近似协调集,且$A$的任意真子集都不是下近似协调集,则称$A$是下近似约简.

若$\lambda_A^{\leqslant} = \lambda_{AT}^{\leqslant}$,则称$A$是上近似协调集;进一步,若$A$是上近似协调集,且$A$的任意真子集都不是上近似协调集,则称$A$是上近似约简.

定理2 设$I = \left(U, AT \cup \{d\}, F, G\right)$为模糊序决策信息系统,$A \subseteq AT$,则:

(1)若属性子集A为下近似协调集,当且仅当对于$\forall D_i \in U/R_d^{\leqslant}$,都有$\underline{R_A^{\leqslant}}\left(D_i\right) = \underline{R_{AT}^{\leqslant}}\left(D_i\right)$时成立.

(2)若属性子集A为上近似协调集,当且仅当对于$\forall D_i \in U/R_d^{\leqslant}$,都有$\overline{R_A^{\leqslant}}\left(D_i\right) = \overline{R_{AT}^{\leqslant}}\left(D_i\right)$时成立.

证明

(1)对于下近似协调集的情况证明如下.

充分性:利用反证法证明.

假设$\exists D_k \in U/R_d^{\leqslant}$,使得$\underline{R_A^{\leqslant}}\left(D_k\right) \neq \underline{R_{AT}^{\leqslant}}\left(D_i\right)$成立,那么,由前文可知$\underline{R_A^{\leqslant}}\left(D_k\right) \subseteq \underline{R_{AT}^{\leqslant}}\left(D_k\right)$,那么只少存在一个$\underline{R_{AT}^{\leqslant}}\left(D_k\right)$,使得$\underline{R_{AT}^{\leqslant}}\left(D_k\right) \in \sigma_{AT}^{\leqslant}$且$\underline{R_{AT}^{\leqslant}}\left(D_k\right) \notin \sigma_A^{\leqslant}$,同样也至少存在一个$\underline{R_A^{\leqslant}}\left(D_k\right)$,使得$\underline{R_A^{\leqslant}}\left(D_k\right) \in \sigma_A^{\leqslant}$且$\underline{R_A^{\leqslant}}\left(D_k\right) \notin \sigma_{AT}^{\leqslant}$,那么$\sigma_A^{\leqslant} \neq \sigma_{AT}^{\leqslant}$.然而,属性子集$A$为下近似协调集,推导结论$\sigma_A^{\leqslant} \neq \sigma_{AT}^{\leqslant}$与定义6相矛盾.所以,充分性即证明.

必要性:同样使用反证法证明.

假设属性子集A不是下近似协调集,那么$\sigma_A^{\leqslant} \neq \sigma_{AT}^{\leqslant}$,然而,对于属性子集$A$,$\forall D_i \in U/R_d^{\leqslant}$,都有$\underline{R_A^{\leqslant}}\left(D_i\right) = \underline{R_{AT}^{\leqslant}}\left(D_i\right)$,显然$\sigma_A^{\leqslant} = \sigma_{AT}^{\leqslant}$,矛盾.所以必要性即证明.

(2)与(1)同理可证.

表1给出了一个模糊序决策信息系统.

计算可知$R_{AT}^{\leqslant} \subseteq R_d^{\leqslant}$是不成立的,因此,该模

表1 一个模糊序决策信息系统

Table 1 A fuzzy ordered decision information system

U	$a_1(\omega_1 = 0.1)$	$a_2(\omega_2 = 0.4)$	$a_3(\omega_3 = 0.5)$	d
x_1	0.7	0.3	0.9	3
x_2	0.5	0.1	0.1	1
x_3	0.5	0.3	0	2
x_4	0.5	0.2	0.4	5
x_5	0.7	0.4	0.5	4

糊序信息系统是不协调的．对属性$\left(a_1,a_2,a_3\right)$加权$\left(\omega_1=0.1,\omega_2=0.4,\omega_3=0.5\right)$. 记：

$$D_1=\left[x_1\right]_d^{\leqslant}=\left\{x_1,x_4,x_5\right\}$$
$$D_2=\left[x_2\right]_d^{\leqslant}=\left\{x_1,x_2,x_3,x_4,x_5\right\}$$
$$D_3=\left[x_3\right]_d^{\leqslant}=\left\{x_1,x_3,x_4,x_5\right\}$$
$$D_4=\left[x_4\right]_d^{\leqslant}=\left\{x_4\right\}$$
$$D_5=\left[x_5\right]_d^{\leqslant}=\left\{x_4,x_5\right\}$$

有：

$$S_{x_1}=0.64,S_{x_2}=0.14,S_{x_3}=0.17,$$
$$S_{x_4}=0.33,S_{x_5}=0.48$$

又：

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

计算可得：

$$\underline{R_{AT}^{\leqslant}}\left(D_1\right)=\left\{x_1,x_4,x_5\right\}$$
$$\underline{R_{AT}^{\leqslant}}\left(D_2\right)=\left\{x_1,x_2,x_3,x_4,x_5\right\}$$
$$\underline{R_{AT}^{\leqslant}}\left(D_3\right)=\left\{x_1,x_3,x_4,x_5\right\}$$
$$\underline{R_{AT}^{\leqslant}}\left(D_4\right)=\varnothing$$
$$\underline{R_{AT}^{\leqslant}}\left(D_5\right)=\varnothing$$
$$\overline{R_{AT}^{\leqslant}}\left(D_1\right)=\left\{x_1,x_2,x_3,x_4,x_5\right\}$$
$$\overline{R_{AT}^{\leqslant}}\left(D_2\right)=\left\{x_1,x_2,x_3,x_4,x_5\right\}$$
$$\overline{R_{AT}^{\leqslant}}\left(D_3\right)=\left\{x_1,x_2,x_3,x_4,x_5\right\}$$
$$\overline{R_{AT}^{\leqslant}}\left(D_4\right)=\left\{x_2,x_3,x_4\right\}$$
$$\overline{R_{AT}^{\leqslant}}\left(D_5\right)=\left\{x_2,x_3,x_4,x_5\right\}$$

若取$A=\left\{a_2,a_3\right\}$,则：

$$S'_{x_1}=0.57,S'_{x_2}=0.09,S'_{x_3}=0.12,$$
$$S'_{x_4}=0.28,S'_{x_5}=0.41$$
$$\underline{R_A^{\leqslant}}\left(D_1\right)=\left\{x_1,x_4,x_5\right\}=\underline{R_{AT}^{\leqslant}}\left(D_1\right)$$
$$\underline{R_A^{\leqslant}}\left(D_2\right)=\left\{x_1,x_3,x_4,x_5\right\}=\underline{R_{AT}^{\leqslant}}\left(D_2\right)$$
$$\underline{R_A^{\leqslant}}\left(D_3\right)=\left\{x_1,x_2,x_3,x_4,x_5\right\}=\underline{R_{AT}^{\leqslant}}\left(D_3\right)$$
$$\underline{R_A^{\leqslant}}\left(D_4\right)=\varnothing=\underline{R_{AT}^{\leqslant}}\left(D_4\right)$$
$$\underline{R_A^{\leqslant}}\left(D_5\right)=\varnothing=\underline{R_{AT}^{\leqslant}}\left(D_5\right)$$

即$\sigma_A^{\leqslant}=\sigma_{AT}^{\leqslant}$,$A=\left\{a_2,a_3\right\}$是下近似协调集．验证可知,$A$的任意真子集都不是下近似协调集,所以,$A$是下近似约简．

同理可得,当$A=\left\{a_3\right\}$时$\lambda_A^{\leqslant}=\lambda_{AT}^{\leqslant}$,所以,$A=\left\{a_3\right\}$是上近似约简．

下面给出模糊序信息系统的上近似约简与下近似约简的判定定理．

定理 3 设$I=\left(U,AT\cup\{d\},F,G\right)$是模糊序信息系统,$A\subseteq AT$,则：

(1)A是下近似协调集,当且仅当对任意$D_i\in U/R_d^{\leqslant}$, 当 $x_i\in\underline{R_{AT}^{\leqslant}}\left(D_i\right),x_j\notin\underline{R_{AT}^{\leqslant}}\left(D_i\right)$ 时 $S\left(x_i\right)<S(x_j)$成立．

(2)A是上近似协调集,当且仅当对任意$D_i\in U/R_d^{\leqslant}$, 当 $x_i\in\overline{R_{AT}^{\leqslant}}\left(D_i\right),x_j\notin\overline{R_{AT}^{\leqslant}}\left(D_i\right)$ 时 $S\left(x_i\right)<S(x_j)$成立．

证明

(1) 必要性：假设存在$D_i\in U/R_d^{\leqslant}$,当$x_i\in\underline{R_{AT}^{\leqslant}}\left(D_i\right),x_j\notin\underline{R_{AT}^{\leqslant}}\left(D_i\right)$时，都有$S\left(x_i\right)<S\left(x_j\right)$. 由优势类的定义可知,$x_j\in\left[x_i\right]_A^{\leqslant}$,而$A$是下近似协调集,所以对于任意的$D_i\in U/R_d^{\leqslant}$,有$\underline{R_A^{\leqslant}}\left(D_i\right)=\underline{R_{AT}^{\leqslant}}\left(D_i\right)$,故必要性得证．

充分性：若A不是下近似协调集,则必然存在$D_i\in U/R_d^{\leqslant}$,使得$\underline{R_A^{\leqslant}}\left(D_i\right)\neq\underline{R_{AT}^{\leqslant}}\left(D_i\right)$,即存在$x\in\underline{R_{AT}^{\leqslant}}\left(D_i\right)$,但$x\notin\underline{R_A^{\leqslant}}\left(D_i\right)$,故有$[x]_{AT}^{\leqslant}\subseteq D_i$,但$[x]_A^{\leqslant}\not\subset D_i$. 而$[x]_{AT}^{\leqslant}\subseteq[x]_A^{\leqslant}$,故存在$x_0\in[x]_A^{\leqslant}$,但$x_0\notin D_i$,从而$x_0\notin\underline{R_{AT}^{\leqslant}}\left(D_i\right)$. 所以有$x\in\underline{R_C^{\leqslant}}\left(D_i\right)$,$x_0\notin\underline{R_{AT}^{\leqslant}}\left(D_i\right)$,故存在$a_k\in A$,使得$S\left(x_i\right)>S\left(x_j\right)$,与$x_0\in[x]_A^{\leqslant}$矛盾．故充分性得证．

(2)方法同(1).

3 上、下近似约简的辨识矩阵及约简方法

前文已经介绍了上近似约简和下近似约简的定义以及判定定理．然而,直接利用定义求解上、下近似的过程过于复杂,现给出辨识矩阵的定义,利用辨识矩阵求出上近似约简和下近似约简．

定义7 设$I=\left(U, AT\cup\{d\}, F, G\right)$是不协调模糊序信息系统,记$D_f=\left\{\left(x_i, x_j\right): x_i\in \underline{R_{AT}^{\leqslant}}\left(D_i\right), x_j\notin \underline{R_{AT}^{\leqslant}}\left(D_i\right), D_i\in R_d^{\leqslant}\right\}$,分别定义上近似辨识矩阵和下近似辨识矩阵:

将属性按照权重ω从大到小的顺序排列,比较$F_{i,n}=\sum_{k=1}^{n}\omega_k\mu_{a_k}\left(x_i\right),(n=1,2,\cdots,m)$与$F_{j,n}=\sum_{k=1}^{n}\omega_k\mu_{a_k}\left(x_j\right)(n=1,2,\cdots,m)$的大小.如果$F_{i,n}<F_{j,n}$,则记录下此时的$a_n$,否则,记为$\varnothing$.

$$D_\lambda\left\{x_i, x_j\right\}=\begin{cases}\left\{a_n\in AT: F_{i,n}<F_{j,n}\right\}, & \left(x_i, x_j\right)\in D_f\\ \varnothing, & \left(x_i, x_j\right)\notin D_f\end{cases}$$

使用与刚才求$D_\lambda\left\{x_i, x_j\right\}$相同的方法,在最后,记录下第一个使得$F_{i,n}>F_{j,n}$的$a_n$.

$$D_\sigma\left\{x_i, x_j\right\}=\begin{cases}\left\{a_n\in AT: F_{i,n}>F_{j,n}\right\} & \left(x_i, x_j\right)\in D_f\\ \varnothing, & \left(x_i, x_j\right)\notin D_f\end{cases}$$

由D_λ和D_σ形成的矩阵$M_\lambda=\left(D_\lambda\left(x_i, x_j\right), x_i, x_j\in U\right)$和$M_\sigma=\left(D_\sigma\left(x_i, x_j\right), x_i, x_j\in U\right)$分别称为上近似辨识矩阵和下近似辨识矩阵.

定理4 设$I=\left(U, AT\cup\{d\}, F, G\right)$是不协调模糊序信息系统,$A\subseteq AT$,则:

(1) A是上近似协调集,当且仅当对$\forall\left(x_i, x_j\right)\in D_f$,有$A\cap D_\lambda\left(x_i, x_j\right)\neq\varnothing$.

(2) A是下近似协调集,当且仅当对$\forall\left(x_i, x_j\right)\in D_f$,有$A\cap D_\lambda\left(x_i, x_j\right)\neq\varnothing$.

证明

(1) 必要性:对于$\forall\left(x_i, x_j\right)\in D_f$,存在$D_i\in U/R_d^{\leqslant}$,使得$x_i\in\underline{R_{AT}^{\leqslant}}\left(D_i\right)$, $x_j\notin\underline{R_{AT}^{\leqslant}}\left(D_i\right)$.由定理1可知,一定存在$a_k\in A$,使得$F\left(x_i\right)<F\left(x_j\right)$,从而有$a_k\in D_\lambda\left(x_i, x_j\right)$.因此,若$A$是上近似协调集,则$\forall\left(x_i, x_j\right)\in D_f$,有$A\cap D_\lambda\left(x_i, x_j\right)\neq\varnothing$,得证.

充分性:若对于$\forall\left(x_i, x_j\right)\in D_f$,都有$A\cap D_\lambda\left(x_i, x_j\right)\neq\varnothing$,则至少存在$a_k\in A$,使得$a_k\in D_\lambda\left(x_i, x_j\right)$,故有$F\left(x_i\right)<F\left(x_j\right)$.而$x_i\in\underline{R_{AT}^{\leqslant}}\left(D_i\right)$, $x_j\notin\underline{R_{AT}^{\leqslant}}\left(D_i\right)$,由定理1可知,$A$是上近似协调集,得证.

(2)证明过程类似(1).

定义8 设$I=\left(U, AT\cup\{d\}, F, G\right)$是不协调模糊序信息系统,$M_\lambda$与$M_\sigma$为其上近似辨识矩阵和下近似辨识矩阵,称:

$$G_\lambda=\wedge\left\{\vee\left\{a_k: a_k\in D_\lambda\left(x_i, x_j\right)\right\}, x_i, x_j\in D_f\right\}$$

$$G_\sigma=\wedge\left\{\vee\left\{a_k: a_k\in D_\sigma\left(x_i, x_j\right)\right\}, x_i, x_j\in D_f\right\}$$

分别为该模糊序信息系统的上近似辨识函数和下近似辨识函数.

定理5 设$I=\left(U, AT\cup\{d\}, F, G\right)$是不协调模糊序信息系统,则有:

上近似辨识函数G_λ的极小析取范式为$G_\lambda=\bigvee_{k=1}^{p}\left(\bigwedge_{s=1}^{q_k}a_s\right)$.若记$B_\lambda^k=\left\{a_s, s=1,2,\cdots,q_k\right\}$,则$\left\{B_\lambda^k, k=1,2,\cdots,p\right\}$是所有上近似约简形式的集合.

下近似辨识函数G_σ的极小析取范式为$G_\sigma=\bigvee_{k=1}^{p}\left(\bigwedge_{s=1}^{q_k}a_s\right)$.若记$B_\sigma^k=\left\{a_s, s=1,2,\cdots,q_k\right\}$,则$\left\{B_\lambda^k, k=1,2,\cdots,p\right\}$是所有下近似约简形式的集合.

证明 对于$\forall\left(x_i, x_j\right)\in D_f$,由极小析取范式的定义可知:

$$B_\lambda^k\cap D_\lambda\left(x_i, x_j\right)\neq\varnothing$$

再由前面的定理可知B_λ^k是上近似协调集.由$G_\lambda=\bigvee_{k=1}^{p}\left(B_\lambda^k\right)$在$B_\lambda^k$中去掉一个元素形成$B_\lambda^{k'}$,必然存在$\left(x_i, x_j\right)\in D_f$,使得$B_\lambda^{k'}\cap D_\lambda\left(x_i, x_j\right)=\varnothing$,故$B_\lambda^{k'}$不是上近似协调集,所以$B_\lambda^k$是上近似约简.而上近似识别函数中包含了所有的$D_\lambda\left(x_i, x_j\right)$.因此,不可能存在其他的上近似约简了.

例1 利用辨识矩阵求表1中属性加权的模糊序信息系统的上近似约简和下近似约简.

上近似辨识矩阵如表2所示.

可以计算得到:

$$G_\lambda=a_3$$

所以上近似约简为$\left\{a_3\right\}$,该结果与第2节的计算结果一致.

下近似辨识矩阵如表3所示.

可以计算得到:

表 2　表 1 的上近似辨识矩阵

Table 2　The upper approximation identification matrix of Table 1

	x_1	x_2	x_3	x_4	x_5
x_1	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
x_2	$\varnothing$	$\varnothing$	a_3	$\varnothing$	$\varnothing$
x_3	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
x_4	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
x_5	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

表 3　表 1 的下近似辨识矩阵

Table 3　The lower approximation identification matrix of Table 1

	x_1	x_2	x_3	x_4	x_5
x_1	$\varnothing$	a_3	a_3	$\varnothing$	$\varnothing$
x_2	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$
x_3	$\varnothing$	a_2	$\varnothing$	$\varnothing$	$\varnothing$
x_4	$\varnothing$	a_3	a_3	$\varnothing$	$\varnothing$
x_5	$\varnothing$	a_3	a_3	$\varnothing$	$\varnothing$

$G_\sigma = a_2 \wedge a_3$

因此，下近似约简为$\left\{a_2, a_3\right\}$，该结果与第 2 节的计算结果一致．

与模糊序信息系统的近似约简相比，同样考虑其上近似约简和下近似约简问题，并研究了两种约简的判定定理和辨识矩阵的计算方法．然而，与模糊序信息系统的近似约简不同的是，本文对属性集增加了权重并对其求和，争取将权值大的属性保留，同时尽力删除权值小的属性．

4　算法与实验

本节利用上述的含属性加权的模糊序信息系统的近似约简的辨识矩阵方法，进行算法设计，下面分别写出上、下近似约简的伪代码，并且给出详细的数值实验．

首先给出上近似约简的具体算法．

算法 1　含属性加权的模糊序信息系统的上近似约简算法

Input：$I = \left(U, C \cup \{d\}, F, G, W\right)$

Output：含属性加权的模糊序信息系统的上近似约简集 Red

1\. Selected feature subset $Red \leftarrow \varnothing$;

2\. Count the number of score function $S(x) = \sum_{i=1}^{n} \omega_i \mu_{a_i}(x), i = 1, 2, \cdots, n$;//计算出得分函数//

3\. for $i=1$ to $|U|$ do:

4\. $\quad dec = \varnothing$

5\. $\quad$ for $j=1$ to $|U|$ do:

6\. $\quad\quad$ if $S_{x_i} \leqslant S_{x_j}$ then do:

7\. $\quad\quad\quad dec \leftarrow x_j$

8\. $\quad\quad$ end if

9\. $\quad$ end for

10\. $\quad Dec = Dec \cup dec$

11\. end for

12\. for $i=1$ to $|U|$ do:

13\. $d = \varnothing$

14\. $\quad$ for $j=1$ to $|U|$ do:

15\. $\quad\quad$ if $d_i \leqslant d_j$ then do:

16\. $\quad\quad\quad d \leftarrow d_j$

17\. $\quad\quad$ end if

18\. $\quad$ end for

19\. $\quad D = D \cup d$

20\. end for

21\. for $i=1$ to $|U|$ do:

22\. $\quad \overline{R_C^{\leqslant}}(i) = \varnothing$

23\. $\quad$ for $j=1$ to $|U|$ do:

24\. $\quad\quad$ if $Dec[i] \cap d[j] \neq \varnothing$ then do:

25\. $\quad\quad\quad \overline{R_C^{\leqslant}}(i) \leftarrow x_i$

26\. $\quad\quad$ end if

27\. $\quad$ end for

28\. $\quad \overline{R_C^{\leqslant}} = \overline{R_C^{\leqslant}} \cup \overline{R_C^{\leqslant}}(i)$

29\. end for

30\. for $k=1$ to $|U|$ do:

31\. $\quad D_f = \varnothing$

32\. $\quad$ for $i=1$ to $|U|$ do:

33\. $\quad\quad$ for $j=1$ to $|U|$ do:

34\. $\quad\quad\quad$ if $x_i \in \overline{R_C^{\leqslant}}(k)$ and $x_j \notin \overline{R_C^{\leqslant}}(k)$ then do:

35\. $\quad\quad\quad\quad x_i = x_j = 0$

36\. $\quad\quad\quad\quad$ for $m = 1$ to $|C|$:

37\. $\quad\quad\quad\quad\quad x_i = x_i + \omega_{i_m}\mu_{i_m}, x_j = x_j + \omega_{j_m}\mu_{j_m}$

38\. $\quad\quad\quad\quad\quad$ if $x_i < x_j$ then do:

39\. $\quad\quad\quad\quad\quad\quad D_f \leftarrow a_m$

40\. $\quad\quad\quad\quad\quad$ end if

41\. $\quad\quad\quad\quad$ end for

42\. $\quad\quad\quad$ end if

43\. $\quad\quad$ end for

44.　　end for
45.　　$M_\lambda = M_\lambda \cup D_f$
46. end for
47. *Red* is the Minimal disjunctive normal form of M_λ
48. end

下面给出算法1的时间复杂度分析：第2步要求出每个对象对应的得分函数，所以时间复杂度为$O\left(|U|^2\right)$；第3步至第11步需要求优势类，时间复杂度为$O\left(|U|^2\right)$；第12步至第20步为求决策类，时间复杂度为$O\left(|U|^2\right)$；第21步至第29步求上近似，时间复杂度为$O\left(|U|^2\right)$；第30步至第46步求辨识矩阵，时间复杂度为$O\left(|U|^3|C|\right)$；第47步为根据辨识矩阵求约简结果，时间复杂度为$O\left(|U|^2\right)$. 因此，算法1总的时间复杂度为$O\left(|U|^2+|U|^2+|U|^2+|U|^3|C|+|U|^2\right)$.

下面给出下近似约简的具体算法.

算法2　含属性加权的模糊序信息系统的下近似约简算法

Input：$I=\left(U, C\cup\{d\}, F, G, W\right)$

Output：含属性加权的模糊序信息系统的下近似约简集 *Red*

1. 与上近似约简的1～29步相同
2. for k to $|U|$ then do：
3.　$D_f = \varnothing$
4.　for i to $|U|$ then do：
5.　　for j to $|U|$ then do：
6.　　　if $x_i \in \underline{R_C^{\leqslant}}(k)$ and $x_j \notin \underline{R_C^{\leqslant}}(k)$ then do：
7.　　　　$x_i = x_j = 0$
8.　　　　for m=1 to $|C|$：
9.　　　　　$x_i = x_i + \omega_{i_m}\mu_{i_m}, x_j = x_j + \omega_{j_m}\mu_{j_m}$
10.　　　　　if $x_i > x_j$ then do：
11.　　　　　　$D_f \leftarrow a_m$
12.　　　　　end if
13.　　　　end for
14.　　　end if
15.　　end for
16.　end for
17.　$M_\sigma = M_\sigma \cup D_f$
18. end for
19. *Red* is the Minimal disjunctive normal form of M_σ
20. end

与上近似约简相同，其时间复杂度为$O\left(|U|^2+|U|^2+|U|^2+|U|^3|C|+|U|^2\right)$.

根据系列实验来分别验证上近似约简算法和下近似约简算法的有效性. 使用的计算机配置如下：Intel（R）Core（TM）i7-9750H CPU @ 2.60 GHz，内存16 GB，操作系统为64位Windows 10，实现程序采用Jupyter平台. 实验中使用的数据集信息如表4所示.

表4　实验数据集总览

Table 4　The experimental datasets

No.	数据集	样本数	特征数	分类数
1	Caesarian Section Classification Dataset	79	4	2
2	iris	150	4	3
3	wine	178	13	3
4	Connectionist Bench	208	60	2
5	seed	210	7	3
6	Blood Transfusion Service Center	748	4	2
7	audit_risk	776	26	2
8	banknote authentication	1372	3	2

根据算法1求出表4数据集的上近似约简，并分别使用KNN和SVM分类器对约简结果进行分类，求出其分类的精度，如表5所示.

根据算法2求出表4数据集的下近似约简，并分别使用KNN和SVM分类器对约简结果进行分类，求出其分类的精度，如表6所示.

通过表5和表6的结果可以看出，利用上述算法1、算法2所求得的上近似约简和下近似约简进行分类，不论是采用SVM算法分类还是KNN算法分类，分类精度都可以尽量保持在一个比较高的数值. 其中，在上近似约简中，SVM算法和KNN算法所求得的平均精度分别为84.27%和81.08%；在下近似约简中，SVM算法和KNN算法所求得的平均精度则分别为87.51%和73.01%，所有的结果都具有一定的可信度.

表 5 在 SVM 与 KNN 下所得上近似约简的分类精度

Table 5 Classification accuracy of upper approximate reduction under SVM and KNN

数据集	SVM	KNN
Caesarian Section Classification Dataset	61.90%±5.83%	68.75%±4.93%
iris	99.17%±0.55%	96.67%±0.72%
wine	95.71%±0.63%	97.14%±0.47%
Connectionist Bench	84.85%±3.26%	76.16%±2.96%
seed	71.86%±4.38%	61.90%±6.47%
Blood Transfusion Service Center	76.76%±4.72%	75.33%±4.37%
audit_risk	86.13%±2.36%	81.41%±2.74%
banknote authentication	97.81%±1.51%	91.27%±1.79%
平均	84.27%±2.91%	81.08%±3.06%

表 6 在 SVM 与 KNN 下所得下近似约简的分类精度

Table 6 Classification accuracy of upper approximate reduction under SVM and KNN

数据集	SVM	KNN
Caesarian Section Classification Dataset	74.60%±5.47%	43.75%±7.46%
iris	74.17%±4.68%	63.33%±4.37%
wine	91.43%±3.62%	65.71%±4.73%
Connectionist Bench	88.79%±4.38%	69.52%±6.24%
seed	99.04%±1.49%	85.71%±2.84%
Blood Transfusion Service Center	78.92%±5.37%	77.33%±5.22%
audit_risk	95.81%±1.74%	87.82%±2.68%
banknote authentication	97.35%±1.37%	90.91%±2.03%
平均	87.51%±2.97%	73.01%±4.45%

5 结论

本文首先介绍了模糊序信息系统,进而对模糊序信息系统进行加权,定义了带有属性加权的模糊序信息系统,并且在该信息系统中引入了近似约简的概念,通过研究得到了近似约简的判定定理和上、下近似约简的可辨识矩阵,建立了在属性加权的模糊序关系下不协调决策信息系统的近似约简方法,同时通过实例证明了该约简方法的有效性.

参考文献

[1] Taj S M, Sudha M, Kumaravel A. Predicting heart failure using data mining with rough set theory and Fuzzy Petri Net. Journal of Physics: Conference Series, 2021(1724): 012033.

[2] Wang G Q, Li T R, Zhang P F, et al. Double-local rough sets for efficient data mining. Information Sciences, 2021(571): 475－498.

[3] Wang Y F. Mining stock price using fuzzy rough set system. Expert Systems with Applications, 2003, 24(1): 13－23.

[4] Ji W T, Pang Y, Jia X Y, et al. Fuzzy rough sets and fuzzy rough neural networks for feature selection: A review. WIREs Data Mining and Knowledge Discovery, 2021, 11(3): e1402.

[5] 高建来,运士伟,张永胜. 融合粗糙集与球形支持向量机的多分类识别. 河南科技大学学报:自然科学版, 2011, 32(5): 77－80. (Gao J L, Yun S W, Zhang Y S. Multi-classification pattern recognition based on fusion of rough set and sphere-structured support vector machines. Journal of Henan University of Science and Technology: Natural Science, 2011, 32(5): 77－80.)

[6] 王江荣,黄建华,罗资琴,等. 基于粗糙集的Logistic回归模型在矿井突水模式识别中的应用. 煤田地质与勘探, 2015, 43(6): 70－74. (Wang J R, Huang J H, Luo Z Q, et al. Application of logistic regression model based on rough set in recognition of mine water inrush pattern. Coal Geology & Exploration, 2015, 43(6): 70－74.)

[7] Csajbók Z. Partial approximative set theory: A generalization of the rough set theory//2010 International Conference of Soft Computing and Pattern Recognition. Cergy-Pontoise, France: IEEE, 2010: 51－56.

[8] Zhou H, Wang Z, Shi C Q, et al. Study of fault diagnosis distribution network based on rough set and artificial intelligence. Journal of Physics: Conference Series, 2021(1754): 012204.

[9] Trabelsi S, Elouedi Z, Lingras P. Belief rough set classifier//Proceedings of the 22nd Canadian Conference on Artificial Intelligence: Advances in Artificial Intelligence. Springer Berlin Heidelberg, 2009: 257－261.

[10] 徐伟华.序信息系统与粗糙集.北京:科学出版社,2013.

[11] 袁修久,张文修.模糊目标信息系统的属性约简.系统工程理论与实践,2004,24(5):116－120,125.(Yuan X J,Zhang W X. Attribute reductions in fuzzy inconsistent information systems. Systems Engineering-Theory & Practice,2004,24(5):116－120,125.)

[12] Jensen R,Shen Q. Fuzzy - rough attribute reduction with application to web categorization. Fuzzy Sets and Systems,2004,141(3):469－485.

[13] Jensen R,Shen Q. Fuzzy-rough data reduction with ant colony optimization. Fuzzy Sets and Systems,2005,149(1):5－20.

[14] Pawlak Z. Rough Sets: Theoretical aspects of reasoning about data. Norwell: Kluwer Academic Publishers,1992.

[15] Pawlak Z, Grzymala - Busse J, Slowinski R, et al. Rough sets. Communications of the ACM,1995,38(11):88－95.

[16] 张文修,米据生,吴伟志.不协调目标信息系统的知识约简.计算机学报,2003,26(1):12－18.(Zhang W X,Mi J S,Wu W Z. Knowledge reductions in inconsistent information systems. Chinese Journal of Computers,2003,26(1):12－18.)

[17] 王珏,苗夺谦,周育健.关于Rough Set理论与应用的综述.模式识别与人工智能,1996,9(4):337－344.(Wang J,Miao D Q,Zhou Y J. Rough set theory and its application: A Survey. PR & AI,1996,9(4):337－344.)

（责任编辑　杨可盛）