计算机研究与发展 ISSN 1000-1239/CN 11-1777/TP
Journal of Computer Research and Development 47(8): 1362-1371, 2010

区间值信息系统的知识约简

张 楠 苗夺谦 岳晓冬
(同济大学计算机科学与技术系 上海 201804)
(嵌入式系统与服务计算教育部重点实验室(同济大学) 上海 201804)
(国家高性能计算机工程技术研究中心同济分中心 上海 201804)
(zhangnan0851@163.com)

Approaches to Knowledge Reduction in Interval-Valued Information Systems

Zhang Nan, Miao Duoqian, Yue Xiaodong
(Department of Computer Science and Technology, Tongji University, Shanghai 201804)
(Key Laboratory of Embedded System and Service Computing (Tongji University), Ministry of Education, Shanghai 201804)
(Tongji Branch, National Engineering and Technology Center of High Performance Computers, Shanghai 201804)

Abstract Knowledge reduction is an important problem of rough set theory. The notion of a reduction plays an essential role in analyzing an information system. A reduction is a minimum subset of attributes that can provide the same descriptive or classification ability as the full set of available attributes. In other words, attributes in a reduction are jointly sufficient and individually necessary in an information system. Original knowledge reduction in Pawlak's rough set theory is mainly based on single-valued information systems. However, there always exist interval numbers in a lot of information systems in reality. Thus, knowledge reduction in interval-valued information systems has been a hot issue in the related area in recent years. Although some research works for knowledge reduction in interval-valued information systems have been done, there are still critical problems, such as redundancy of classification results and low accuracy of rough approximation to be dealt with. Aiming at the existing problems, the authors introduce the α-maximal consistent class into interval-valued information systems to improve the accuracy of knowledge classification and rough approximation. A novel measurement for the similarity degree of different interval numbers is proposed in this paper. Furthermore, to simplify knowledge representation, the authors present the methods for constructing the discernibility matrix and attribute reduction in interval-valued information systems or decision systems. Finally, the latent knowledge can be discovered by knowledge reduction.

Key words interval-valued information system; similarity rate; α-maximal consistent class; rough approximation; knowledge reduction

摘 要 知识约简是粗糙集理论的重要研究内容之一.传统的知识约简主要针对单值信息系统,但在许多实际问题中,信息系统中的数据往往以区间值的形式存在,因此,区间值信息系统的知识约简研究具有重要意义.现有工作中,论域的分类结果存在冗余度大、误分率高等问题.针对上述问题,在区间值信息系统中引入了α-极大相容类的概念,并提出了新的粗糙上下近似算子,α-极大相容类的采用有效地提高了分类和粗糙近似精度.最后,给出了区间值信息系统知识约简的定义和相应区分函数的计算方法,为区间值信息系统的知识获取提供了一条新的途径.

关键词 区间值信息系统;相似率;α-极大相容类;粗糙近似;知识约简

中图法分类号 TP18

收稿日期: 2009-05-11; 修回日期: 2009-11-12
基金项目: 国家自然科学基金项目(60475019, 60775036); 教育部博士学科点专项科研基金项目(20060247039)

0 引　言

粗糙集理论(rough set theory, RST)是波兰科学家Pawlak教授[1]提出的一种处理模糊、不确定性知识的数学工具, 是经典集合论的推广. 经过20多年的研究与发展, 粗糙集理论已在决策分析、故障诊断、模式识别、数据挖掘与人工智能等领域取得了巨大成功.

知识约简[3-10]是粗糙集理论的核心内容之一. 众所周知, 知识库中描述知识的属性并不是同等重要的, 甚至某些属性是冗余的. 这一方面造成了资源的浪费; 另一方面干扰人们作出正确的决策. 所谓知识约简, 就是保持知识库分类能力不变的条件下, 删除其中不相关或不重要的属性, 使知识的表示简化, 这正是人们所期望的[2,4-5].

目前, 许多学者对区间值信息系统的知识约简作了深入的研究, 并取得了很多成果[11-19]. 但是, 这些研究大多是以相似类[11-13]为基础的, 这种分类方法主要存在两个问题: 第一, 分类粒度粗、数目多、冗余度高; 第二, 同一类中的任意两个对象不一定具有相同的属性特征, 误分率高. 针对上述问题, 文章提出了α-极大相容类, 使同一类中的对象具有相同的属性特征, 从而细化了相似类, 提高了区间值信息系统的分类和粗糙近似精度. 本文还进一步给出了基于α-极大相容类的区间值信息系统知识约简的定义与计算方法, 为区间值信息系统的知识获取提供了一条有效途径.

1 基本概念和性质

本节中, 1.1, 1.2节定义了区间值信息系统和相似率等基本概念; 1.3, 1.4节首先分析了基于相似类的分类和粗糙近似存在的问题, 并在此基础上提出了基于α-极大相容类的分类和粗糙近似. 通过证明可得, 新方法提高了分类和粗糙近似精度.

1.1 区间值信息系统

区间值信息系统是属性值以区间数形式存在的信息系统. 特别地, 在区间值决策系统中条件属性值为区间值, 决策属性值为单值. 区间值信息系统的形式化定义如下:

定义 **1**[11]. 区间值信息系统为四元组 $\zeta=(U, AT, V, f)$, 其中 $U=\{u_1, u_2, \cdots, u_n\}$ 是非空有限论域; $AT=\{a_1, a_2, \cdots, a_m\}$ 是非空有限属性集; $V=\bigcup_{a_k\in AT} V_{\{a_k\}}$, $V_{\{a_k\}}$ 表示属性 $a_k\in AT$ 的值域; $f: U\times AT\rightarrow V$ 是一个信息函数, 它指定论域 U 中每一个对象 u_i 在属性 a_k 上的区间属性值, 即对任意的 $u_i\in U$, $a_k\in A$, 有 $f(u_i, a_k)=a_k(u_i)=[l_i^k, u_i^k]$.

区间值决策系统 $\zeta=(U, AT\cup D, V, f)$ 是一类特殊的区间值信息系统, 非空有限属性集 $AT\cup D$ 包括条件属性集 $AT=\{a_1, a_2, \cdots, a_m\}$ 和决策属性集 $D=\{d\}$ 两部分; $V=V_{AT}\cup V_D$, 其中 V_{AT} 为条件属性值集合, V_D 为决策属性值集合; $f: U\times AT\rightarrow V_{AT}$ 为区间值映射, $f: U\times D\rightarrow V_D$ 为单值映射.

一个区间值决策系统如表1所示, 其中论域 $U=\{u_1, u_2, \cdots, u_{10}\}$, 条件属性集 $AT=\{a_1, a_2, a_3, a_4, a_5\}$, 决策属性集 $D=\{d\}$; 条件属性值 $a_k(u_i)=[l_i^k, u_i^k]$ 是区间值, 如 $a_1(u_2)=[3.38, 4.50]$; 决策属性值 $d(u_i)$ 是单值, 如 $d(u_2)=2$.

Table 1　An Interval-Valued Decision System

表 **1**　区间值决策系统

U	a_1	a_2	a_3	a_4	a_5	d
u_1	[2.17, 2.96]	[5.32, 7.23]	[3.35, 5.59]	[3.21, 4.37]	[2.46, 3.59]	1
u_2	[3.38, 4.50]	[3.38, 5.29]	[1.48, 3.58]	[2.36, 3.52]	[1.29, 2.42]	2
u_3	[2.09, 2.89]	[7.03, 8.94]	[3.47, 5.69]	[3.31, 4.46]	[3.48, 4.61]	2
u_4	[3.39, 4.51]	[3.21, 5.12]	[0.68, 1.77]	[1.10, 2.26]	[0.51, 1.67]	3
u_5	[3.70, 4.82]	[2.98, 4.89]	[1.12, 3.21]	[2.07, 3.23]	[0.97, 2.10]	2
u_6	[4.53, 5.63]	[5.51, 7.42]	[3.50, 5.74]	[3.27, 4.43]	[2.49, 3.62]	2
u_7	[2.03, 2.84]	[5.72, 7.65]	[3.68, 5.91]	[3.47, 4.61]	[2.53, 3.71]	1
u_8	[3.06, 4.18]	[3.11, 5.02]	[1.26, 3.36]	[2.25, 3.41]	[1.13, 2.25]	3
u_9	[3.38, 4.50]	[3.27, 5.18]	[1.30, 3.40]	[4.21, 5.36]	[1.11, 2.23]	1
u_{10}	[1.11, 2.26]	[2.51, 3.61]	[0.76, 1.85]	[1.30, 2.46]	[0.42, 1.57]	4

1.2 相似率

经典粗糙集理论[1-2]采用等价关系对论域进行划分,同一等价类中的对象具有相同的属性值.然而,区间值信息系统中相同区间值形成的等价类很难对论域形成合理的划分[18-19].因此,在区间值信息系统中引入相似率来表示2个区间值的相似程度,为论域的分类提供了度量标准.

为了方便描述相似率的概念,首先介绍区间值的相关基本运算[17,20]:

1) 交运算(meet).

$$a_k(u_i)\cap a_k(u_j)=\begin{cases}[l_i^k,u_i^k], & [l_i^k,u_i^k]\subseteq[l_j^k,u_j^k];\\ [l_j^k,u_j^k], & [l_j^k,u_j^k]\subseteq[l_i^k,u_i^k];\\ [l_i^k,u_j^k], & l_i^k\in[l_j^k,u_j^k]\wedge u_j^k\in[l_i^k,u_i^k];\\ [l_j^k,u_i^k], & l_j^k\in[l_i^k,u_i^k]\wedge u_i^k\in[l_j^k,u_j^k];\\ \varnothing, & \text{其他}.\end{cases}$$

2) 并运算(join).

$$a_k(u_i)\cup a_k(u_j)=[\min(l_i^k,l_j^k),\max(u_i^k,u_j^k)].$$

3) 补运算(complement).

$$a_k^c(u_i)=(-\infty,l_i^k)\cup(u_i^k,+\infty);$$

4) 排斥并运算(exclusive join).

$$a_k(u_i)\oplus a_k(u_j)=\begin{cases}\{a_k^c(u_i)\cap a_k^c(u_j)\}^c, & a_k(u_i)\cap a_k(u_j)=\varnothing;\\ \{a_k(u_i)\cup a_k(u_j)\}\cap\{a_k(u_i)\cap a_k(u_j)\}^c, & \text{其他}.\end{cases}$$

下面,给出区间值相似率的概念.

定义 **2**. 设区间值信息系统 $\zeta=(U,AT,V,f)$,对任意的 $u_i,u_j\in U$, $a_k\in AT$,区间值 $a_k(u_i)=[l_i^k,u_i^k]$ 和 $a_k(u_j)=[l_j^k,u_j^k]$ 的相似率 α_{ij}^k 定义为:

$$\alpha_{ij}^k=\begin{cases}\dfrac{\min\{u_i^k-l_i^k,u_j^k-l_j^k\}}{\max\{u_i^k-l_i^k,u_j^k-l_j^k\}}, & [l_i^k,u_i^k]\subseteq[l_j^k,u_j^k]\vee[l_j^k,u_j^k]\subseteq[l_i^k,u_i^k];\\ 1-\min\left\{\dfrac{H\{a_k(u_i),a_k(u_j)\}}{\max\{u_i^k-l_i^k,u_j^k-l_j^k\}},1\right\}, & \text{其他}.\end{cases}$$

其中:$H\{a_k(u_i),a_k(u_j)\}=\max\{D_{ij}^k,D_{ji}^k\}$;

$$D_{ij}^k=\max_{a\in a_k(u_i)}\left\{\min_{b\in a_k(u_j)}\{d(a,b)\}\right\};$$

$$D_{ji}^k=\max_{a\in a_k(u_j)}\left\{\min_{b\in a_k(u_i)}\{d(a,b)\}\right\};$$

$d(a,b)$ 为 a,b 两点间的欧氏距离,显然有 $\alpha_{ij}^k\in[0,1]$.

关于相似率 α_{ij}^k 的取值与区间值位置关系的几点讨论:

1) 若 $\alpha_{ij}^k=0$,区间值 $a_k(u_i)$ 和 $a_k(u_j)$ 相离;

2) 若 $0<\alpha_{ij}^k<1$,区间值 $a_k(u_i)$ 和 $a_k(u_j)$ 部分相交或真包含;

3) 若 $\alpha_{ij}^k=1$,区间值 $a_k(u_i)$ 和 $a_k(u_j)$ 相同,即 $l_i^k=l_j^k$, $u_i^k=u_j^k$.

1.3 α-极大相容类

定义 **3**[11-13]. 设区间值信息系统 $\zeta=(U,AT,V,f)$, $A\subseteq AT$, $\alpha\in[0,1]$,则 ζ 中关于属性集 A 的α-相容关系定义为

$$T_A^\alpha=\{(u_i,u_j)\in U\times U:\alpha_{ij}^k>\alpha,\forall a_k\in A\}.$$

性质 **1**. 设区间值信息系统 $\zeta=(U,AT,V,f)$, $A\subseteq AT$, $\alpha\in[0,1]$,则:

1) T_A^α 具有自反性,即对任意的 $u_i\in U$,有 $(u_i,u_i)\in T_A^\alpha$;

2) T_A^α 具有对称性,即对任意的 $u_i,u_j\in U$,若 $(u_i,u_j)\in T_A^\alpha$,有 $(u_j,u_i)\in T_A^\alpha$;

3) T_A^α 不一定具有传递性,即对任意的 $u_i,u_j,u_m\in U$,若 $(u_i,u_m)\in T_A^\alpha$, $(u_m,u_j)\in T_A^\alpha$,不一定有 $(u_i,u_j)\in T_A^\alpha$.

性质 **2**. 设区间值信息系统 $\zeta=(U,AT,V,f)$, $A\subseteq AT$, $\alpha\in[0,1]$,则

$$T_A^\alpha=\bigcap_{a_k\in A}T_{\{a_k\}}^\alpha.$$

现有文献[11-13]中,往往采用相似类作为论域 U 的分类,定义如下:

定义 **4**. 设区间值信息系统 $\zeta=(U,AT,V,f)$, $A\subseteq AT$, $\alpha\in[0,1]$,定义对象 u_i 关于属性集 A 的α-相似类为

$$S_A^\alpha(u_i)=\{u_j\in U:(u_i,u_j)\in T_A^\alpha\}.$$

$S_A^\alpha(u_i)(u_i\in U)$ 形成对论域 U 的覆盖:

$$S^\alpha(A)=\{S_A^\alpha(u_1),S_A^\alpha(u_2),\dots,S_A^\alpha(u_n)\}.$$

采用相似类作为论域 U 的分类是将与对象 u_i 满足相容关系 T_A^α 的所有对象分为一类,这种分类方法存在2个问题:1)分类粒度粗、数目多、冗余度高;2)同一类中任意2个对象不一定满足相容关系 T_A^α,即同一类中任意2个对象不一定具有相同的属性特征,容易造成误分类.

为了解决以上2个问题,提高分类精度,在对象集的分类中引入了α-极大相容类的概念.

定义 **5**. 设 $\zeta=(U,AT,V,f)$ 是区间值信息系统, $A\subseteq AT$.若 $M\subseteq U$,对任意的 $u_i,u_j\in M$ 均满足 $(u_i,u_j)\in T_A^\alpha$,则称 M 是 ζ 中关于属性集 A 的α-相容类.若对任意的 $u_m\in U-M$,必存在 $u_i\in M$ 使得 $(u_i,u_m)\notin T_A^\alpha$.此时,称 M 为 ζ 中关于属性集 A 的α-极大相容类.

定义5说明:1)α-极大相容类是论域中对象满足两两相容的最大集合,任意α-相容类必可扩展成α-极大相容类;2)α-极大相容类是论域 U 中对象满

足传递关系的最大子集; 3) α-极大相容类的集合构成了论域 U 上的一个完全覆盖 $\xi^{\alpha}(A)=\{M_A^{\alpha}(u_1), M_A^{\alpha}(u_2), \cdots, M_A^{\alpha}(u_n)\}$, 其中 $M_A^{\alpha}(u_i)$ 为对象 u_i 关于属性集 A 的α-极大相容类.

采用α-极大相容类作为论域 U 的分类克服了采用相似类作为分类产生的 2 个问题, 提高了分类精度. 同时, 采用α-极大相容类可根据不同的用户需求和数据集的分布特点对阈值α进行动态调整, 亦可作为经典极大相容类应用[9]的有益扩展.

注 1. 粗糙集是粒计算[21]的重要分支理论之一, 粒计算是粗糙集、模糊集等理论的超集. 在基于相容关系的粒计算研究工作中, 郑征等人[22-24]提出了相容粒度空间模型, 研究了相容关系下粒度的理论与方法, 并将其应用于图像纹理识别中. 覆盖广义粗糙集, 不完备信息系统等相容关系粗糙集模型均可以归结在相容粒度空间计算的框架之内. 郑征等人提出的相容粒度空间模型为本文的论域分类工作提供了研究基础.

例 **1.** 表 1 所示的区间值信息系统 $\zeta=(U, AT, V, f)$, 令 $\alpha=0.7$, $\boldsymbol{T}_{AT}^{0.7}$ 对应的相似率布尔矩阵[11]如下:

$$\boldsymbol{T}_{AT}^{0.7}=\begin{pmatrix}1&0&0&0&0&0&1&0&0&0\\0&1&0&0&1&0&0&1&0&0\\0&0&1&0&0&0&0&0&0&0\\0&0&0&1&0&0&0&0&0&0\\0&1&0&0&1&0&0&0&0&0\\0&0&0&0&0&1&0&0&0&0\\1&0&0&0&0&0&1&0&0&0\\0&1&0&0&0&0&0&1&0&0\\0&0&0&0&0&0&0&0&1&0\\0&0&0&0&0&0&0&0&0&1\end{pmatrix}.$$

则 $S^{0.7}(AT)=\{S_{AT}^{0.7}(u_1), S_{AT}^{0.7}(u_2), \cdots, S_{AT}^{0.7}(u_{10})\}$.

其中:

$S_{AT}^{0.7}(u_1)=\{u_1, u_7\}, S_{AT}^{0.7}(u_2)=\{u_2, u_5, u_8\}$;

$S_{AT}^{0.7}(u_3)=\{u_3\}, S_{AT}^{0.7}(u_4)=\{u_4\}$;

$S_{AT}^{0.7}(u_5)=\{u_2, u_5\}, S_{AT}^{0.7}(u_6)=\{u_6\}$;

$S_{AT}^{0.7}(u_7)=\{u_1, u_7\}, S_{AT}^{0.7}(u_8)=\{u_2, u_8\}$;

$S_{AT}^{0.7}(u_9)=\{u_9\}, S_{AT}^{0.7}(u_{10})=\{u_{10}\}$.

$\xi^{0.7}(AT)=\{M_{AT}^{0.7}(u_1), M_{AT}^{0.7}(u_2), \cdots, M_{AT}^{0.7}(u_{10})\}$.

其中:

$M_{AT}^{0.7}(u_1)=\{u_1, u_7\}, M_{AT}^{0.7}(u_2)=\{u_2, u_5\}$;

$M_{AT}^{0.7}(u_3)=\{u_3\}, M_{AT}^{0.7}(u_4)=\{u_4\}$;

$M_{AT}^{0.7}(u_6)=\{u_6\}, M_{AT}^{0.7}(u_8)=\{u_2, u_8\}$;

$M_{AT}^{0.7}(u_9)=\{u_9\}, M_{AT}^{0.7}(u_{10})=\{u_{10}\}$.

$\xi_{AT}^{\alpha}(u_i)$ 是区间值信息系统 ζ 中关于属性集 AT 的所有包含对象 u_i 的α-极大相容类的集合.

$\xi_{AT}^{0.7}(u_1)=\{M_{AT}^{0.7}(u_1)=\{u_1, u_7\}\}$;

$\xi_{AT}^{0.7}(u_2)=\{M_{AT}^{0.7}(u_2)=\{u_2, u_5\}$;

$M_{AT}^{0.7}(u_8)=\{u_2, u_8\}\}$;

$\xi_{AT}^{0.7}(u_3)=\{M_{AT}^{0.7}(u_3)=\{u_3\}\}$;

$\xi_{AT}^{0.7}(u_4)=\{M_{AT}^{0.7}(u_4)=\{u_4\}\}$;

$\xi_{AT}^{0.7}(u_5)=\{M_{AT}^{0.7}(u_2)=\{u_2, u_5\}\}$;

$\xi_{AT}^{0.7}(u_6)=\{M_{AT}^{0.7}(u_6)=\{u_6\}\}$;

$\xi_{AT}^{0.7}(u_7)=\{M_{AT}^{0.7}(u_1)=\{u_1, u_7\}\}$;

$\xi_{AT}^{0.7}(u_8)=\{M_{AT}^{0.7}(u_8)=\{u_2, u_8\}\}$;

$\xi_{AT}^{0.7}(u_9)=\{M_{AT}^{0.7}(u_9)=\{u_9\}\}$;

$\xi_{AT}^{0.7}(u_{10})=\{M_{AT}^{0.7}(u_{10})=\{u_{10}\}\}$.

$S_A^{\alpha}(u_i)$ 和 $\xi_A^{\alpha}(u_i)$ 具有如下性质:

性质 **3.** 设区间值信息系统 $\zeta=(U, AT, V, f)$, 对任意的 $u_i\in U$, $A\subseteq AT$, 则有 $S_A^{\alpha}(u_i)=\bigcup\{M\in\xi_A^{\alpha}(u_i)\}$.

证明.

1) 由定义 4, 5, 对任意的 $u_j\in S_A^{\alpha}(u_i)$, 易得 $u_j\in\bigcup\{M\in\xi_A^{\alpha}(u_i)\}$, 故 $S_A^{\alpha}(u_i)\subseteq\bigcup\{M\in\xi_A^{\alpha}(u_i)\}$.

2) 对任意的 $u_j\in M$, $M\in\xi_A^{\alpha}(u_i)$, 即 u_i, u_j 满足α-相容关系 T_A^{α}. 故 $u_j\in S_A^{\alpha}(u_i)$. 进而 $\bigcup\{M\in\xi_A^{\alpha}(u_i)\}\subseteq S_A^{\alpha}(u_i)$.

由 1) 2) 得 $S_A^{\alpha}(u_i)=\bigcup\{M\in\xi_A^{\alpha}(u_i)\}$.　　证毕.

性质 **4.** 设 $\zeta=(U, AT, V, f)$ 是区间值信息系统, $B\subseteq A\subseteq AT$. 对任意的 $M\in\xi_A^{\alpha}(u_i)$, $u_i\in U$, 总存在 $M'\in\xi_B^{\alpha}(u_i)$ 满足 $M\subseteq M'$.

性质 **5.** 设 $\zeta=(U, AT, V, f)$ 是区间值信息系统, $B\subseteq A\subseteq AT$. 对任意的 $M\in\xi^{\alpha}(A)$, 总存在 $M'\in\xi^{\alpha}(B)$ 满足 $M\subseteq M'$.

性质 **6.** 设区间值信息系统 $\zeta=(U, AT, V, f)$, $A\subseteq AT$, $0\leqslant\alpha\leqslant\beta\leqslant1$. 对任意的 $M\in\xi_A^{\beta}(u_i)$, 总存在 $M'\in\xi_A^{\alpha}(u_i)$ 满足 $M\subseteq M'$.

性质 **7.** 设区间值信息系统 $\zeta=(U, AT, V, f)$, $A\subseteq AT$, $0\leqslant\alpha\leqslant\beta\leqslant1$. 对任意的 $M\in\xi^{\beta}(A)$, 总存在 $M'\in\xi^{\alpha}(A)$ 满足 $M\subseteq M'$.

性质 4~7 说明, 增加区间值信息系统的属性数目或者提高α值, 论域 U 的分类变细, 知识粒度减小, 分辨度增加.

1.4　区间值信息系统中的粗糙近似

首先, 给出由相似类定义的粗糙近似算子[12-13]如下.

定义 **6.** 区间值信息系统 $\zeta=(U, AT, V, f)$, 对任意的 $X\subseteq U$, $A\subseteq AT$, 粗糙上下近似算子为:

$\overline{APR}_A^{\alpha}(X)=\{u_i\in U: S_A^{\alpha}(u_i)\cap X\neq\varnothing\}$;

$\underline{APR}_A^{\alpha}(X)=\{u_i\in U: S_A^{\alpha}(u_i)\subseteq X\}$.

由定义6, 可得:

性质**8**. 设区间值信息系统 $\zeta=(U, AT, V, f)$, 对任意的 $X\subseteq U$, $A\subseteq AT$, 有:

$$\overline{APR}_A^{\alpha}(X)=\{u_i\in U: S_A^{\alpha}(u_i)\cap X\neq\varnothing\}=\bigcup\{S_A^{\alpha}(u_i): u_i\in X\}\neq\bigcup\{S_A^{\alpha}(u_i): S_A^{\alpha}(u_i)\cap X\neq\varnothing\};$$

$$\underline{APR}_A^{\alpha}(X)=\{u_i\in U: S_A^{\alpha}(u_i)\subseteq X\}=\{u_i\in X: S_A^{\alpha}(u_i)\subseteq X\}\neq\bigcup\{S_A^{\alpha}(u_i): S_A^{\alpha}(u_i)\subseteq X\}.$$

由性质3可知, 在论域 U 的分类中, α-极大相容类的采用使同一类中的对象具有相同的属性特征, 细分了α-相似类, 提高了分类精度, 从而使上下近似对任意集合 X 的刻画更加准确. 为此, 提出了由α-极大相容类定义的区间值信息系统的粗糙上下近似算子.

定义**7**. 设区间值信息系统 $\zeta=(U, AT, V, f)$, 对任意的 $X\subseteq U$, $A\subseteq AT$, 定义集合 X 的粗糙上下近似算子为:

$$\overline{apr}_A^{\alpha}(X)=\{u_i\in U: M_A^{\alpha}(u_i)\cap X\neq\varnothing\};$$

$$\underline{apr}_A^{\alpha}(X)=\{u_i\in U: M_A^{\alpha}(u_i)\subseteq X\}.$$

正域、边界域和负域, 记作:

$$pos_A^{\alpha}(X)=\underline{apr}_A^{\alpha}(X);$$

$$bnr_A^{\alpha}(X)=\overline{apr}_A^{\alpha}(X)-\underline{apr}_A^{\alpha}(X);$$

$$neg_A^{\alpha}(X)=U-\overline{apr}_A^{\alpha}(X)=U-pos_A^{\alpha}(X)\cup bnr_A^{\alpha}(X).$$

定义7的重要性质介绍如下.

性质**9**. 设区间值信息系统 $\zeta=(U, AT, V, f)$, 对任意的 $X\subseteq U$, $A\subseteq AT$, 则:

$$\overline{apr}_A^{\alpha}(X)=\{u_i\in U: M_A^{\alpha}(u_i)\cap X\neq\varnothing\}=\bigcup\{M_A^{\alpha}(u_i): M_A^{\alpha}(u_i)\cap X\neq\varnothing\};$$

$$\underline{apr}_A^{\alpha}(X)=\{u_i\in U: M_A^{\alpha}(u_i)\subseteq X\}=\bigcup\{M_A^{\alpha}(u_i): M_A^{\alpha}(u_i)\subseteq X\}.$$

定义7给出了区间值信息系统中基于对象的粗糙上下近似算子的定义. 由性质9可知, 基于对象和粒定义的粗糙上下近似算子是等价的.

性质**10**. 设区间值信息系统 $\zeta=(U, AT, V, f)$, 对任意的 $X\subseteq Y\subseteq U$, $A\subseteq AT$, $\alpha\in[0, 1]$, 则:

1) $\underline{apr}_A^{\alpha}(X)\subseteq X\subseteq\overline{apr}_A^{\alpha}(X)$;

2) $\underline{apr}_A^{\alpha}(\varnothing)=\overline{apr}_A^{\alpha}(\varnothing)=\varnothing$, $\underline{apr}_A^{\alpha}(U)=\overline{apr}_A^{\alpha}(U)=U$;

3) $\overline{apr}_A^{\alpha}(\overline{apr}_A^{\alpha}(X))=\underline{apr}_A^{\alpha}(\overline{apr}_A^{\alpha}(X))=\overline{apr}_A^{\alpha}(X)$, $\underline{apr}_A^{\alpha}(\underline{apr}_A^{\alpha}(X))=\overline{apr}_A^{\alpha}(\underline{apr}_A^{\alpha}(X))=\underline{apr}_A^{\alpha}(X)$;

4) $\overline{apr}_A^{\alpha}(X)=\sim\underline{apr}_A^{\alpha}(\sim X)$, $\underline{apr}_A^{\alpha}(X)=\sim\overline{apr}_A^{\alpha}(\sim X)$;

5) $\underline{apr}_A^{\alpha}(X)\subseteq\underline{apr}_A^{\alpha}(Y)$; $\overline{apr}_A^{\alpha}(X)\subseteq\overline{apr}_A^{\alpha}(Y)$;

6) $\underline{apr}_A^{\alpha}(X\cap Y)=\underline{apr}_A^{\alpha}(X)\cap\underline{apr}_A^{\alpha}(Y)$, $\underline{apr}_A^{\alpha}(X\cup Y)\supseteq\underline{apr}_A^{\alpha}(X)\cup\underline{apr}_A^{\alpha}(Y)$;

7) $\overline{apr}_A^{\alpha}(X\cup Y)=\overline{apr}_A^{\alpha}(X)\cup\overline{apr}_A^{\alpha}(Y)$, $\overline{apr}_A^{\alpha}(X\cap Y)\subseteq\overline{apr}_A^{\alpha}(X)\cap\overline{apr}_A^{\alpha}(Y)$.

性质**11**. 设区间值信息系统 $\zeta=(U, AT, V, f)$, 对任意的 $A\subseteq B\subseteq AT$, $0\leqslant\alpha\leqslant\beta\leqslant1$, 则:

1) $\overline{apr}_A^{\beta}(X)\subseteq\overline{apr}_A^{\alpha}(X)$, 但不一定总有 $\underline{apr}_A^{\alpha}(X)\subseteq\underline{apr}_A^{\beta}(X)$ 成立;

2) $\overline{apr}_B^{\alpha}(X)\subseteq\overline{apr}_A^{\alpha}(X)$, 但不一定总有 $\underline{apr}_A^{\alpha}(X)\subseteq\underline{apr}_B^{\alpha}(X)$ 成立.

集合的不精确性是由边界域的存在而引起的. 集合的边界域越大, 其精确性则越低. 为了准确表达这一点, 我们引入了由α-极大相容类定义的集合 X 的粗糙近似精度为:

定义**8**. 设区间值信息系统 $\zeta=(U, AT, V, f)$, 对任意的 $X\subseteq U$, $A\subseteq AT$, $\alpha\in[0, 1]$, 定义集合 X 的粗糙近似精度为

$$\mu_A^{\alpha}(X)=\frac{|\underline{apr}_A^{\alpha}(X)|}{|\overline{apr}_A^{\alpha}(X)|}=\frac{|\underline{apr}_A^{\alpha}(X)|}{|U|-|\underline{apr}_A^{\alpha}(\sim X)|},$$

其中, $|S|$ 表示集合 S 的基数.

定理**1**. 设区间值信息系统 $\zeta=(U, AT, V, f)$, $A\subseteq AT$, $X\subseteq U$, 则

$$\mu_A'^{\alpha}(X)=\frac{|\underline{APR}_A^{\alpha}(X)|}{|\overline{APR}_A^{\alpha}(X)|}\leqslant\frac{|\underline{apr}_A^{\alpha}(X)|}{|\overline{apr}_A^{\alpha}(X)|}=\mu_A^{\alpha}(X).$$

证明.

1) 首先, 证明 $\underline{ARP}v_A(X)\subseteq\underline{apr}_A^{\alpha}(X)$. 由定义6、性质8, 得:

$$\underline{APR}_A^{\alpha}(X)=\{u_i\in U: S_A^{\alpha}(u_i)\subseteq X\}=\{u_i\in X: S_A^{\alpha}(u_i)\subseteq X\}=\{u_i\in X: \bigcup\{M\in\xi_A^{\alpha}(u_i)\}\subseteq X\};$$

$$\underline{apr}_A^{\alpha}(X)=\bigcup\{M_A^{\alpha}(u_i): M_A^{\alpha}(u_i)\subseteq X\}=\bigcup\{M\in\xi_A^{\alpha}(u_i): M\subseteq X\}=\{u_i\in X: M\in\xi_A^{\alpha}(u_i)\subseteq X\}.$$

又由

$$\{u_i\in X: \bigcup\{M\in\xi_A^{\alpha}(u_i)\}\subseteq X\}\subseteq\{u_i\in X: M\in\xi_A^{\alpha}(u_i)\subseteq X\}$$

得

$$\underline{APR}_A^{\alpha}(X)\subseteq\underline{apr}_A^{\alpha}(X).$$

2) 其次, 证明 $\overline{APR}_A^{\alpha}(X)=\overline{apr}_A^{\alpha}(X)$.

由

$$\overline{APR}_A^{\alpha}(X)=\{u_i\in U: S_A^{\alpha}(u_i)\cap X\neq\varnothing\}=$$
$$\{u_i\in U:\ \cup\{M\in\xi_A^{\alpha}(u_i)\}\cap X\neq\varnothing\}=$$
$$\cup\{M\in\xi^{\alpha}(A): M\cap X\neq\varnothing\}=\overline{apr}_A^{\alpha}(X).$$

综上, 有$\underline{APR}_A^{\alpha}(X)\subseteq\underline{apr}_A^{\alpha}(X)$, $\overline{APR}_A^{\alpha}(X)=\overline{apr}_A^{\alpha}(X)$成立, 故

$$\frac{|\underline{APR}_A^{\alpha}(X)|}{|\overline{APR}_A^{\alpha}(X)|}\leqslant\frac{|\underline{apr}_A^{\alpha}(X)|}{|\overline{apr}_A^{\alpha}(X)|}.$$　证毕.

定理 1 表明相对于定义 6, 定义 7 给出的粗糙上下近似算子可以获得更高的粗糙近似精度. 下面, 给出一个例子来说明定理 1.

例 **2.** 如表 1 所示的区间值信息系统 $\zeta=(U, AT, V, f)$, 令 $X=\{u_2, u_4, u_5, u_6, u_7\}$, 集合 X 的粗糙上下近似为:

$$\overline{APR}_{AT}^{0.7}(X)=\{u_i\in U: S_{AT}^{0.7}(u_i)\cap X\neq\varnothing\}=$$
$$\{u_1, u_2, u_4, u_5, u_6, u_7, u_8\},$$
$$\underline{APR}_{AT}^{0.7}(X)=\{u_i\in U: S_{AT}^{0.7}(u_i)\subseteq X\}=\{u_4, u_5, u_6\};$$
$$\overline{apr}_{AT}^{0.7}(X)=\{u_i: M_{AT}^{0.7}(u_i)\cap X\neq\varnothing\}=$$
$$\{u_1, u_2, u_4, u_5, u_6, u_7, u_8\},$$
$$\underline{apr}_{AT}^{0.7}(X)=\{u_i: M_A^{\alpha}(u_i)\subseteq X\}=\{u_2, u_4, u_5, u_6\}.$$

则集合 X 的粗糙近似精度 $\mu_{AT}^{0.7}(X)$, $\mu_{AT}^{0.7}(X)$为:

$$\mu_{AT}^{0.7}(X)=\frac{|\underline{APR}_{AT}^{0.7}(X)|}{|\overline{APR}_{AT}^{0.7}(X)|}=\frac{3}{7};$$

$$\mu_{AT}^{0.7}(X)=\frac{|\underline{apr}_{AT}^{0.7}(X)|}{|\overline{apr}_{AT}^{0.7}(X)|}=\frac{4}{7}.$$

显然, 有 $\mu_{AT}^{0.7}(X)<\mu_{AT}^{0.7}(X)$.

2　区间值信息系统的知识约简方法

在基于相容关系的广义粗糙集模型研究中, Leung 和 Fischer 等人[11] 提出了 α-区分集(α-discernibility set), 用于构造区间值信息系统的 α-区分函数(α-discernibility function). 文献[9] 给出了一种基于极大相容块(maximal consistent block, MCB) 的不完备信息系统的知识获取模型. 结合上述工作, 在本节, 我们在定义 9, 10 给出了两种区间值信息系统的知识约简定义, 并在定理 2, 3 提出了对应的知识约简方法. 最后, 在定理 4 讨论了 2 种知识约简之间的关系.

定义 **9.** 设 $\zeta=(U, AT, V, f)$ 是区间值信息系统, $A\subseteq AT$. 若属性子集 A 满足:

1) $\xi^{\alpha}(A)=\xi^{\alpha}(AT)$;

2) 不存在任何 $B\subset A$, 满足 $\xi^{\alpha}(B)=\xi^{\alpha}(AT)$.

则称属性子集 A 为区间值信息系统 ζ 的一个约简. ζ 中所有约简的集合, 记作 $red^{\alpha}(AT)$; 所有约简的交集称为核属性集, 记作 $cor^{\alpha}(AT)$.

实际上, 定义 9 给出了判断任意属性子集 A 是否为区间值信息系统 ζ 的约简的方法, 由此可以进一步得到采用区分函数进行区间值信息系统知识约简的计算方法.

定理 **2.** 设区间值信息系统 $\zeta=(U, AT, V, f)$, 若 $A\subseteq AT$ 是 ζ 的一个约简, 当且仅当 $\wedge A$ 是区分函数

$$\psi^{\alpha}(AT)=\bigwedge_{\substack{(u_i, M_{AT}^{\alpha}(u_j))\in U\times\{M_{AT}^{\alpha}(u_j)\in\\ \xi^{\alpha}(AT)-\xi_{AT}^{\alpha}(u_i)\}}}\alpha_{AT}\{u_i, M_{AT}^{\alpha}(u_j)\}$$

的一个素蕴涵.

其中:

$$\alpha_{AT}\{u_i, M_{AT}^{\alpha}(u_j)\}=\bigwedge_{u_j\in M_{AT}^{\alpha}(u_j)}\vee\alpha(u_i, u_j),$$

$$\alpha(u_i, u_j)=\{a_k\in AT: \alpha_{ij}^k\leqslant\alpha\}.$$

证明.

设

$$C=\{A\subseteq AT: \text{对 } \forall u_i\in U, \text{有}$$
$$\forall M'\in\xi^{\alpha}(AT)-\xi_{AT}^{\alpha}(u_i), \forall u_j\in M',$$
$$A\cap\alpha(u_i, u_j)\neq\varnothing;$$
$$\forall M_{AT}^{\alpha}(u_m), M_{AT}^{\alpha}(u_n)\in\xi_{AT}^{\alpha}(u_i),$$
$$M_{AT}^{\alpha}(u_m)\neq M_{AT}^{\alpha}(u_n),$$
$$\forall(u_m, u_n)\in M_{AT}^{\alpha}(u_m)\times M_{AT}^{\alpha}(u_n),$$
$$A\cap\alpha(u_m, u_n)\neq\varnothing, \alpha(u_m, u_n)\neq\varnothing\}$$
$$R=\{A\subseteq AT: \xi^{\alpha}(A)=\xi^{\alpha}(AT)\}.$$

“$C\Rightarrow R$”, 反证法. 若 $M=M_A^{\alpha}(u_i)\in\xi^{\alpha}(A)$, 对任意的 u_j, $u_m\in M$, 则属性集 A 中任何一个属性都不能区分对象 u_i, u_j 和 u_m, 故 $A\cap\alpha(u_i, u_j)=\varnothing$, $A\cap\alpha(u_j, u_m)=\varnothing$. 假设 $M\notin\xi^{\alpha}(AT)$, 存在 u_m 满足 $u_m\in M'$, $M'\in\xi(AT)-\xi_{AT}(u_i)$ 或者 $u_m\in M''$, $M''\in\xi_{AT}(u_i)$, $M''\neq M$, 则 $A\cap\alpha(u_i, u_j)\neq\varnothing$, 或 $A\cap\alpha(u_j, u_m)\neq\varnothing$, 推出矛盾, 故 $\xi^{\alpha}(A)\subseteq\xi^{\alpha}(AT)$. 又由性质 5, 易得 $\xi^{\alpha}(AT)\subseteq\xi^{\alpha}(A)$, 故 $\xi^{\alpha}(A)=\xi^{\alpha}(AT)$.

“$R\Rightarrow C$”, 对任意的 $u_i\in U$, 若 $M=M_{AT}^{\alpha}(u_i)\in\xi_{AT}(u_i)$, u_i, $u_j\in M$, $u_m\notin M$. 则有 $u_m\in M'$, $M'\in\xi^{\alpha}(AT)-\xi_{AT}^{\alpha}(u_i)$ 或 $u_m\in M''$, $M''\in\xi_{AT}^{\alpha}(u_i)$, $M''\neq M$. 两种情况下, 分别有 $AT\cap\alpha(u_i, u_j)\neq\varnothing$, $AT\cap\alpha(u_j, u_m)\neq\varnothing$. 由 $\xi^{\alpha}(A)=\xi^{\alpha}(AT)$, 即属性集 A 与属性集 AT 对任意对象 $u_i\in U$ 有相同的区分能力, 故 $A\cap\alpha(u_i, u_j)\neq\varnothing$, $A\cap\alpha(u_j, u_m)\neq\varnothing$.　证毕.

在实际应用中, 常常需要将论域中某个对象与剩余对象区分开, 不考虑剩余对象之间的差异. 为此, 引入区间值信息系统中关于对象 u_i 的知识约简定义.

定义 **10.** 设 $\zeta=(U,AT,V,f)$ 是区间值信息系统, $A\subseteq AT$. 若属性子集 A 满足:

1) $\xi_A^{\alpha}(u_i)=\xi_{AT}^{\alpha}(u_i)$;

2) 不存在任何 $B\subset A$, 满足 $\xi_B^{\alpha}(u_i)=\xi_{AT}^{\alpha}(u_i)$.

则称属性子集 A 是区间值信息系统 ζ 中关于对象 u_i 的一个约简. 关于对象 u_i 所有约简的集合, 记作 $red_{AT}^{\alpha}(u_i)$; 所有约简的交集称为核属性集, 记作 $cor_{AT}^{\alpha}(u_i)$.

定理 **3.** 设 $\zeta=(U,AT,V,f)$ 是区间值信息系统, 若 $A\subseteq AT$ 是 ζ 中关于 u_i 的一个约简, 当且仅当 $\wedge A$ 是区分函数

$$\psi_{AT}^{\alpha}(u_i)=\bigwedge_{M_{AT}^{\alpha}(u_j)\in\xi^{\alpha}(AT)-\xi_{AT}^{\alpha}(u_i)}\alpha_{AT}\{u_i,M_{AT}^{\alpha}(u_j)\}$$
$$\bigwedge_{\{M_{AT}^{\alpha}(u_m),M_{AT}^{\alpha}(u_n)\}\in\xi_{AT}^{\alpha}(u_i)\times\xi_{AT}^{\alpha}(u_i)}\alpha_{AT}\{M_{AT}^{\alpha}(u_m),M_{AT}^{\alpha}(u_n)\}$$

的一个素蕴涵.

其中:

$$\alpha_{AT}\{u_i,M_{AT}^{\alpha}(u_j)\}=\bigwedge_{u_j\in M_{AT}^{\alpha}(u_j)}\vee\alpha(u_i,u_j),$$

$$\alpha(u_i,u_j)=\{a_k\in AT:\alpha_{ij}^k\leqslant\alpha\};$$

$$\alpha_{AT}\{M_{AT}^{\alpha}(u_m),M_{AT}^{\alpha}(u_n)\}=$$
$$\bigwedge_{(u_m,u_n)\in\{M_{AT}^{\alpha}(u_m)-\{u_i\}\}\times\{M_{AT}^{\alpha}(u_n)-\{u_i\}\}}\vee\alpha(u_m,u_n),$$

$$\alpha(u_m,u_n)=\{a_k\in AT:\alpha_{mn}^k\leqslant\alpha\}.$$

证明.

设:

$$\begin{aligned}C=\{A\subseteq AT:&\\ &\forall M'\in\xi^{\alpha}(AT)-\xi_{AT}^{\alpha}(u_i),\ \forall u_j\in M',\\ &A\cap\alpha(u_i,u_j)\neq\varnothing;\\ &\forall M_{AT}^{\alpha}(u_m),M_{AT}^{\alpha}(u_n)\in\xi_{AT}^{\alpha}(u_i),\\ &M_{AT}^{\alpha}(u_m)\neq M_{AT}^{\alpha}(u_n),\\ &\forall(u_m,u_n)\in\{M_{AT}^{\alpha}(u_m)-\{u_i\}\}\times\\ &\{M_{AT}^{\alpha}(u_n)-\{u_i\}\},\\ &A\cap\alpha(u_m,u_n)\neq\varnothing,\alpha(u_m,u_n)\neq\varnothing\};\end{aligned}$$
$$R=\{A\subseteq AT:\xi_A^{\alpha}(u_i)=\xi_{AT}^{\alpha}(u_i)\}.$$

“$C\Rightarrow R$”, 反证法. 如果 $M\in\xi_A^{\alpha}(u_i)$, $u_i,u_m,u_n\in M$, 即属性集 A 中任何属性都不能区分对象 u_i, u_m 和 u_n, 即 $A\cap\alpha(u_i,u_n)=\varnothing$, $A\cap\alpha(u_m,u_n)=\varnothing$. 假设 $M\notin\xi_{AT}^{\alpha}(u_i)$, 存在 u_n 满足 $M-\{u_n\}\in\xi_{AT}^{\alpha}(u_i)$. 有如下 2 种情况: ①若 $u_n\in M'$, 这里 $M'\in\xi^{\alpha}(AT)-\xi_{AT}^{\alpha}(u_i)$. 则 $A\cap\alpha(u_i,u_n)\neq\varnothing$, 推出矛盾; ②若 $u_n\in M''$, $M''\in\xi_{AT}^{\alpha}(u_i)$, 则 $A\cap\alpha(u_m,u_n)\neq\varnothing$, 推出矛盾, 故 $\xi_A^{\alpha}(u_i)\subseteq\xi_{AT}^{\alpha}(u_i)$. 由性质 4, 易得 $\xi_{AT}^{\alpha}(u_i)\subseteq\xi_A^{\alpha}(u_i)$, 故 $\xi_A^{\alpha}(u_i)=\xi_{AT}^{\alpha}(u_i)$.

“$R\Rightarrow C$”, 若 $M\in\xi_{AT}^{\alpha}(u_i)$, $u_i,u_m\in M$, $u_n\notin M$, 则有 $u_n\in M'$, $M'\in\xi^{\alpha}(AT)-\xi_{AT}^{\alpha}(u_i)$ 或 $u_n\in M''$, $M''\in\xi_{AT}^{\alpha}(u_i)$, $M''\neq M$. 两种情况下, 分别有 $AT\cap\alpha(u_i,u_n)\neq\varnothing$, $AT\cap\alpha(u_m,u_n)\neq\varnothing$. 由 $\xi_{AT}^{\alpha}(u_i)=\xi_A^{\alpha}(u_i)$, 即属性集 A 与属性集 AT 对 u_i 有相同的区分能力, 故 $A\cap\alpha(u_i,u_n)\neq\varnothing$, $A\cap\alpha(u_m,u_n)\neq\varnothing$.

证毕.

定理 **4.** 设区间值信息系统 $\zeta=(U,AT,V,f)$, 有
$$\psi^{\alpha}(AT)=\bigwedge_{u_i\in U}\psi_{AT}^{\alpha}(u_i).$$

证明. 由定理 2, 3 易证.

定理 2, 3 给出了区间值信息系统知识约简的计算方法, 下面给出一个数值计算的例子.

例 **3.** 表 1 所示区间值信息系统 $\zeta=(U,AT,V,f)$ 的知识约简 $\psi^{0.7}(AT)$ 为:

$$\psi^{0.7}(AT)=a_1\wedge a_4\wedge(a_2\vee a_5)=$$
$$(a_1\wedge a_2\wedge a_4)\vee(a_1\wedge a_4\wedge a_5),$$
$$red^{0.7}(AT)=\{\{a_1,a_2,a_4\},\{a_1,a_4,a_5\}\},$$
$$cor^{0.7}(AT)=\{a_1,a_2,a_4\}\cap\{a_1,a_4,a_5\}=\{a_1,a_4\}.$$

论域 U 中任意对象 u_i 的约简为:

$$\psi_{AT}^{0.7}(u_1)=a_1\wedge(a_2\vee a_5)=$$
$$(a_1\wedge a_2)\vee(a_1\wedge a_5),$$
$$red_{AT}^{0.7}(u_1)=\{\{a_1,a_2\},\{a_1,a_5\}\};$$
$$\psi_{AT}^{0.7}(u_2)=a_1\wedge a_4,red_{AT}^{0.7}(u_2)=\{\{a_1,a_4\}\};$$
$$\psi_{AT}^{0.7}(u_3)=a_2\vee a_5,red_{AT}^{0.7}(u_3)=\{\{a_2\},\{a_5\}\};$$
$$\psi_{AT}^{0.7}(u_4)=(a_1\vee a_2)\wedge(a_3\vee a_4\vee a_5)=$$
$$(a_1\wedge a_3)\vee(a_1\wedge a_4)\vee(a_1\wedge a_5)\wedge$$
$$(a_2\wedge a_3)\vee(a_2\wedge a_4)\vee(a_2\wedge a_5),$$
$$red_{AT}^{0.7}(u_4)=\{\{a_1,a_3\},\{a_1,a_4\},\{a_1,a_5\},$$
$$\{a_2,a_3\},\{a_2,a_4\},\{a_2,a_5\}\};$$
$$\psi_{AT}^{0.7}(u_5)=a_1\wedge a_4,red_{AT}^{0.7}(u_5)=\{\{a_1,a_4\}\};$$
$$\psi_{AT}^{0.7}(u_6)=a_1,red_{AT}^{0.7}(u_6)=\{\{a_1\}\};$$
$$\psi_{AT}^{0.7}(u_7)=a_1\wedge(a_2\vee a_5)=(a_1\wedge a_2)\vee(a_1\wedge a_5),$$
$$red_{AT}^{0.7}(u_7)=\{\{a_1,a_2\},\{a_1,a_5\}\};$$
$$\psi_{AT}^{0.7}(u_8)=a_1\wedge a_4,red_{AT}^{0.7}(u_8)=\{\{a_1,a_4\}\};$$
$$\psi_{AT}^{0.7}(u_9)=a_4,red_{AT}^{0.7}(u_9)=\{\{a_4\}\};$$
$$\psi_{AT}^{0.7}(u_{10})=a_1\vee a_2,red_{AT}^{0.7}(u_{10})=\{\{a_1\},\{a_2\}\}.$$

3 区间值决策系统的知识约简方法

作为区间值信息系统知识约简的特例, 本节将重点讨论区间值决策系统的知识约简.

为了解决不完备决策系统中决策规则的非协调性, Kryszkiewicz[25-26] 提出了广义决策函数. 由于区间值的不确定性以及对阈值 α 不恰当的选取也造成了大量的非协调性规则. 为此, 定义区间值决策系统

中的α-广义决策函数如下:

定义 **11.** 区间值决策系统 $\zeta=(U, AT\cup D, V, f)$, 其中 $AT=\{a_1, a_2, \cdots, a_m\}$ 为条件属性集, $D=\{d\}$ 为决策属性集, 对任意的 $A\subseteq AT$. 称

$$\partial_A^\alpha(u_i)=\{f_d(u_j): u_j\in S_A^\alpha(u_i)\}$$

为区间值决策系统 ζ 的α-广义决策函数, 且

$$\partial^\alpha(A)=\{\partial_A^\alpha(u_1), \partial_A^\alpha(u_2), \cdots, \partial_A^\alpha(u_n)\}.$$

类似地, 由α-极大相容类定义的α-广义决策函数 $\tau_A^\alpha(u_i)$ 为:

$$\tau_A^\alpha(u_i)=\{f_d(u_j): u_j\in M_A^\alpha(u_i)\}.$$

性质 **12.** 设区间值决策系统 $\zeta=(U, AT\cup D, V, f)$, 对任意的 $A\subseteq B\subseteq AT$, $0\leqslant\alpha\leqslant\beta\leqslant 1$, 则:

1) $\partial_A^\beta(u_i)\subseteq\partial_A^\alpha(u_i)$, $\tau_A^\beta(u_i)\subseteq\tau_A^\alpha(u_i)$;

2) $\partial_B^\alpha(u_i)\subseteq\partial_A^\alpha(u_i)$, $\tau_B^\alpha(u_i)\subseteq\tau_A^\alpha(u_i)$.

定义 **12.** 设区间值决策系统 $\zeta=(U, AT\cup D, V, f)$, 对于 $A\subseteq AT$. 若属性子集 A 满足:

1) $\partial_A^\alpha(u_i)=\partial_{AT}^\alpha(u_i)$;

2) 不存在任何 $B\subset A$, 满足 $\partial_B^\alpha(u_i)=\partial_{AT}^\alpha(u_i)$.

则称属性子集 A 为区间值决策系统 ζ 中关于对象 u_i 的一个约简. 关于对象 u_i 所有约简的集合, 记作 $red_{AT\cup D}^\alpha(u_i)$; 所有约简的交集称为核属性集, 记作 $cor_{AT\cup D}^\alpha(u_i)$.

定理 **5.** 设 $\zeta=(U, AT\cup D, V, f)$ 是区间值决策系统, 若 $A\subseteq AT$ 是 ζ 中关于 u_i 的一个约简, 当且仅当 $\wedge A$ 是区分函数

$$\varphi_{AT}^\alpha(u_i)=\bigwedge_{\substack{M_{AT}^\alpha(u_j)\in\xi^\alpha(AT)-\\ \xi_{AT}^\alpha(u_i)(\tau_{AT}^\alpha(u_j)\not\subseteq\partial_{AT}^\alpha(u_i))}}\alpha_{AT}\{u_i, M_{AT}^\alpha(u_j)\}$$

的一个素蕴涵.

其中:

$$\alpha_{AT}\{u_i, M_{AT}^\alpha(u_j)\}=\bigwedge_{u_j\in M_{AT}^\alpha(u_j)(d(u_j)\notin\partial_{AT}^\alpha(u_i))}\vee\alpha(u_i, u_j),$$

$$\alpha(u_i, u_j)=\{a_k\in AT: \alpha_{ij}^k\leqslant\alpha\}.$$

证明.

设:

$$C=\{A\subseteq AT:$$
$$\forall M'\in\xi^\alpha(AT)-\xi_{AT}^\alpha(u_i), \forall u_j\in M',$$
$$A\cap\alpha(u_i, u_j)\neq\varnothing;$$
$$d(u_j)\notin\partial_{AT}^\alpha(u_i), \tau_{AT}^\alpha(u_j)\not\subseteq\partial_{AT}^\alpha(u_i)\};$$
$$R=\{A\subseteq AT: \partial_A^\alpha(u_i)=\partial_{AT}^\alpha(u_i)\}.$$

"$C\Rightarrow R$", 反证法. 若 $v\in\partial_A^\alpha(u_i)$, 则存在 $u_j\in M$, $M\in\xi_A^\alpha(u_i)$ 满足 $d(u_j)=v\in\partial_A^\alpha(u_i)$, 此时 $A\cap\alpha(u_i, u_j)=\varnothing$. 如果假设 $v\notin\partial_{AT}^\alpha(u_i)$, 则 $u_j\in M'$, $M'\in\xi^\alpha(AT)-\xi_{AT}^\alpha(u_i)$, 且 $\tau_{AT}^\alpha(u_j)\not\subseteq\partial_{AT}^\alpha(u_i)$. 由定义 C, 有 $A\cap\alpha(u_i, u_j)\neq\varnothing$, 推出矛盾. 则 $\partial_A^\alpha(u_i)\subseteq\partial_{AT}^\alpha(u_i)$. 由性质 12, 易得 $\partial_{AT}^\alpha(u_i)\subseteq\partial_A^\alpha(u_i)$, 故 $\partial_A^\alpha(u_i)=\partial_{AT}^\alpha(u_i)$.

"$R\Rightarrow C$", 设 $v=d(u_j)\notin\partial_{AT}^\alpha(u_i)$, 那么存在 $u_j\in M'$, $M'\in\xi^\alpha(AT)-\xi_{AT}^\alpha(u_i)$. 此时 $\tau_{AT}^\alpha(u_j)\not\subseteq\partial_{AT}^\alpha(u_i)$. 由 $\partial_A^\alpha(u_i)=\partial_{AT}^\alpha(u_i)$, 得 $v=d(u_j)\notin\partial_A^\alpha(u_i)$. 表明属性集 A 中存在属性可以区分对象 u_i 和 u_j, 即 $A\cap\alpha(u_i, u_j)\neq\varnothing$. 　证毕.

由定义 12、定理 5 易得区间值决策系统 ζ 约简的定义和计算方法如下.

定义 **13.** 设 $\zeta=(U, AT\cup D, V, f)$ 是区间值决策系统, $A\subseteq AT$, 若属性子集 A 满足:

1) $\partial^\alpha(A)=\partial^\alpha(AT)$;

2) 不存在任何 $B\subset A$, 满足 $\partial^\alpha(B)=\partial^\alpha(AT)$.

则称属性子集 A 为区间值决策系统 ζ 的一个约简. ζ 中所有约简的集合, 记作 $red^\alpha(AT\cup D)$; 所有约简的交集称为核属性集, 记作 $cor^\alpha(AT\cup D)$.

定理 **6.** 设 $\zeta=(U, AT\cup D, V, f)$ 是区间值决策系统, 若 $A\subseteq AT$ 是 ζ 的一个约简, 当且仅当 $\wedge A$ 是区分函数

$$\varphi^\alpha(AT)=\bigwedge_{\substack{(u_i, M_{AT}^\alpha(u_j))\in U\times\{M_{AT}^\alpha(u_j)\in\\ \xi^\alpha(AT)-\xi_{AT}^\alpha(u_i)\}(\tau_A^\alpha(u_j)\not\subseteq\partial_A^\alpha(u_i))}}\alpha_{AT}\{u_i, M_{AT}^\alpha(u_j)\}$$

的一个素蕴涵.

其中:

$$\alpha_{AT}\{u_i, M_{AT}^\alpha(u_j)\}=\bigwedge_{u_j\in M_{AT}^\alpha(u_j)(d(u_j)\notin\partial_{AT}(u_i))}\vee\alpha(u_i, u_j),$$

$$\alpha(u_i, u_j)=\{a_k\in AT: \alpha_{ij}^k\leqslant\alpha\}.$$

证明略.

定理 **7.** 设区间值决策系统 $\zeta=(U, AT\cup D, V, f)$, 有

$$\varphi^\alpha(AT)=\bigwedge_{u_i\in U}\varphi_{AT}^\alpha(u_i).$$

证明. 由定理 5, 6 易证.

例 **4.** 表 1 中所示的区间值决策系统 $\zeta=(U, AT\cup D, V, f)$ 的知识约简 $\varphi^{0.7}(AT)$ 为:

$$\varphi^{0.7}(AT)=a_1\wedge a_4\wedge(a_2\vee a_5)=$$
$$(a_1\wedge a_2\wedge a_4)\vee(a_1\wedge a_4\wedge a_5),$$
$$red^{0.7}(AT\cup D)=\{\{a_1, a_2, a_4\}, \{a_1, a_4, a_5\}\},$$
$$cor^{0.7}(AT\cup D)=\{a_1, a_2, a_4\}\cap$$
$$\{a_1, a_4, a_5\}=\{a_1, a_4\}.$$

ζ 中关于任意对象 u_i 的约简分别为:

$\varphi_{AT}^{0.7}(u_1)=a_1\wedge(a_2\vee a_5)=(a_1\wedge a_2)\vee(a_1\wedge a_5)$,

$red_{AT\cup D}^{0.7}(u_1)=\{\{a_1, a_2\}, \{a_1, a_5\}\}$;

$\varphi_{AT}^{0.7}(u_2)=a_4$, $red_{AT\cup D}^{0.7}(u_2)=\{\{a_4\}\}$;

$\varphi_{AT}^{0.7}(u_3)=a_2\vee a_5$, $red_{AT\cup D}^{0.7}(u_3)=\{\{a_2\}, \{a_5\}\}$;

$\varphi_{AT}^{0.7}(u_4)=a_3\vee a_4\vee a_5$,

$red_{AT\cup D}^{0.7}(u_4)=\{\{a_3\}, \{a_4\}, \{a_5\}\}$;

$\varphi_{AT}^{0.7}(u_5)=a_1\wedge a_4$, $red_{AT\cup D}^{0.7}(u_5)=\{\{a_1, a_4\}\}$;

$\varphi_{AT}^{0.7}(u_6)=a_1$, $red_{AT\cup D}^{0.7}(u_6)=\{\{a_1\}\}$;

$\varphi_{AT}^{0.7}(u_7) = a_1 \wedge (a_2 \vee a_5) = (a_1 \wedge a_2) \vee (a_1 \wedge a_5)$,
$red_{AT \cup D}^{0.7}(u_7) = \{\{a_1, a_2\}, \{a_1, a_5\}\}$;
$\varphi_{AT}^{0.7}(u_8) = a_4, red_{AT \cup D}^{0.7}(u_8) = \{\{a_4\}\}$;
$\varphi_{AT}^{0.7}(u_9) = a_4, red_{AT \cup D}^{0.7}(u_9) = \{\{a_4\}\}$;
$\varphi_{AT}^{0.7}(u_{10}) = a_1 \vee a_2, red_{AT \cup D}^{0.7}(u_{10}) = \{\{a_1\}, \{a_2\}\}$.

4 结束语

由于客观事物的不确定性以及人类思维的模糊性,现实中的大量数据往往以区间值的形式来表示,因此,如何从区间值信息系统中获取知识逐渐成为一个研究热点.从理论研究的角度来看,相容关系是等价关系的拓广;从应用问题的角度来看,区间值信息系统是单值信息系统的拓广[27-29].文章首先引入了区间值相似率的概念,采用 α-极大相容类对论域 U 进行分类,得到了论域中唯一确定的完全覆盖,并使具有相同属性特征的对象分在同一类中.然后,定义了区间值信息系统的粗糙上下近似算子,并证明其较现有方法提高了粗糙近似精度.最后,给出了区间值信息系统的知识约简定义和方法,实例分析表明了该方法的有效性,为区间值信息系统的知识发现提供了一条思路.

致谢 衷心感谢《计算机研究与发展》审稿人提出的宝贵意见!在本文的撰写、修改过程中,得到了冯琴荣、周杰和高灿同学的无私帮助,在此表示感谢!

参 考 文 献

[1] Pawlak Z. Rough sets [J]. International Journal of Computer and Information Sciences, 1982, 11(5): 341-356

[2] Zhang Wenxiu, Wu Weizhi, Liang Jiye, et al. Rough Set Theory and Approach [M]. Beijing: Science Press, 2001 (in Chinese)
(张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001)

[3] Skowron A, Rauszer C. The discernibility matrices and functions in information system [C] //Intelligent Decision Support Handbook of Applications and Advances of the Rough Sets Theory. Amsterdan, Netherlands: Kluwer Academic Publishers, 1992: 331-362

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

[5] Miao Duoqian, Hu Guirong. A heuristic algorithm for reduction of knowledge [J]. Journal of Computer Research and Development, 1999, 36(6): 681-684 (in Chinese)
(苗夺谦, 胡贵荣. 知识约简的一种启发式算法[J]. 计算机研究与发展, 1999, 36(6): 681-684)

[6] Wang Guoyin. Extension of rough set under incomplete information systems [J]. Journal of Computer Research and Development, 2002, 39(10): 1238-1243 (in Chinese)
(王国胤. Rough 集理论在不完备信息系统中的扩充[J]. 计算机研究与发展, 2002, 39(10): 1238-1243)

[7] Wang Guoyin, Yu Hong, Yang Dachun. Decision table reduction based on conditional information entropy [J]. Chinese Journal of Computers, 2002, 25(7): 759-766 (in Chinese)
(王国胤, 于洪, 杨大春. 基于条件信息熵的决策表约简[J]. 计算机学报, 2002, 25(7): 759-766)

[8] Wang Jue, Wang Ren, Miao Duoqian, et al. Data enriching based on rough set theory [J]. Chinese Journal of Computers, 1998, 21(5): 393-400 (in Chinese)
(王珏, 王任, 苗夺谦, 等. 基于 Rough Set 理论的"数据浓缩"[J]. 计算机学报, 1998, 21(5): 393-400)

[9] Leung Yee, Li Deyu. Maximal consistent block technique for rule acquisition in incomplete information systems [J]. Information Sciences, 2003, 153(1): 85-106

[10] Guan Yanyong, Wang Hongkai. Set-valued information systems [J]. Information Sciences, 2006, 176(17): 2507-2525

[11] Leung Yee, Fischer M, Wu Weizhi, et al. A rough set approach for the discovery of classification rules in interval-valued information systems [J]. International Journal of Approximate Reasoning, 2008, 47(2): 233-246

[12] Wu Shunxiang, Huang Zhiyong, Luo Delin, et al. A grey rough set model based on (α, β)-grey similarity relation [C] //Proc of 2007 IEEE Int Conf on Grey Systems and Intelligent Services. Piscataway, NJ: IEEE, 2007: 18-20

[13] Wu S, Shi S, Liu S, et al. Study of grey rough set model based on tolerance relation [C] //Proc of the 9th Int Conf on Control, Automation, Robotics and Vision. Piscataway, NJ: IEEE, 2006: 1-6

[14] Qian Yuhua, Liang Jiye, Dang Chuangyin. Interval ordered information systems [J]. Computer and Mathematics with Applications, 2008, 56(8): 1994-2009

[15] Wu Qiang, Liu Zongtian. Real formal concept analysis based on grey-rough set theory [J]. Knowledge-Based Systems, 2008, 22(1): 38-45

[16] Li Guodong, Yamaguchi Daisuke, Nagai Masatake. A grey-based rough decision-making approach to supplier selection problem [J]. Mathematical and Computer Modeling, 2007, 46(3/4): 573-581

[17] Yamaguchi Daisuke, Li Guodong, Nagai Masatake. A grey-based rough approximation model for interval data processing [J]. Information Sciences, 2007, 177(21): 4727-4744

[18] Yamaguchi Daisuke, Li Guodong, Nagai Masatake. A grey-rough set approach for interval data reduction of attributes [G] //LNCS 4585: Proc of Int Conf of Rough Sets and Emerging Intelligent Systems Paradigms. Berlin: Springer, 2007: 400-410

[19] Yamaguchi Daisuke, Li Guodong, Nagai Masatake. On the combination of rough set theory and grey theory based on grey lattice operations [G] //LNCS 4259: Proc of the 5th Int Conf on Rough Sets and Current Trends in Computing. Berlin: Springer, 2006: 507- 516

[20] Sengupta A, Pal T. On comparing interval numbers [J]. European Journal of Operational Research, 2000, 127(1): 28- 43

[21] Miao Duoqian, Wang Guoyin, Liu Qing, et al. Granular Computing: Past, Present and Future[M]. Beijing: Science Press, 2007 (in Chinese)
(苗夺谦, 王国胤, 刘清, 等. 粒计算: 过去、现在与展望[M]. 北京: 科学出版社, 2007)

[22] Zheng Zheng. Research on tolerance granular space models and its applications [D]. Beijing: Institute of Computing Technology, Chinese Academy of Sciences, 2006 (in Chinese)
(郑征. 相容粒度空间模型及其应用研究[D]. 北京: 中国科学院计算技术研究所, 2006)

[23] Zheng Zheng, Hu Hong. Shi Zhongzhi. Tolerance relation based granular space [G] //LNCS 3641: Proc of Int Conf of Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. Berlin: Springer, 2005: 682- 691

[24] Zheng Zheng, Hu Hong, Shi Zhongzhi. Tolerance spaces and its applications [C] //Proc of 2005 IEEE Int Conf on Granular Computing. Piscataway, NJ: IEEE, 2005: 367- 372

[25] Kryszkiewicz M. Rough set approach to incomplete information systems [J]. Information Sciences, 1998, 112(1/2/3/4): 39- 49

[26] Kryszkiewicz M. Rules in incomplete information Systems [J]. Information Sciences, 1999, 113(3/4): 271- 292

[27] Yao Yiyu, Liu Qing. A generalized decision logic in interval-set-valued information tables [G] //LNAI 1711: Proc of the 7th Int Workshop on Rough Sets, Fuzzy Sets, Data Mining and Granular Soft Computing. Berlin: Springer, 1999: 285 - 293

[28] Yao Yiyu, Li Xining. Comparison of rough-set and interval-set models for uncertain reasoning [J]. Fundamenta Informaticae, 1996, 27(2/3): 289- 298

[29] Yao Yiyu. Interval-set algebra for qualitative knowledge representation [C] //Proc of the 5th Int Conf of Computing and Information. Piscataway, NJ: IEEE, 1993: 370- 374

Zhang Nan, born in 1979. PhD candidate in the Department of Computer Science and Technology at Tongji University. His main research interests include rough set theory, knowledge spaces, cognitive informatics and artificial intelligence.

张　楠, 1979 年生, 博士研究生, 主要研究方向为粗糙集理论、知识空间、认知信息学与人工智能.

Miao Duoqian, born in 1964. Professor and PhD supervisor in the Department of Computer Science and Technology at Tongji University. He is a senior member of China Computer Federation, and his main research interests include rough set theory, granular computing, principal curve, artificial intelligence, etc.

苗夺谦, 1964 年生, 教授、博士生导师, 中国计算机学会高级会员, 主要研究方向为粗糙集理论、粒计算、主曲线与人工智能等.

Yue Xiaodong, born in 1980. PhD candidate in the Department of Computer Science and Technology at Tongji University. His main research interests include image processing, autonomy-oriented computing, pattern recognition and data mining technology.

岳晓冬, 1980 年生, 博士研究生, 主要研究方向为图像处理、面向自治计算、模式识别与数据挖掘技术.

Research Background

Rough set theory, introduced by Z. Pawlak in 1982, has been conceived as a useful mathematical tool to conceptualize and analyze different types of data. It is an extension of classical set theory for the research of systems marked by uncertain and incomplete information, and has been successfully applied in many fields, such as data classification, medical diagnosis, pattern recognition, image processing, decision analysis, process control, information retrieval, and conflict analysis.

Due to rampant existence of interval numbers in real life, the research on knowledge reduction in interval-valued information systems becomes necessary. In this paper, a new measurement for computing the similarity degree of different interval numbers is proposed. Aiming at these existing problems in current research, we introduce the α-maximal consistent class into interval-valued information systems to improve the accuracy of knowledge classification and rough approximation. The α-maximal consistent class can provide discernibility matrices and discernibility functions in knowledge reduction of interval-valued information systems and decision systems. Finally, the knowledge hidden in interval-valued information systems can be obtained by knowledge reduction.

This research is partially supported by National Natural Science Foundation of China (No. 60775036, No. 60475019) and the Research Foundation of the Doctoral Program of Higher Education of China (No. 20060247039).