软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2012,23(7):1745−1759 [doi: 10.3724/SP.J.1001.2012.04226]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

粗糙集的近似集*

张清华[1,2+], 王国胤[2], 肖 雨[2]

[1](重庆邮电大学 系统理论及其应用研究中心,重庆 400065)
[2](计算智能重庆市重点实验室(重庆邮电大学),重庆 400065)

Approximation Sets of Rough Sets

ZHANG Qing-Hua[1,2+], WANG Guo-Yin[2], XIAO Yu[2]

[1](Research Center for System Theory and Application, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

[2](The Chongqing Key Laboratory of Computational Intelligence (Chongqing University of Posts and Telecommunications), Chongqing 400065, China)

+ Corresponding author: E-mail: zhangqh@cqupt.edu.cn

Zhang QH, Wang GY, Xiao Y. Approximation sets of rough sets. *Journal of Software*, 2012,23(7):1745−1759 (in Chinese). http://www.jos.org.cn/1000-9825/4226.htm

Abstract: Rough sets proposed by professor Pawlak in 1982 is an important tool to process the uncertainty of a set's boundary, and it describes the uncertainty of set X (or concept) with two crisp boundaries that are upper-approximation set and lower-approximation set of X. However, a rough set does not give out the method for precisely, or approximately describe the uncertain set X (or concept) with existing knowledge base. In this paper, the similaritybetween two sets is proposed at first, the disadvantages of using upper-approximation set $\overline{R}(X)$ or lower- approximation set $\underline{R}(X)$ as an approximation set of the uncertain set X (or concept) are analyzed, and then amethod for building an approximation set of the uncertain set X is presented, the conclusion that the set $R_{0.5}(X)$ is the optimal approximation set is proved. Finally, the changing regularities of similarity between $R_{0.5}(X)$ and X with the change of knowledge granulatity in knowledge space are disscussed in detail. From the new viewpoint, this paper presents a new method for building an approximation set of the uncertain set X, and it will promote the development of rough set model.
Key words: rough set; approximation set; granular computing; knowledge space; similarity

摘 要: 粗糙集是 1982 年由 Pawlak 教授提出的解决集合边界不确定的重要方法,它通过两个精确的上、下近似集作为边界线来刻画目标集合(概念)X 的不确定性,但它没有给出如何用已知的知识基(知识粒)来精确或近似地描述边界不确定的目标集合(概念)X 的方法.首先给出了集合之间的相似度概念,然后分析了分别用上近似集 $\overline{R}(X)$ 和下近似集 $\underline{R}(X)$ 作为目标集合(概念)X 近似描述的不足,提出了在已有知识基(粒)空间下寻找目标集合(概念)X 的近似集的方法,并分析了用 $R_{0.5}(X)$作为 X(概念)的近似集的优越性.最后讨论了不同知识粒度空间下 $R_{0.5}(X)$与 X 的相似度随知识粒度的变化关系.从新的角度提出了目标集合(概念)X 近似集的构造方法,促进了粗糙集模型的发展.
关键词: 粗糙集;近似集;粒计算;知识空间;相似度

* 基金项目: 国家自然科学基金(61073146); 重庆市教委科学研究项目(KJ110512, KJ110522); 重庆邮电大学博士启动基金(A2010-06)

收稿时间: 2012-02-13; 定稿时间: 2012-03-27

中图法分类号: TP181 **文献标识码**: A

进入 21 世纪以来,不确定性问题的研究工作受到越来越多的关注[1].如何对不确定性信息和数据进行更加有效的处理,从而发现不确定性信息中蕴涵的知识和规律,是一个重要的研究课题[2],不确定性知识发现的各种不同方法应运而生.Zadeh 在 1965 年提出的模糊集(fuzzy set)理论[3]、Pawlak 在 1982 年提出的粗糙集(rough set)理论[4]和张钹、张铃在 1990 年提出的商空间理论[5]是粒计算(granular computing)的三大基础数学理论,是处理不确定性问题的有效方法.粗糙集(也称为 rough 集或粗集)理论是一种能够定量分析处理不精确、不一致、不完整信息与知识的数学工具,它最初的原型来源于比较简单的信息模型,其基本思想是,通过关系数据库分类归纳形成概念和规则,通过不分明关系(等价关系)的分类以及分类对于目标的近似实现知识发现.粗糙集理论是继概率论、模糊集理论、证据理论之后的又一个处理不确定性问题的数学工具.由于粗糙集理论思想新颖、方法独特、计算简便,因此已成为一种重要的智能信息处理技术[6−8].粗糙集理论的本质思想是,利用不可分辨关系(等价关系)来建立一个划分空间(知识基),在知识粒空间上用两个精确的集合(即上近似集和下近似集)来近似刻画一个边界模糊的目标概念(X).如果知识空间中的知识粒度较粗,被刻画的目标概念的边界域较宽,则近似精度相对较低;如果知识粒度较细,被刻画的目标概念的边界域较窄,则近似精度相对较高.

在给定一个知识基$(U,R)=\{[x]_R|x\in U\}$的条件下,如果 $X=\bigcup\{[x]_R \mid x\in X\}$,则称集合(概念)$X$ 在知识基(U,R)下是精确的,即 X 正好是(U,R)中的一些知识粒的并集;如果 X 不能表示为(U,R)中的知识粒的并集,则称 X 是粗糙的,即在当前知识粒空间下,无法用精确的知识粒表示概念 X.那么在当前的知识粒空间中,如何利用现有知识粒来构建目标概念 X 较好的近似集,是人们进一步关注的问题[2,8].Pawlak 教授试图用两个集合来逼近 X,在他的粗糙集理论模型中,X 的上近似集 $\overline{R}(X)$ 是由在知识空间(U,R)中与 X 有非空交集的知识粒的并集构成的集合,X 的下近似集 $\underline{R}(X)$ 是由在知识空间(U,R)中所有包含于 X 的知识粒的并集构成的集合.因此,$\underline{R}(X)\subseteq X\subseteq\overline{R}(X)$.称 $\underline{R}(X)$ 和 $\overline{R}(X)$ 是不精确概念 X 的两条精确边界线.

换个角度思考,我们是否可以用 $\underline{R}(X)$ 或 $\overline{R}(X)$ 来作为不精确概念 X 的近似集呢?有的研究者分别用 $\underline{R}(X)$ 和 $\overline{R}(X)$ 作为不精确概念 X 的近似概念,可以得到一些近似的规则或知识.特别是一些研究者利用下近似集 $\underline{R}(X)$ 作为目标概念 X 的近似集,即通过寻找正区域来实现规则的获取,并计算出规则的精度或不确定性[9−12].但除了用 $\underline{R}(X)$ 或 $\overline{R}(X)$ 作为 X 的近似集以外,在当前知识基下,是否有更为合适的集合(它也是当前知识基下的知识粒的并集)作为 X 的近似集,该近似集介于 $\underline{R}(X)$ 和 $\overline{R}(X)$ 之间,而近似程度更好?如果有这样的近似集,我们也可以利用这个集合来提取近似的规则,其近似度将会更高.Pawlak 和 Skowron 对粗糙集的扩展型进行了总结与分析[13],很多研究者对此进行了讨论[14,15],比较典型的有变精度粗糙集模型和概率粗糙集模型.米据生和吴伟志等人讨论了变精度粗糙集模型,并利用它进行知识约简[16];张贤勇等人构建了变精度粗糙集的上、下近似算子以及相关的阈值,得到较好的结论[17];Ziarko[18]和 Yao[19−21]等研究者结合概率论和包含度等理论,讨论了概率粗糙集模型,并取得较好的理论模型.但这些方法主要是侧重于构建扩展的Pawlak近似算子,从而利用它们来处理噪声数据.

本文从另外一个角度试图利用当条件属性对论域形成的知识粒空间(知识基)直接构建目标概念 X 的近似集,这个近似集相对于 $\underline{R}(X)$ 或 $\overline{R}(X)$ 可能有更好的近似度.为此,本文首先将粗糙集转化为模糊集,然后针对粗糙集边界域中元素的隶属度的不同,利用截集的方式来构造 X 的近似集,从而用该近似集代替 X 来获取在当前知识基下的近似规则(知识).另外,我们还讨论了随着知识空间中知识粒度的变化、粗糙集的近似集与目标概念 X 的近似度的变化规律.

本文第 1 节介绍相关的基本概念.第 2 节提出粗糙集的相似度.第 3 节给出粗糙集的近似集及其相关性质.第 4 节讨论不同知识粒度下 X 的近似集与 X 的相似度变化关系.第 5 节是结束语.

1 相关基本概念

为了更清楚地阐述本文的思想,首先给出相关的基本概念.

经典 Pawlak 粗糙集模型处理的不分明关系是一种等价关系,建立一对近似算子,即上近似算子和下近似算子(又称为上、下近似集).下面介绍粗糙集理论的几个基本概念.

定义 1(信息表知识表达系统[4,22]). 一个信息表知识表达系统 S 可以表示为 $S=\langle U,A,V,f\rangle$,其中,U 是对象全集,也称为论域;$A=C\cup D$ 是属性全集,子集 C 和 D 分别称为条件属性集和结果属性集;$V=\bigcup_{r\in A}V_r$ 是属性值的集合,V_r 表示属性 $r\in A$ 的属性值范围,即属性 r 的值域;$f:U\times A\to V$ 是一个信息函数,它指定 U 中每一个对象 x 的属性值.

定义 2(不分明关系(indiscernible relations)[4,22]). 对于任一属性集合 $R\subseteq A$,定义一个不可分辨二元关系(不分明关系)$IND(R)=\{(x,y)|(x,y)\in U^2,\forall_{b\in R}(b(x)=b(y))\}$.

定义 3(上近似集(upper approximation set)与下近似集(lower approximation set)[4,22]). 给定信息表知识表达系统 $S=\langle U,A,V,f\rangle$,对于任一对象集合 $X\subseteq U$ 和属性集合 $R\subseteq A$,X 关于 R 的上近似集 $\overline{R}(X)$ 和下近似集 $\underline{R}(X)$ 分别定义如下:

$$\overline{R}(X)=\bigcup\{Y_i\,|\,Y_i\in U/IND(R)\wedge Y_i\cap X\neq\varnothing\},$$
$$\underline{R}(X)=\bigcup\{Y_i\,|\,Y_i\in U/IND(R)\wedge Y_i\subseteq X\},$$

其中,$U/IND(R)=\{X|(X\subseteq U\wedge\forall_{x\in X,y\in X,b\in R}(b(x)=b(y)))\}$是不分明关系 R 在 U 上的划分.

上近似集和下近似集也可以通过集合的形式作如下定义:

$$\underline{R}(X)=\{x\,|\,x\in U\wedge[x]_R\subseteq X\},$$
$$\overline{R}(X)=\{x\,|\,x\in U\wedge[x]_R\cap X\neq\varnothing\},$$

其中,$[x]_R\in U/IND(R)$,即$[x]_R$ 表示对象 x 在不可分辨关系 R 上形成的等价类(划分块).

$\underline{R}(X)$ 表示根据知识 R,U 中一定能归入目标概念(集合)X 的所有对象集合;$\overline{R}(X)$ 表示根据知识 R,U 中可能归入目标概念(集合)X的所有对象集合.集合 $BN_R(X)=\overline{R}(X)-\underline{R}(X)$ 称为目标概念(集合)X关于属性集合 R 的边界域(boundary region); $POS_R(X)=\underline{R}(X)$ 称为目标概念(集合)X 关于属性集合 R 的正域(positive region); $NEG_R(X)=U-\overline{R}(X)$ 称为 X 的负域(negative region).$BN_R(X)$是根据知识 R,U 中既不能肯定归入目标概念(集合)X,又不能肯定归入 $\overline{X}$ (X 的补集)的所有对象集合.

定义 4(可定义集[4,22]). 给定信息表知识表达系统 $S=\langle U,A,V,f\rangle$,对于任一对象子集 $X\subseteq U$ 和属性子集 $R\subseteq A$,当且仅当 $\overline{R}(X)=\underline{R}(X)$ 时,称集合 X 是 R 可定义集.

定义 5(粗糙集(rough 集)[4,22]). 给定信息表知识表达系统 $S=\langle U,A,V,f\rangle$,对于任一对象子集 $X\subseteq U$ 和属性子集 $R\subseteq A$,当且仅当 $\overline{R}(X)\neq\underline{R}(X)$ 时,称集合 X 是 R 粗糙集(rough 集).

由图 1 可知,在一定知识粒度的知识空间中,假设黑线勾勒出的知识表示概念 X,那么概念 X 的下近似集就是由完全属于边界域范围内粒子组成,在图 1 中用黑体部分表示;概念 X 的上近似集由灰色部分和黑色部分组成;概念 X 的边界域由灰色部分组成.由此可以看出,知识粒度越小,则下近似集可能越大,边界域可能越小.

模糊集理论主要是处理边界“含糊”的问题.如图 2 所示,在已知一定知识粒度的情况下,椭圆不能被精确地描述.因此,粗糙集方法通过上、下近似集不断地逼近目标概念(集合),可以很好地解决这一问题.又由于粗糙集是在一定知识粒度下的研究,所以它的集合运算与经典集合存在一些差异.

在一定的知识粒度下,如果集合 X 是粗糙的,则记为 $R(X)$.为了叙述方便,我们令 $R(X)=[\underline{R}(X),\overline{R}(X)]$,这时,$\underline{R}(X)\subseteq X\subseteq\overline{R}(X)$,即我们构造了 X 的两个边界集合.在当前知识粒度情况下,无法用现有知识粒来精确描述 X,因此,X 是不确定的.但是又该如何构造 X 的近似集呢?用 $\underline{R}(X)$ 作为 X 的近似集,还是用 $\overline{R}(X)$ 作为 X 的近似集?除了 $\underline{R}(X)$ 和 $\overline{R}(X)$ 以外,X 有没有更好的近似集?换言之,在一定的知识粒度下,如果 X 是一个目标概念,我们如何寻找其更好的近似概念?面对这些问题,必须重新审视粗糙集对 X 的近似描述原理.粗糙集虽然给出了 X 的近似描述边界,但并没有给出在当前知识粒度下 X 的最佳近似描述方法.本文将粗糙集给出的下近似集和上近似集作为 X 近似描述的两个边界,然后结合粗糙集的模糊性,从模糊集截集的角度来寻找目标概念 X 的更好的近似集.

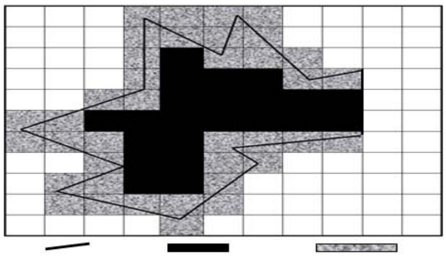


Fig.1 Schematic diagram of rough sets
图 1 粗糙集示意图

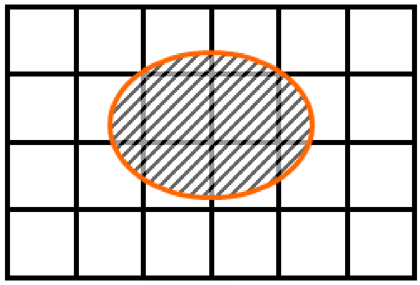


Fig.2 Set under certain knowledge granularity
图 2 一定知识粒度下的集合

定义 6(粗糙集的粗糙度[22]**).** 在一个信息系统中,$IND(R)$是 U 上的一个不可分辨关系,$[x]_R$ 表示对象 x 的等价类,对象子集 $X\subseteq U$,X 的粗糙精度和粗糙度:

- 粗糙精度: $\alpha_R(X)=\dfrac{\underline{R}(X)}{\overline{R}(X)}$;
- 粗糙度: $\rho_R(X)=1-\alpha_R(X)=1-\dfrac{\underline{R}(X)}{\overline{R}(X)}=\dfrac{BN_R(X)}{\overline{R}(X)}$.

显然,对于任意的 $X\subseteq U$,都有 $0\leqslant\alpha_R(X)\leqslant 1$ 且 $0\leqslant\rho_R(X)\leqslant 1$.

- 如果 $\overline{R}(X)=\underline{R}(X)=X$, 此时,$\rho_R(X)=0$(或$\alpha_R(X)=1$),即精确集 X 的粗糙度为 0;
- 如果 $\underline{R}(X)\neq\overline{R}(X)$, 此时,$0<\rho_R(X)\leqslant 1$(或 $0\leqslant\alpha_R(X)<1$),即粗糙集 X 的粗糙度大于 0 且小于等于 1.

为了更好地度量集合之间的相似度,我们给出集合相似度的概念.

定义 7(集合的近似度(或相似度)). 设 A,B 是有限论域 U 上的两个子集,即 $A\subseteq U,B\subseteq U$,定义映射:$S:U\times U\rightarrow[0,1]$,即$(A,B)\rightarrow S(A,B)$.称 $S(A,B)$是集合 A,B 的近似度,如果 $S(A,B)$满足如下条件:

(1) 对任意的 $A,B\in U,0\leqslant S(A,B)\leqslant 1$(有界性);

(2) 对称性:对任意的 $A,B\in U,S(A,B)=S(B,A)$(对称性);

(3) 对任意的 $A,B\in U,S(A,A)=1;S(A,B)=0$ 的充要条件是 $A\cap B=\varnothing$.

对于任意满足条件(1)~条件(3)的公式,都是集合 A,B 的近似度公式.

为简化起见,本文构造近似度公式 $S(A,B)=\dfrac{|A\cap B|}{|A\cup B|}$,其中,$|\cdot|$表示有限子集的元素个数.

显然, $S(A,B)=\dfrac{|A\cap B|}{|A\cup B|}$ 满足定义 3 的 3 个条件.

2 粗糙集的近似集 $R_\lambda(X)$

如果 X 是粗糙的,则 $\underline{R}(X)\subset X\subset\overline{R}(X)$. 因此,在关系 R 下如何寻求得到目标概念(集合)X 的近似集合,是值得探讨的问题.粗糙集给出了 X 的近似的边界集合 $\underline{R}(X)$ 和 $\overline{R}(X)$,但却没有给出 X 的近似集合如何构造.

- 如果直接用 $\underline{R}(X)$ 作为 X 的近似集,则近似度

$$S(X,\underline{R}(X))=\frac{|X\cap\underline{R}(X)|}{|X\cup\underline{R}(X)|}=\frac{|\underline{R}(X)|}{|X|}.$$

- 如果直接用 $\overline{R}(X)$ 作为 X 的近似集,则近似度

$$S(X,\overline{R}(X))=\frac{|X\cap\overline{R}(X)|}{|X\cup\overline{R}(X)|}=\frac{|X|}{|\overline{R}(X)|}.$$

在知识粒空间不变的情况下,是否有更合适的集合作为 X 的近似集呢?即在 $U/IND(R)$知识空间下,是否存在与目标概念(集合)X 近似度更高的集合?本文从集合近似集的角度,在 $U/IND(R)$知识下,利用定义 7 定义的集合

近似度概念和近似度公式,给出了目标概念 X 的近似集,进一步促进了粗糙集理论的发展与完善.

设 U 是非空对象集,知识空间为 $U/IND(R)$,对象子集 $X\subseteq U$,则对于任意的 $x(x\in U)$,x 属于集合 X 的隶属函数为[24,25]

$$\mu_X^R(x)=\frac{|X\cap[x]_R|}{|[x]_R|}.$$

显然,$0\leqslant\mu_X^R(x)\leqslant 1$,它表示任意一个元素 x 属于集合 X 的程度.

设 $X=\{x_1,x_2,\dots,x_n\}$,令 $F_X^R=\{\mu_X^R(x_1),\mu_X^R(x_2),\dots,\mu_X^R(x_n)\}$,则 F_X^B 是集合 U 上的一个模糊集(即 $F_X^R\in F(U)$).在该模糊集上,位于同一等价类(分块)的两个元素的隶属度值是相同的.

由粗糙集上、下近似和边界的概念,不难得出:

$$\underline{R}(X)=\{x\,|\,x\in U\wedge\mu_X^R(x)=1\},$$
$$\overline{R}(X)=\{x\,|\,x\in U\wedge 0<\mu_X^R(x)\leqslant 1\},$$
$$BN_R(X)=\overline{R}(X)-\underline{R}(X)=\{x\,|\,x\in U\wedge 0<\mu_X^R(x)<1\}.$$

由此可以看出,对于粗糙集的边界域上的元素,它们对集合 X 的隶属程度不同,有的元素的隶属度非常小,可能接近于0;有的元素的隶属度非常大,可能接近于1.隶属度为1的元素在正区域(或下近似集)中,而隶属度小于1且大于0的元素位于粗糙集的边界域上.这时,我们根据隶属度的大小,通过模糊集截集的方法将其分为两个部分,从而构建一个介于上近似集和下近似集之间的集合 $R_\lambda(X)$.

定义 8(X 的λ近似集). 设 X 是论域的一个集合(目标概念),令

$$R_\lambda(X)=\{x\in U\,|\,\mu_X^R(x)\geqslant\lambda\},\ 1\geqslant\lambda>0,$$

称 $R_\lambda(X)$为 X 的λ近似集.

定义 9(X 的强λ近似集). 设 X 是论域的一个集合(目标概念),令

$$\underset{\cdot}{R}_\lambda(X)=\{x\in U\,|\,\mu_X^R(x)>\lambda\},\ 1>\lambda>0,$$

称 $\underset{\cdot}{R}_\lambda(X)$ 为 X 的强λ近似集.

显然,$\underset{\cdot}{R}_\lambda(X)\subseteq R_\lambda(X)$.用 $R_\lambda(X)$作为 X 的近似集会有何性质和特点?显然,$R_1(X)=\underline{R}(X)$,$R_0(X)=\overline{R}(X)$,即粗糙集定义的上、下近似集是 $R_\lambda(X)$的两个特例,且 $\underline{R}(X)\subseteq R_\lambda(X)\subseteq\overline{R}(X)$.若 $1\geqslant\lambda_1>\lambda_2>0$,显然有

$$\underline{R}(X)\subseteq R_{\lambda_1}(X)\subseteq R_{\lambda_2}(X)\subseteq\overline{R}(X).$$

为了更加直观地理解 $R_\lambda(X)$,并且结合人们的认知直观性更好地研究 X 的近似集,我们通常取λ=0.5.因为λ=0.5 时,属于 X 的隶属度大于等于 0.5(即超过半数以上的元素属于概念 X),符合人们的认知习惯,如图 3 所示.

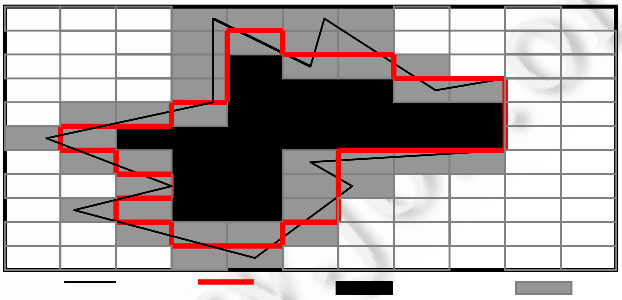


Fig.3 Relation chart among the sets X, $R_{0.5}(X)$, $\underline{R}(X)$ and $\overline{R}(X)$

图 3 X,$R_{0.5}(X)$, $\underline{R}(X)$ 和 $\overline{R}(X)$ 之间的关系图

接下来,我们重点研究当λ=0.5 时近似集 $R_{0.5}(X)$与 X 的近似度.

3 粗糙集的近似集 $R_{0.5}(X)$

从图 3 可以看出,$R_{0.5}(X)$介于 $\underline{R}(X)$ 和 $\overline{R}(X)$ 之间,即 $\underline{R}(X)\subseteq R_{0.5}(X)\subseteq \overline{R}(X)$.下面我们来讨论 $R_{0.5}(X)$的运算性质及它与目标概念(集合)X 近似程度的一些性质.

3.1 $R_{0.5}(X)$的运算性质

粗糙集的上、下近似集有如下性质[22,23]:

(1) $\underline{R}(\sim X)=\sim\overline{R}(X),\overline{R}(\sim X)=\sim\underline{R}(X)$;

(2) 若 $X\subseteq Y$,则 $\underline{R}(X)\subseteq\underline{R}(Y),\overline{R}(X)\subseteq\overline{R}(Y)$;

(3) $\underline{R}(X\cup Y)\supseteq\underline{R}(X)\cup\underline{R}(Y),\overline{R}(X\cup Y)=\overline{R}(X)\cup\overline{R}(Y)$;

(4) $\underline{R}(X\cap Y)=\underline{R}(X)\cap\underline{R}(Y),\overline{R}(X\cap Y)\subseteq\overline{R}(X)\cup\overline{R}(Y)$.

那么,$R_{0.5}(X)$对集合的交、并、补运算有类似性质.

性质 1. 设 X,Y 是两个集合(概念),则:

(1) $\underset{\cdot}{R}_{0.5}(\sim X)=\sim R_{0.5}(X)$;

(2) 若 $X\subseteq Y$,则 $R_{0.5}(X)\subseteq R_{0.5}(Y)$;

(3) $R_{0.5}(X\cup Y)\supseteq R_{0.5}(X)\cup R_{0.5}(Y)$;

(4) $R_{0.5}(X\cap Y)\subseteq R_{0.5}(X)\cap R_{0.5}(Y)$.

证明:

(1) 因为

$$\underset{\cdot}{R}_{0.5}(\sim X)=\{x\mid \mu_X^R(x)>0.5\}=\left\{x\left|\frac{|[x]_R\cap(\sim X)|}{|[x]_R|}>0.5\right.\right\}=\left\{x\left|\frac{|[x]_R\cap(U-X)|}{|[x]_R|}>0.5\right.\right\}=$$

$$\left\{x\left|\left(1-\frac{|([x]_R\cap X)|}{|[x]_R|}\right)>0.5\right.\right\}=\left\{x\left|\frac{|[x]_R\cap X|}{|[x]_R|}\leqslant 0.5\right.\right\}=\sim R_{0.5}(X).$$

因此,性质 1(1)成立.值得注意的是,$R_{0.5}(\sim X)=\sim R_{0.5}(X)$ 不一定成立,读者可以自己证明;

(2) 对于元素 $x\in R_{0.5}(X)$,有 $x\in\left\{x\left|\frac{|[x]_R\cap X|}{|[x]_R|}\geqslant 0.5\right.\right\}\subseteq X$; 又因为 $X\subseteq Y$,有 $\frac{|[x]_R\cap X|}{|[x]_R|}\leqslant\frac{|[x]_R\cap Y|}{|[x]_R|}$.故

$$x\in\left\{x\left|\frac{|[x]_R\cap Y|}{|[x]_R|}\geqslant 0.5\right.\right\}\subseteq Y.$$

所以,$x\in R_{0.5}(Y)$.因此,$R_{0.5}(X)\subseteq R_{0.5}(Y)$;

(3) 因 $X\subseteq X\cup Y$ 且 $Y\subseteq X\cup Y$,由性质 1(2)可得,$R_{0.5}(X)\subseteq R_{0.5}(X\cup Y)$且 $R_{0.5}(Y)\subseteq R_{0.5}(X\cup Y)$.所以,

$$R_{0.5}(X\cup Y)\supseteq R_{0.5}(X)\cup R_{0.5}(Y).$$

值得指出的是,$R_{0.5}(X\cup Y)\subseteq R_{0.5}(X)\cup R_{0.5}(Y)$不一定成立;

(4) 因 $X\cap Y\subseteq X$ 且 $X\cap Y\subseteq Y$,由性质 1(2)可得,$R_{0.5}(X\cap Y)\subseteq R_{0.5}(X)$且 $R_{0.5}(X\cap Y)\subseteq R_{0.5}(Y)$.所以,

$$R_{0.5}(X\cap Y)\subseteq R_{0.5}(X)\cap R_{0.5}(Y).$$

值得指出的是,$R_{0.5}(X\cap Y)\supseteq R_{0.5}(X)\cap R_{0.5}(Y)$也不一定成立. □

3.2 $R_{0.5}(X)$与X的近似性(相似性)

在讨论目标概念(集合)X 的近似集 $R_{0.5}(X)$及其性质之前,首先给出几个数学上的基本结论,便于证明相关后继定理.

引理 1. 设 a,b,c 和 d 均是实数,且 $0<a<b,0<c<d$,则 $\frac{a}{b}<\frac{a+d}{b+c}$.

证明:因为 $0<a<b,0<c<d$,所以 $ac<bd$,因此 $ab+ac<ab+bd,a(b+c)<b(a+d)$,即 $\frac{a}{b}<\frac{a+d}{b+c}$. □

引理 2. 设 a,b,c 和 d 均是实数,且 $0<c<a,0<d<b$,若 $\frac{a}{b}\geqslant\frac{c}{d}$,则 $\frac{a}{b}\leqslant\frac{a-c}{b-d}$;若 $\frac{a}{b}\leqslant\frac{c}{d}$,则 $\frac{a}{b}\geqslant\frac{a-c}{b-d}$.

证明:因为 $0<c<a,0<d<b$,若 $\frac{a}{b}\geqslant\frac{c}{d}$,则 $ad\geqslant bc$,因此,$ab-ad\leqslant ab-bc$,$a(b-d)\leqslant b(a-c)$,所以 $\frac{a}{b}\leqslant\frac{a-c}{b-d}$. 同理,若 $\frac{a}{b}\leqslant\frac{c}{d}$,则 $\frac{a}{b}\geqslant\frac{a-c}{b-d}$. □

为了更加详细地了解 $R_{0.5}(X)$与 X 的近似度(相似度),我们给出定理 1 和定理 2.

定理 1. 设论域 U 是一个有限论域,$X\subseteq U$,R 是 U 上的等价关系,则 $S(X,R_{0.5}(X))\geqslant S(X,\underline{R}(X))$.

证明:对任意 $x\in R_{0.5}(X)$,有 $\mu_X^R(x)\geqslant 0.5$, 即 $\mu_X^R(x)=\frac{|[x]_R\cap X|}{|[x]_R|}\geqslant 0.5$. 因为 R 是 U 上的等价关系,由 R 在 U 上形成的等价类记为$[x_1]_R,[x_2]_R,\ldots,[x_n]_R$,则

$$R_{0.5}(X)=\{x\,|\,\mu_X^R(x)\geqslant 0.5\}=\{x\,|\,\mu_X^R(x)=1\}\cup\{x\,|\,0.5\leqslant\mu_X^R(x)<1\}.$$

显然,$\{x\,|\,\mu_X^R(x)=1\}=\underline{R}(X)$. 而令 $\{x\,|\,0.5\leqslant\mu_X^R(x)<1\}=[x_{i_1}]_R\cup[x_{i_2}]_R\cup\ldots\cup[x_{i_k}]_R$,所以,

$$X\cap R_{0.5}(X)=X\cap(\underline{R}(X)\cup[x_{i_1}]_R\cup[x_{i_2}]_R\cup\ldots\cup[x_{i_k}]_R).$$

又因为 $\underline{R}(X),[x_{i_1}]_R,[x_{i_2}]_R,\ldots,[x_{i_k}]_R$ 中任意两个集合的交集为空集,因此,

$$\begin{aligned}|X\cap R_{0.5}(X)|&=|X\cap\underline{R}(X)|+|X\cap[x_{i_1}]_R|+|X\cap[x_{i_2}]_R|+\ldots+|X\cap[x_{i_k}]_R|\\&=|\underline{R}(X)|+|X\cap[x_{i_1}]_R|+|X\cap[x_{i_2}]_R|+\ldots+|X\cap[x_{i_k}]_R|.\end{aligned}$$

而 $X\cup R_{0.5}(X)=X\cup([x_{i_1}]_R-X)\cup([x_{i_2}]_R-X)\cup\ldots\cup([x_{i_k}]_R-X)$,

同时, $X,([x_{i_1}]_R-X),([x_{i_2}]_R-X),\ldots,([x_{i_k}]_R-X)$ 中任意两个交集为空集,所以,

$$|X\cup R_{0.5}(X)|=|X|+|([x_{i_1}]_R-X)|+|([x_{i_2}]_R-X)|+\ldots+|([x_{i_k}]_R-X)|.$$

因此,

$$S(X,R_{0.5}(X))=\frac{|X\cap R_{0.5}(X)|}{|X\cup R_{0.5}(X)|}=\frac{|\underline{R}(X)|+|X\cap[x_{i_1}]_R|+|X\cap[x_{i_2}]_R|+\ldots+|X\cap[x_{i_k}]_R|}{|X|+|([x_{i_1}]_R-X)|+|([x_{i_2}]_R-X)|+\ldots+|([x_{i_k}]_R-X)|}.$$

又因为 $\mu_X^R(x_{i_1})=\frac{|[x_{i_1}]_R\cap X|}{|[x_{i_1}]_R|}=\frac{|[x_{i_1}]_R\cap X|}{|[x_{i_1}]_R\cap X|+|[x_{i_1}]_R-X|}\geqslant 0.5$, 所以, $|[x_{i_1}]_R\cap X|\geqslant|[x_{i_1}]_R-X|$;

同理, $|[x_{i_2}]_R\cap X|\geqslant|[x_{i_2}]_R-X|,\ldots,|[x_{i_k}]_R\cap X|\geqslant|[x_{i_k}]_R-X|$. 由引理 1 容易得到:

$$S(X,R_{0.5}(X))=\frac{|\underline{R}(X)|+|X\cap[x_{i_1}]_R|+|X\cap[x_{i_2}]_R|+\ldots+|X\cap[x_{i_k}]_R|}{|X|+|([x_{i_1}]_R-X)|+|([x_{i_2}]_R-X)|+\ldots+|([x_{i_k}]_R-X)|}\geqslant\frac{|\underline{R}(X)|}{|X|}=S(X,\underline{R}(X)).$$ □

定理 1 表明,用 $R_{0.5}(X)$作为目标概念 X 的近似集,比用 $\underline{R}(X)$ 作为目标概念的近似集的近似度更高.接下来讨论 $R_{0.5}(X)$和 $\overline{R}(X)$ 的目标概念 X 的近似度.

在 $\overline{R}(X)$ 中存在一些知识粒,它们都不包含于 $R_{0.5}(X)$中,不妨设 $\overline{R}(X)-R_{0.5}(X)=[x_{j_1}]_R\cup[x_{j_2}]_R\cup\ldots\cup[x_{j_s}]_R$,其中, $[x_{j_1}]_R,[x_{j_2}]_R,\ldots,[x_{j_s}]_R$ 两两不相交.而这些知识粒都与 X 有非空交集,不难得出:

$$([x_{j_1}]_R\cap X)\cup([x_{j_2}]_R\cap X)\cup\ldots\cup([x_{j_s}]_R\cap X)=X-R_{0.5}(X).$$

并且, $([x_{j_1}]_R-X)\cup([x_{j_2}]_R-X)\cup\ldots\cup([x_{j_s}]_R-X)=\overline{R}(X)-R_{0.5}(X)-X$.

定理 2. 设论域 U 是一个有限论域,$X\subseteq U$,R 是 U 上的等价关系,若 $\frac{|X|}{|\overline{R}(X)|}>\frac{|X-R_{0.5}(X)|}{|\overline{R}(X)-R_{0.5}(X)-X|}$,则

$$S(X,R_{0.5}(X))\geqslant S(X,\overline{R}(X)).$$

证明:设 $\overline{R}(X)-R_{0.5}(X)=[x_{j_1}]_R\cup[x_{j_2}]_R\cup\ldots\cup[x_{j_s}]_R$,其中 $[x_{j_1}]_R,[x_{j_2}]_R,\ldots,[x_{j_s}]_R$ 两两不相交.

因为 $0<\mu_X^R(x_{j_1})=\frac{|[x_{j_1}]_R\cap X|}{|[x_{j_1}]_R|}<0.5$,所以 $[x_{j_1}]_R\cap X\neq\varnothing$;同理, $[x_{j_2}]_R\cap X\neq\varnothing,\ldots,[x_{j_s}]_R\cap X\neq\varnothing$. 所以,

$$X\cap R_{0.5}(X)=X-([x_{j_1}]_R\cap X)-([x_{j_2}]_R\cap X)-\ldots-([x_{j_s}]_R\cap X)=X-(X-R_{0.5}(X)).$$

又因为$[x_{j_1}]_R\cap X,[x_{j_2}]_R\cap X,...,[x_{j_s}]_R\cap X$两两交集为空,所以,

$$X\cap R_{0.5}(X)=X-(X-R_{0.5}(X)),|X\cap R_{0.5}(X)|=|X|-|X-R_{0.5}(X)|.$$

且$X\cup R_{0.5}(X)=\bar{R}(X)-(([x_{j_1}]_R-X)\cup([x_{j_2}]_R-X)\cup...\cup([x_{j_s}]_R-X))$.

又因为$([x_{j_1}]_R-X),([x_{j_2}]_R-X),...,([x_{j_s}]_R-X)$两两交集为空,所以,

$$X\cup R_{0.5}(X)=\bar{R}(X)-(\bar{R}(X)-R_{0.5}(X)-X),|X\cup R_{0.5}(X)|=|\bar{R}(X)|-|(\bar{R}(X)-R_{0.5}(X)-X)|.$$

因此,

$$S(X,R_{0.5}(X))=\frac{|X\cap R_{0.5}(X)|}{|X\cup R_{0.5}(X)|}=\frac{|X|-|X-R_{0.5}(X)|}{|\bar{R}(X)|-|(\bar{R}(X)-R_{0.5}(X)-X)|},$$

又$\frac{|X|}{|\bar{R}(X)|}>\frac{|X-R_{0.5}(X)|}{|\bar{R}(X)-R_{0.5}(X)-X|}$,根据引理2,$\frac{|X|}{|\bar{R}(X)|}\leqslant\frac{|X|-|X-R_{0.5}(X)|}{|\bar{R}(X)|-|(\bar{R}(X)-R_{0.5}(X)-X)|}$成立,即

$$S(X,R_{0.5}(X))\geqslant S(X,\bar{R}(X)).$$ □

由定理2可以看出,用$R_{0.5}(X)$作为目标概念X的近似集,在一定条件下比用$\bar{R}(X)$作为X的近似集效果更好.即,只要$\frac{|X|}{|\bar{R}(X)|}>\frac{|X-R_{0.5}(X)|}{|\bar{R}(X)-R_{0.5}(X)-X|}$成立,则$S(X,R_{0.5}(X))\geqslant S(X,\bar{R}(X))$成立.

为了更清楚地了解在阈值$1\geqslant\lambda>0.5$时,$S(X,R_{0.5}(X)),S(X,R_\lambda(X))$和$S(X,\underline{R}(X))$三者之间的关系,我们给出定理3和定理4.

定理3. 设论域U是一个有限论域,$X\subseteq U$,R是U上的等价关系,对任意的$1\geqslant\lambda>0.5$,则有

$$S(X,R_{0.5}(X))\geqslant S(X,R_\lambda(X))\geqslant S(X,\underline{R}(X)).$$

证明:对任意$x\in R_{0.5}(X)$,有$\mu_X^R(x)\geqslant 0.5$,即$\mu_X^R(x)=\frac{|[x]_R\cap X|}{|[x]_R|}\geqslant 0.5$.

$$R_{0.5}(X)=\{x\,|\,\mu_X^R(x)\geqslant 0.5\}=\{x\,|\,\mu_X^R(x)=1\}\cup\{x\,|\,0.5\leqslant\mu_X^R(x)<1\}.$$

显然,$\{x\,|\,\mu_X^R(x)=1\}=\underline{R}(X)$.而令

$$\{x\,|\,0.5\leqslant\mu_X^R(x)<1\}=[x_{i_1}]_R\cup[x_{i_2}]_R\cup...\cup[x_{i_k}]_R,R_\lambda(X)=\{x\,|\,\mu_X^R(x)\geqslant\lambda>0.5\}$$
$$=\{x\,|\,\mu_X^R(x)=1\}\cup\{x\,|\,0.5<\lambda\leqslant\mu_X^R(x)<1\}.$$

显然,$R_\lambda(X)\subseteq R_{0.5}$.为了便于讨论,我们令$\{x\,|\,0.5<\lambda\leqslant\mu_X^R(x)<1\}=[x_{i_1}]_R\cup[x_{i_2}]_R\cup...\cup[x_{i_q}]_R$.这里,$q\leqslant k$.所以,

$$X\cap R_\lambda(X)=X\cap(\underline{R}(X)\cup[x_{i_1}]_R\cup[x_{i_2}]_R\cup...\cup[x_{i_q}]_R).$$

又因为$\underline{R}(X),[x_{i_1}]_R,[x_{i_2}]_R,...,[x_{i_k}]_R$中任意两个集合的交集为空集,因此,

$$|X\cap R_{0.5}(X)|=|X\cap\underline{R}(X)|+|X\cap[x_{i_1}]_R|+|X\cap[x_{i_2}]_R|+...+|X\cap[x_{i_k}]_R|$$
$$=|\underline{R}(X)|+|X\cap[x_{i_1}]_R|+|X\cap[x_{i_2}]_R|+...+|X\cap[x_{i_k}]_R|,$$

而

$$|X\cap R_\lambda(X)|=|X\cap\underline{R}(X)|+|X\cap[x_{i_1}]_R|+|X\cap[x_{i_2}]_R|+...+|X\cap[x_{i_q}]_R|$$
$$=|\underline{R}(X)|+|X\cap[x_{i_1}]_R|+|X\cap[x_{i_2}]_R|+...+|X\cap[x_{i_q}]_R|.$$

因此,

$$S(X,R_{0.5}(X))=\frac{|X\cap R_{0.5}(X)|}{|X\cup R_{0.5}(X)|}$$
$$=\frac{|\underline{R}(X)|+|X\cap[x_{i_1}]_R|+|X\cap[x_{i_2}]_R|+...+|X\cap[x_{i_q}]_R|+|X\cap[x_{i_{q+1}}]_R|+...+|X\cap[x_{i_k}]_R|}{|X|+|([x_{i_1}]_R-X)|+|([x_{i_2}]_R-X)|+...+|([x_{i_q}]_R-X)|+|([x_{i_{q+1}}]_R-X)|+...+|([x_{i_k}]_R-X)|}.$$

于是,

$$S(X,R_\lambda(X))=\frac{|X\cap R_\lambda(X)|}{|X\cup R_\lambda(X)|}=\frac{|\underline{R}(X)|+|X\cap[x_{i_1}]_R|+|X\cap[x_{i_2}]_R|+...+|X\cap[x_{i_q}]_R|}{|X|+|([x_{i_1}]_R-X)|+|([x_{i_2}]_R-X)|+...+|([x_{i_q}]_R-X)|}.$$

又因为 $|X\cap[x_{i_{q+1}}]_R|+...+|X\cap[x_{i_k}]_R|\geqslant|([x_{i_{q+1}}]_R-X)|+...+|([x_{i_k}]_R-X)|$,根据引理 1,

$$S(X,R_{0.5}(X))\geqslant S(X,R_\lambda(X))\geqslant S(X,\underline{R}(X)).$$

证毕. □

由定理 2 和定理 3,容易得到推论 1.

推论 1. 设论域 U 是一个有限论域,$X\subseteq U$,R 是 U 上的等价关系.对任意的 $1\geqslant\lambda>0.5$,如果

$$\frac{|X|}{|\overline{R}(X)|}>\frac{|X-R_\lambda(X)|}{|\overline{R}(X)-R_\lambda(X)-X|},$$

则 $S(X,R_{0.5}(X))\geqslant S(X,R_\lambda(X))\geqslant S(X,\overline{R}(X))$.

定理 4. 设论域 U 是一个有限论域,$X\subseteq U$,R 是 U 上的等价关系.对任意的 $0.5\leqslant\lambda_1<\lambda_2\leqslant 1$,则有

$$S(X,R_{\lambda_1}(X))\geqslant S(X,R_{\lambda_2}(X)).$$

证明:对任意 $x\in R_{\lambda_1}(X)$,有 $\mu_X^R(x)\geqslant\lambda_1$,即 $\mu_X^R(x)=\dfrac{|[x]_R\cap X|}{|[x]_R|}\geqslant\lambda_1\geqslant 0.5$.因为 R 是 U 上的等价关系,设 R 在 U 上形成的等价类记为 $[x_1]_R,[x_2]_R,...,[x_n]_R$.

$R_{\lambda_1}(X)=\{x\,|\,\mu_X^R(x)\geqslant\lambda_1\}=\{x\,|\,\mu_X^R(x)=1\}\cup\{x\,|\,\lambda_1\leqslant\mu_X^R(x)<1\}$,由定理 1 可知 $\{x\,|\,\mu_X^R(x)=1\}=\underline{R}(X)$,而令

$$\{x\,|\,\lambda_1\leqslant\mu_X^R(x)<1\}=[x_{i_1}]_R\cup[x_{i_2}]_R\cup...\cup[x_{i_u}]_R.$$

同理,$R_{\lambda_2}(X)=\{x\,|\,\mu_X^R(x)\geqslant\lambda_2\}=\{x\,|\,\mu_X^R(x)=1\}\cup\{x\,|\,\lambda_2\leqslant\mu_X^R(x)<1\}$.令

$$\{x\,|\,\lambda_2\leqslant\mu_X^R(x)<1\}=[x_{i_1}]_R\cup[x_{i_2}]_R\cup...\cup[x_{i_v}]_R.$$

这里,因为 $0.5\leqslant\lambda_1<\lambda_2\leqslant 1$,所以,$v\leqslant u$,则

$$X\cap R_{\lambda_1}(X)=X\cap(\underline{R}(X)\cup[x_{i_1}]_R\cup[x_{i_2}]_R\cup...\cup[x_{i_v}]_R\cup...\cup[x_{i_u}]_R),$$
$$X\cap R_{\lambda_2}(X)=X\cap(\underline{R}(X)\cup[x_{i_1}]_R\cup[x_{i_2}]_R\cup...\cup[x_{i_l}]_R).$$

又因为 $\underline{R}(X),[x_{i_1}]_R,[x_{i_2}]_R,...,[x_{i_u}]_R$ 中任意两个集合的交集为空集,因此,

$$|X\cap R_{\lambda_1}(X)|=|X\cap\underline{R}(X)|+|X\cap[x_{i_1}]_R|+|X\cap[x_{i_2}]_R|+...+|X\cap[x_{i_v}]_R|+...+|X\cap[x_{i_u}]_R|$$
$$=|\underline{R}(X)|+|X\cap[x_{i_1}]_R|+|X\cap[x_{i_2}]_R|+...+|X\cap[x_{i_v}]_R|+...+|X\cap[x_{i_u}]_R|,$$
$$|X\cap R_{\lambda_2}(X)|=|\underline{R}(X)|+|X\cap[x_{i_1}]_R|+|X\cap[x_{i_2}]_R|+...+|X\cap[x_{i_v}]_R|.$$

而 $X\cup R_{\lambda_1}(X)=X\cup([x_{i_1}]_R-X)\cup([x_{i_2}]_R-X)\cup...\cup([x_{i_v}]_R-X)\cup...\cup([x_{i_u}]_R-X)$,

同时,$X,([x_{i_1}]_R-X),([x_{i_2}]_R-X),...,([x_{i_v}]_R-X),...,([x_{i_u}]_R-X)$ 中任意两个交集为空集,所以,

$$|X\cup R_{\lambda_1}(X)|=|X|+|([x_{i_1}]_R-X)|+|([x_{i_2}]_R-X)|+...+|([x_{i_v}]_R-X)|+...+|([x_{i_u}]_R-X)|,$$
$$|X\cup R_{\lambda_2}(X)|=|X|+|([x_{i_1}]_R-X)|+|([x_{i_2}]_R-X)|+...+|([x_{i_v}]_R-X)|.$$

因此,

$$S(X,R_{\lambda_1}(X))=\frac{|X\cap R_{\lambda_1}(X)|}{|X\cup R_{\lambda_1}(X)|}=\frac{|\underline{R}(X)|+|X\cap[x_{i_1}]_R|+...+|X\cap[x_{i_v}]_R|+|X\cap[x_{i_{v+1}}]_R|+...+|X\cap[x_{i_u}]_R|}{|X|+|([x_{i_1}]_R-X)|+...+|([x_{i_v}]_R-X)|+|([x_{i_{v+1}}]_R-X)|+...+|([x_{i_u}]_R-X)|},$$
$$S(X,R_{\lambda_2}(X))=\frac{|X\cap R_{\lambda_2}(X)|}{|X\cup R_{\lambda_2}(X)|}=\frac{|\underline{R}(X)|+|X\cap[x_{i_1}]_R|+...+|X\cap[x_{i_v}]_R|}{|X|+|([x_{i_1}]_R-X)|+...+|([x_{i_v}]_R-X)|}.$$

因为 $\mu_X^R(x_{i_{v+1}})=\dfrac{|[x_{i_{v+1}}]_R\cap X|}{|[x_{i_{v+1}}]_R|}=\dfrac{|[x_{i_{v+1}}]_R\cap X|}{|[x_{i_{v+1}}]_R\cap X|+|[x_{i_{v+1}}]_R-X|}\geqslant\lambda_1\geqslant 0.5$,所以,$|[x_{i_{v+1}}]_R\cap X|\geqslant|[x_{i_{v+1}}]_R-X|$.

同理,$|[x_{i_{v+2}}]_R\cap X|\geqslant|[x_{i_{v+2}}]_R-X|,...,|[x_{i_u}]_R\cap X|\geqslant|[x_{i_u}]_R-X|$,所以,

$$|X\cap[x_{i_{v+1}}]_R|+...+|X\cap[x_{i_u}]_R|\geqslant|([x_{i_{v+1}}]_R-X)|+...+|([x_{i_u}]_R-X)|.$$

由引理 1,我们容易得到:

$$S(X,R_{\lambda_2}(X))=\frac{|X\cap R_{\lambda_2}(X)|}{|X\cup R_{\lambda_2}(X)|}=\frac{|\underline{R}(X)|+|X\cap[x_{i_1}]_R|+...+|X\cap[x_{i_v}]_R|}{|X|+|([x_{i_1}]_R-X)|+...+|([x_{i_v}]_R-X)|}\leqslant$$
$$\frac{|\underline{R}(X)|+|X\cap[x_{i_1}]_R|+...+|X\cap[x_{i_v}]_R|+|X\cap[x_{i_{v+1}}]_R|+...+|X\cap[x_{i_u}]_R|}{|X|+|([x_{i_1}]_R-X)|+...+|([x_{i_v}]_R-X)|+|([x_{i_{v+1}}]_R-X)|+...+|([x_{i_u}]_R-X)|}=S(X,R_{\lambda_1}(X)).$$ □

定理 4 表明,当阈值λ大于 0.5 时,阈值越大,X 与 $R_\lambda(X)$的近似程度越小.

即,当 $0.5\leqslant\lambda_1<\lambda_2\leqslant 1$ 时, $S(X,R_{\lambda_1}(X))\geqslant S(X,R_{\lambda_2}(X))$.

3.3 基于$R_{0.5}(X)$提取规则的实例分析

为了更清晰地说明如何用$R_{0.5}(X)$提取规则,我们以决策信息表(见表 1)为例来加以分析.这里,对象集 $U=\{x_1, x_2,\dots,x_{16}\}$,条件属性集 $C=\{a,b,c\}$,决策属性集 $D=\{d\}$.

Table 1 Decision information table

表 1 决策信息表

	a	b	c	d
x_1	1	1	0	1
x_2	1	1	0	1
x_3	1	1	2	1
x_4	1	1	2	1
x_5	1	1	2	1
x_6	2	0	1	1
x_7	1	1	2	2
x_8	2	0	1	2
x_9	1	0	1	2
x_{10}	1	0	1	2
x_{11}	3	2	1	2
x_{12}	3	2	1	2
x_{13}	2	0	1	3
x_{14}	3	2	2	3
x_{15}	3	2	1	3
x_{16}	3	2	2	3

根据粗糙集理论的基本知识,易知:

$U/IND(C)=\{\{x_1,x_2\},\{x_3,x_4,x_5,x_7\},\{x_6,x_8,x_{13}\},\{x_9,x_{10}\},\{x_{11},x_{12},x_{15}\},\{x_{14},x_{16}\}\}$,

$U/IND(D)=\{\{x_1,x_2,x_3,x_4,x_5,x_6\},\{x_7,x_8,x_9,x_{10},x_{11},x_{12}\},\{x_{13},x_{14},x_{15},x_{16}\}\}$.

令决策属性形成的 3 个决策概念分别是 $X_1=\{x_1,x_2,x_3,x_4,x_5,x_6\}$,$X_2=\{x_7,x_8,x_9,x_{10},x_{11},x_{12}\}$和 $X_3=\{x_{13},x_{14},x_{15},x_{16}\}$.因此,这 3 个决策概念在知识粒空间(条件属性形成的划分空间,$U/IND(C)$)下的下近似集、上近似集和近似集分别是

$$\underline{R}(X_1)=\{x_1,x_2\},$$
$$\underline{R}(X_2)=\{x_9,x_{10}\},$$
$$\underline{R}(X_3)=\{x_{14},x_{16}\};$$
$$\overline{R}(X_1)=\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_{13}\},$$
$$\overline{R}(X_2)=\{x_3,x_4,x_5,x_6,x_7,x_8,x_9,x_{10},x_{11},x_{12},x_{13},x_{15}\},$$
$$\overline{R}(X_3)=\{x_6,x_8,x_{11},x_{12},x_{13},x_{14},x_{15},x_{16}\};$$
$$R_{0.5}(X_1)=\{x_1,x_2,x_3,x_4,x_5,x_7\},$$
$$R_{0.5}(X_2)=\{x_9,x_{10},x_{11},x_{12},x_{15}\},$$
$$R_{0.5}(X_3)=\{x_{14},x_{16}\}.$$

从表 1 容易求得基于 $\underline{R}(X)$ 提取的决策规则如下:

(1) 针对概念 X_1:$(a=1\wedge b=1\wedge c=0)\rightarrow d=1$,支持规则的对象集合$\{x_1,x_2\}$;

(2) 针对概念 X_2:$(a=1\wedge b=0\wedge c=1)\rightarrow d=2$,支持规则的对象集合$\{x_9,x_{10}\}$;

(3) 针对概念 X_3:$(a=3\wedge b=2\wedge c=2)\rightarrow d=3$,支持规则的对象集合$\{x_{14},x_{16}\}$.

基于 $\overline{R}(X)$ 提取的决策规则如下:

(1) 针对概念 X_1:$(a=1\wedge b=1\wedge c=0)\vee(a=1\wedge b=1\wedge c=2)\vee(a=2\wedge b=0\wedge c=1)\rightarrow d=1$;支持规则的对象集合$\{x_1,x_2,x_3,x_4,x_5,x_6\}$,错误的支持规则的对象有$\{x_7,x_8,x_{13}\}$;

(2) 针对概念 X_2:$(a=1\wedge b=1\wedge c=2)\vee(a=2\wedge b=0\wedge c=1)\vee(a=1\wedge b=0\wedge c=1)\vee(a=3\wedge b=2\wedge c=1)\rightarrow d=2$,支持规则的对象集合$\{x_7,x_8,x_9,x_{10},x_{11},x_{12}\}$,错误的支持规则的对象有$\{x_3,x_4,x_5,x_6,x_{13},x_{15}\}$;

(3) 针对概念 X_3:$(a=2\wedge b=0\wedge c=1)\vee(a=1\wedge b=0\wedge c=1)\vee(a=3\wedge b=2\wedge c=1)\vee(a=3\wedge b=2\wedge c=2)\rightarrow d=3$,支持的对象集合$\{x_{13},x_{14},x_{15},x_{16}\}$,错误的支持规则的对象有$\{x_6,x_8,x_{11},x_{12}\}$.

基于 $R_{0.5}(X)$提取的决策规则如下:

(1) 针对概念 X_1:$(a=1\wedge b=1\wedge(c=0\vee c=2))\rightarrow d=1$,支持规则的对象集合$\{x_1,x_2,x_3,x_4,x_5\}$,错误的支持规则的对象有$\{x_7\}$;

(2) 针对概念 X_2:$(a=1\wedge b=0\wedge c=1)\vee(a=3\wedge b=2\wedge c=1)\rightarrow d=2$,支持规则的对象集合$\{x_9,x_{10},x_{11},x_{12}\}$,错误的支持规则的对象有$\{x_{15}\}$;

(3) 针对概念 X_3:$(a=3\wedge b=2\wedge c=2)\rightarrow d=3$,支持的对象集合$\{x_{14},x_{16}\}$.

为此,根据这些规则情况得出对比分析表(见表 2).根据表 2 可以看出,基于 $\underline{R}(X)$ 提取决策规则,很多对象无法确定决策结果,这是实际决策问题中非常忌讳的问题;基于 $\overline{R}(X)$ 提取决策规则,每个对象都有决策分类,但是处于边界域上的对象无法具体分类哪个决策类,因此产生很多错误的决策分类结果;基于 $R_{0.5}(X)$提取决策规则,正确分类的对象增多,无法分类的对象减少.相对于前面两种情况,基于 $R_{0.5}(X)$来提取决策规则具有更好的分类效果.

Table 2 Comparative analysis

表 2 对比分析

		支持规则的对象个数	错误的对象个数	无法识别的对象个数
基于 $\underline{R}(X)$ 提取规则	X_1	2	0	4
	X_2	2	0	4
	X_3	2	0	2
基于 $\overline{R}(X)$ 提取规则	X_1	6	3	0
	X_2	6	6	0
	X_3	4	4	0
基于 $R_{0.5}(X)$提取规则	X_1	5	1	1
	X_2	4	1	1
	X_3	2	0	1

定理 1~定理 4 及其推论 1 揭示了 $S(X,R_{0.5}(X))$,$S(X,R_\lambda(X))$, $S(X,\underline{R}(X))$ 和 $S(X,\overline{R}(X))$ 之间的关系.在知识基中,如果知识粒度逐渐减小(增大),则 $S(X,R_{0.5}(X))$, $S(X,\underline{R}(X))$ 和 $S(X,\overline{R}(X))$ 之间的关系有何变化?下面,我们来讨论这一问题.

4 近似集 $R_{0.5}(X)$随知识粒度的变化关系

在不同知识粒度空间,粗糙集的不确定性变化是人们研究的热点问题[11,12],很多研究者试图发现粗糙集的不确定性(如粗糙度、模糊度等)随知识的变化规律[23−28].因此,在不同知识粒度空间下,$S(X,R_{0.5}(X))$的变化规律如何,是我们关心的问题.如果知识粒被细化,即知识颗粒被细分,那么 X 和 $R_{0.5}(X)$的相似度 $S(X,R_{0.5}(X))$会有怎样的变化规律?换言之,在不同知识粒度下,$S(X,R_{0.5}(X))$将会如何变化?

设 U 在关系 R 下形成的等价类分别是$[x_1]_R,[x_2]_R,\ldots,[x_n]_R$,而 U 在关系 R'下形成的等价类分别是$[x_1]_{R'},[x_2]_{R'},\ldots,[x_n]_{R'}$.如果 $R'\subseteq R$,则$[x_i]_{R'}\subseteq[x_i]_R$,称 R'在 U 上形成的划分 U/R'是 R 在 U 上形成的划分 U/R 的细分,记为 $U/R'\preceq U/R$;如果存在 $x_j\in U$,有$[x_j]_{R'}\subset[x_j]_R$,则称划分 U/R'是划分 U/R 的严格细分,记为 $U/R'\prec U/R$.

接下来,分情况讨论在不同知识粒度下,$S(X,R_{0.5}(X))$与 $S(X,R'_{0.5}(X))$ 的关系.设 $U/R'\prec U/R$,则对任意的 $x\in U$,必有$[x]_{R'}\subseteq[x]_R$,且存在 $y\in U$,必有$[y]_{R'}\subset[y]_R$.所以,在 U/R'至少存在两个及其以上的知识粒的并集等于$[y]_R$.为了便

于描述,我们记 $R_{0.5}(X)=\underline{R}(X)\cup[x_{i_1}]_R\cup[x_{i_2}]_R\cup...\cup[x_{i_k}]_R$, $BN_R(X)=[x_{i_1}]_R\cup[x_{i_2}]_R\cup...\cup[x_{i_m}]_R$,这里,$m\geqslant k$.如果是 $NEG_R(X)$或者 $POS_R(X)=\underline{R}(X)$ 中的知识粒被细化,显然 $S(X,R_{0.5}(X))=S(X,R'_{0.5}(X))$.接下来我们主要讨论边界域上的某个知识粒被细化后 $S(X,R'_{0.5}(X))$ 的变化情况.为了简化起见,不妨设边界域上有且仅有一个知识粒(等价类) $[x_{i_t}]_R$ 在 U/R'中被细分成两个知识粒 $[x_{i_t}^1]_{R'}$ 和 $[x_{i_t}^2]_{R'}$,其他知识粒不变.下面我们分情况加以讨论.

定理 5. 如果 $k<t\leqslant m$,即 $[x_{i_t}]_R\not\subset R_{0.5}(X)$:

① 若 $[x_{i_1}^1]_{R'}\subseteq R'_{0.5}(X),[x_{i_1}^2]_{R'}\not\subset R'_{0.5}(X)$, 则 $S(X,R_{0.5}(X))\leqslant S(X,R'_{0.5}(X))$;

② 若 $[x_{i_1}^1]_{R'}\not\subset R'_{0.5}(X),[x_{i_1}^2]_{R'}\not\subset R'_{0.5}(X)$, 则 $S(X,R_{0.5}(X))=S(X,R'_{0.5}(X))$.

证明:根据定理 1 的证明,

$$S(X,R_{0.5}(X))=\frac{|X\cap R_{0.5}(X)|}{|X\cup R_{0.5}(X)|}=\frac{|\underline{R}(X)|+|X\cap[x_{i_1}]_R|+|X\cap[x_{i_2}]_R|+...+|X\cap[x_{i_k}]_R|}{|X|+|([x_{i_1}]_R-X)|+|([x_{i_2}]_R-X)|+...+|([x_{i_k}]_R-X)|}.$$

因为 $[x_{i_t}^1]_{R'}\cup[x_{i_t}^2]_{R'}=[x_{i_t}]_R,[x_{i_t}^1]_{R'}\subseteq R'_{0.5}(X)$, 且 $[x_{i_t}^2]_{R'}\not\subset R'_{0.5}(X)$, 所以,

$$\begin{aligned}S(X,R'_{0.5}(X))&=\frac{|X\cap R'_{0.5}(X)|}{|X\cup R'_{0.5}(X)|}\\&=\frac{|\underline{R}'(X)|+|X\cap[x_{i_1}]_{R'}|+|X\cap[x_{i_2}]_{R'}|+...+|X\cap[x_{i_k}]_{R'}|+|X\cap[x_{i_t}^1]_{R'}|}{|X|+|([x_{i_1}]_{R'}-X)|+|([x_{i_2}]_{R'}-X)|+...+|([x_{i_k}]_{R'}-X)|+|([x_{i_t}^1]_{R'}-X)|}\\&=\frac{|\underline{R}(X)|+|X\cap[x_{i_1}]_R|+|X\cap[x_{i_2}]_R|+...+|X\cap[x_{i_k}]_R|+|X\cap[x_{i_t}^1]_{R'}|}{|X|+|([x_{i_1}]_R-X)|+|([x_{i_2}]_R-X)|+...+|([x_{i_k}]_R-X)|+|([x_{i_t}^1]_{R'}-X)|}.\end{aligned}$$

因为 $[x_{i_1}^1]_{R'}\subseteq R'_{0.5}(X)$, 所以 $\dfrac{|[x_{i_t}^1]_{R'}\cap X|}{|[x_{i_t}^1]_{R'}\cap X|+|[x_{i_t}^1]_{R'}-X|}\geqslant 0.5$, 即 $|[x_{i_t}^1]_{R'}\cap X|\geqslant|[x_{i_t}^1]_{R'}-X|$.

根据引理 1, $S(X,R_{0.5}(X))\leqslant S(X,R'_{0.5}(X))$.

另外,若 $[x_{i_1}^1]_{R'}\not\subset R'_{0.5}(X),[x_{i_1}^2]_{R'}\not\subset R'_{0.5}(X)$, 则 $S(X,R_{0.5}(X))=S(X,R'_{0.5}(X))$.

当然,由于 $[x_{i_t}]_R\not\subset R_{0.5}(X)$,所以, $[x_{i_1}^1]_{R'}\subseteq R'_{0.5}(X)$ 且 $[x_{i_1}^2]_{R'}\subseteq R'_{0.5}(X)$ 的情况不可能发生. □

定理 6. 如果 $1\leqslant t\leqslant k$,即 $[x_{i_t}]_R\subseteq R_{0.5}(X)$:

① 若 $[x_{i_t}^1]_{R'}\subseteq R'_{0.5}(X),[x_{i_t}^2]_{R'}\subseteq R'_{0.5}(X)$, 则 $S(X,R_{0.5}(X))=S(X,R'_{0.5}(X))$;

② 若 $[x_{i_t}^1]_{R'}\subseteq R'_{0.5}(X),[x_{i_t}^2]_{R'}\not\subset R'_{0.5}(X)$, 且 $S(X,R_{0.5}(X))\geqslant\dfrac{|[x_{i_t}^2]\cap X|}{|[x_{i_t}^2]-X|}$, 则 $S(X,R_{0.5}(X))\leqslant S(X,R'_{0.5}(X))$.

证明:情形①的证明显然,这里主要证明情形②.因为 $[x_{i_1}^2]_{R'}\not\subset R'_{0.5}(X)$:

(i) 若 $[x_{i_1}^2]_{R'}\cap X=\varnothing$,则 $|X\cap[x_{i_t}^1]_R|=|X\cap[x_{i_t}]_R|$,且 $|([x_{i_t}^1]_{R'}-X)|<|([x_{i_t}]_R-X)|$.因此,

$$\begin{aligned}S(X,R'_{0.5}(X))&=\frac{|X\cap R'_{0.5}(X)|}{|X\cup R'_{0.5}(X)|}\\&=\frac{|\underline{R}'(X)|+|X\cap[x_{i_1}]_{R'}|+|X\cap[x_{i_2}]_{R'}|+...+|X\cap[x_{i_t}^1]_{R'}|+...+|X\cap[x_{i_k}]_{R'}|}{|X|+|([x_{i_1}]_{R'}-X)|+|([x_{i_2}]_{R'}-X)|+...+|([x_{i_t}^1]_{R'}-X)|+...+|([x_{i_k}]_{R'}-X)|}\\&=\frac{|\underline{R}(X)|+|X\cap[x_{i_1}]_R|+|X\cap[x_{i_2}]_R|+...+|X\cap[x_{i_t}]_R|+...+|X\cap[x_{i_k}]_R|}{|X|+|([x_{i_1}]_R-X)|+|([x_{i_2}]_R-X)|+...+|([x_{i_t}^1]_{R'}-X)|+...+|([x_{i_k}]_R-X)|}\\&>\frac{|\underline{R}(X)|+|X\cap[x_{i_1}]_R|+|X\cap[x_{i_2}]_R|+...+|X\cap[x_{i_k}]_R|}{|X|+|([x_{i_1}]_R-X)|+|([x_{i_2}]_R-X)|+...+|([x_{i_k}]_R-X)|}\\&=S(X,R_{0.5}(X)).\end{aligned}$$

(ii) 若 $[x_{i_t}^1]_{R'}\subseteq X$,则 $|[x_{i_t}^1]_{R'}\cap X|=|[x_{i_t}^1]_{R'}|$, 因此,

$$
\begin{aligned}
S(X,R'_{0.5}(X)) &= \frac{|X \cap R'_{0.5}(X)|}{|X \cup R'_{0.5}(X)|} \\
&= \frac{|\underline{R}'(X)| + |X \cap [x_{i_1}]_{R'}| + |X \cap [x_{i_2}]_{R'}| + \ldots + |X \cap [x_{i_t}^1]_{R'}| + \ldots + |X \cap [x_{i_k}]_{R'}|}{|X| + |([x_{i_1}]_{R'} - X)| + |([x_{i_2}]_{R'} - X)| + \ldots + |([x_{i_t}^1]_{R'} - X)| + \ldots + |([x_{i_k}]_{R'} - X)|} \\
&= \frac{|\underline{R}(X)| + |X \cap [x_{i_1}]_R| + |X \cap [x_{i_2}]_R| + \ldots + |[x_{i_t}^1]_{R'}| + \ldots + |X \cap [x_{i_k}]_R|}{|X| + |([x_{i_1}]_R - X)| + |([x_{i_2}]_R - X)| + \ldots + |[x_{i_{t-1}}]_R - X| + |[x_{i_{t+1}}]_R - X| + \ldots + |([x_{i_k}]_R - X)|} \\
&= \frac{|X \cap R_{0.5}(X)| - |[x_{i_t}^2]_{R'} \cap X|}{|X \cup R_{0.5}(X)| - |[x_{i_t}^2]_{R'} - X|}.
\end{aligned}
$$

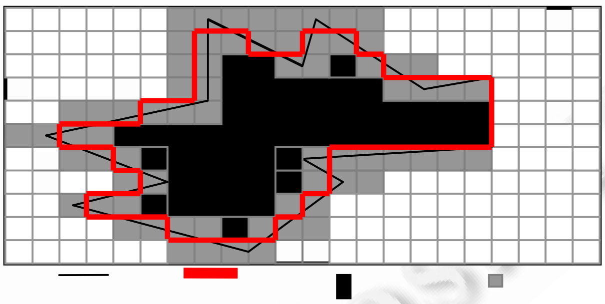
因为 $S(X,R_{0.5}(X)) \geqslant \frac{|[x_{i_t}^2] \cap X|}{|[x_{i_t}^2] - X|}$,根据引理 2,有 $S(X,R_{0.5}(X)) \leqslant S(X,R'_{0.5}(X))$.

(iii) 若 $[x_{i_t}^1]_{R'} \subseteq BN_{R'}(X)$ 且 $[x_{i_t}^2]_{R'} \subseteq BN_{R'}(X)$, 因为 $[x_{i_t}^1]_{R'} \subseteq R'_{0.5}(X), [x_{i_t}^2]_{R'} \not\subset R'_{0.5}(X)$, 则

$$
\begin{aligned}
S(X,R'_{0.5}(X)) &= \frac{|X \cap R'_{0.5}(X)|}{|X \cup R'_{0.5}(X)|} \\
&= \frac{|\underline{R}'(X)| + |X \cap [x_{i_1}]_{R'}| + |X \cap [x_{i_2}]_{R'}| + \ldots + |X \cap [x_{i_t}^1]_{R'}| + \ldots + |X \cap [x_{i_k}]_{R'}|}{|X| + |([x_{i_1}]_{R'} - X)| + |([x_{i_2}]_{R'} - X)| + \ldots + |([x_{i_t}^1]_{R'} - X)| + \ldots + |([x_{i_k}]_{R'} - X)|} \\
&= \frac{|\underline{R}(X)| + |X \cap [x_{i_1}]_R| + |X \cap [x_{i_2}]_R| + \ldots + |[x_{i_t}^1]_{R'}| + \ldots + |X \cap [x_{i_k}]_R|}{|X| + |([x_{i_1}]_R - X)| + |([x_{i_2}]_R - X)| + \ldots + |[x_{i_t}^1]_{R'} - X| + \ldots + |([x_{i_k}]_R - X)|} \\
&= \frac{|X \cap R_{0.5}(X)| - |[x_{i_t}^2]_{R'} \cap X|}{|X \cup R_{0.5}(X)| - |[x_{i_t}^2]_{R'} - X|}.
\end{aligned}
$$

因为 $S(X,R_{0.5}(X)) \geqslant \frac{|[x_{i_t}^2] \cap X|}{|[x_{i_t}^2] - X|}$,根据引理 2,有 $S(X,R_{0.5}(X)) \leqslant S(X,R'_{0.5}(X))$. □

定理 5 和定理 6 揭示了在不同情况下,当边界域上有知识粒(等价类)在被等价关系 R'细分时,$R'_{0.5}(X)$ 与目标概念(集合)X 的近似程度一般不小于 $R_{0.5}(X)$目标概念(集合)X 的近似程度,如图 4 所示.



集合 X 的边界 $R_{0.5}(X)$的边界 X 的下近似 X 的上、下近似之差

Fig.4 Changing chart of $R_{0.5}(X)$ with different knowledge granularity

图 4 $R_{0.5}(X)$随知识粒度细化的变化图

5 结束语

粗糙集理论的研究已经经历了 20 多年的时间,无论是在系统理论、计算模型的建立和应用系统的研制开发上,都已经取得了很多成果,也建立了一套较为完善的粗糙集理论体系.它在机器学习、知识获取、决策分析、数据库的知识发现、专家系统、决策支持系统、归纳推理、矛盾归结、模式识别、模糊控制和医疗诊断等应用领域取得了不少的成果,业已成为粒计算研究的主要工具之一[29,30].粗糙集理论用两个精确的集合作为边界不确定集合的两条边界线,用精确的方法处理不精确的问题.但粗糙集理论本身没有提出如何用已知的知识基

来构建目标概念(集合)X的近似集.虽然很多文献研究了利用粗糙集的扩展模型来处理有噪声数据的信息系统,但是关于直接给出目标概念(集合)X的近似集的方法的研究甚少.本文从另外一个角度,利用当前知识空间中的知识粒构建目标概念(集合)X 的近似集 $R_{0.5}(X)$,并研究它的相关性质.我们首先将粗糙集转化为模糊集,然后利用截集对边界不确定的元素进行分类,并提出集合相似度的概念,构建目标概念(集合)X 的近似集.研究表明,用 $R_{0.5}(X)$作为 X 的近似集,其近似度优于直接用 $\overline{R}(X)$ 或 $\underline{R}(X)$ 作为 X 的近似集,并讨论了在不同知识粒度下 $S(X,R_{0.5}(X))$的变化规律.这些研究工作从新的视角提出描述不确定性概念的方法,希望这些工作能够推动不确定性人工智能的发展,扩展粗糙集理论模型及其应用.在接下来的研究中,我们将用这样的近似集来获取近似规则,实现知识获取.

致谢 在此,我们向对本文工作给予支持和建议的同行,尤其是对本文给予评审并提出宝贵意见的专家们表示感谢.

References:

[1] Li DY, Liu CY, Du Y, Han X. Artificial intelligence with uncertainty. Journal of Software, 2004,15(11):1583−1594 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/15/1583.htm

[2] Liu J, Chen XP, Cai QS, Fan Y. Recognition structure of uncertainty: A unified framework for representation, reasoning and learning. Journal of Software, 2002,13(4):649–651 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/13/649.htm

[3] Zadeh LA. Fuzzy sets. Information and Control, 1965,8(1):338−353. [doi: 10.1016/S0019-9958(65)90241-X]

[4] Pawlak Z. Rough sets. International Journal of Computer and Information Science, 1982,11(5):341−356. [doi: 10.1007/BF01001956]

[5] Zhang L, Zhang B. The Theory and Applications of Problem Solving-Quotient Space Based Granular Computing. 2nd ed., Beijing: Tsinghua University Press, 2007 (in Chinese).

[6] Wang GY, Peters JF, Skowron A, Yao Y. Rough sets and knowledge technology. In: Proc. of the RSKT 2006. LNCS 4062, Berlin: Springer-Verlag, 2006. 1–32.

[7] Wang GY, Li TR, Grzymala BJ, Miao DQ, Skowron A, Yao YY. Rough sets and knowledge technology. In: Proc. of the RSKT 2008. LNCS 5009, Berlin: Springer-Verlag, 2008. 1–18.

[8] Wang GY, Yao YY, Yu H. A survey on rough set theory and its application. Chinese Journal of Computers, 2009,32(7):1229−1246 (in Chinese with English abstract). [doi: 10.3724/SP.J.1016.2009.01229]

[9] Liu SH, Sheng QJ, Shi ZZ. A new method for fast computing positive region. Journal of Computer Research and Development, 2003,40(5):637−642 (in Chinese with English abstract).

[10] Liu Y, Xiong R, Chu J. Quick attribute reduction algorithm with Hash. Chinese Journal of Computers, 2009,32(8):1493−1499 (in Chinese with English abstract).

[11] Hu F, Wang GY. Quick algorithm for certain rule acquisition based on divide and conquer method. Pattern Recognition and Artificial Intelligence, 2010,23(3):349−356 (in Chinese with English abstract).

[12] Guan LH, Wang GY, Yu H. Incremental algorithm of Pawlak reduction based on attribute order. Jouranl of Southwest Jiaotong University, 2011,46(3):461−468 (in Chinese with English abstract).

[13] Pawlak Z, Skowron A. Rough sets: Some extensions. Information Sciences, 2007,177(1):28−40. [doi: 10.1016/j.ins.2006.06.006]

[14] Inuiguchi M, Yoshioka Y, Kusunoki Y. Variable-Precision dominance-based rough set approach and attribute reduction. Int'l Journal of Approximate Reasoning, 2009,50(8):1199–1214. [doi: 10.1016/j.ijar.2009.02.003]

[15] Xie G, Zhang JL, La KK, Yu L. Variable precision rough set for group decision-making: An application. Int'l Journal of Approximate Reasoning, 2007,49(2):331–343. [doi: 10.1016/j.ijar.2007.04.005]

[16] Mi JS, Wu WZ, Zhang WX. Approaches to knowledge reduction based on variable precision rough set model. Information Sciences, 2004,159(3):255−272. [doi: 10.1016/j.ins.2003.07.004]

[17] Zhang XY, Mo ZW, Variable precision rough sets. Pattern Recognition and Artificial Intelligence, 2004,17(2):151−155 (in Chinese with English abstract).

[18] Ziarko W. Probabilistic approach to rough sets. Int'l Journal of Approximate Reasoning, 2008,49(2):272–284. [doi: 10.1016/j.ijar.2007.06.014]

[19] Yao YY. Probabilistic rough set approximations. Int'l Journal of Approximate Reasoning, 2008,49(2):255–271. [doi: 10.1016/j.ijar.2007.05.019]

[20] Yao YY, Zhou B. Naive Bayesian rough sets. In: Yu J, Greco S, Lingras P, Wang G Y, Skowron A, eds. Proc. of the RSKT 2010: Rough Set and Knowledge Technology, the 5th Int'l Conf. LNCS 6401, Berlin: Springer-Verlag, 2010. 719−726.

[21] Yao YY. Two semantic issues in a probabilistic rough set model. Fundamenta Informaticae, 2011,108(3):249–265.
[22] Wang GY. Rough Set Theory and Knowledge Discovery. Xi'an: Xi'an Jiaotong University Press, 2001 (in Chinese).
[23] Zhang WX, Wu WZ. An introduction and a survey for the studies of rough set theory. Fuzzy Systems and Mathematics, 2000,14(4): 1–12 (in Chinese with English abstract).
[24] Miao DQ, Wang GY, Liu Q, Lin ZY, Yao YY. Granular Computing: Past, Present and Future Prospects. Beijing: Science Press, 2007 (in Chinese).
[25] Wang GY, Zhang QH. Uncertainty of rough sets in different knowledge granularities. Chinese Journal of Computers, 2008,31(9): 1588–1598 (in Chinese with English abstract).
[26] Liang JY, Qian YH. Granulation monotonicity of entropy measure in information systems. Journal of Shanxi University (Natural Science Edition), 2007, 30(2):156–162 (in Chinese with English abstract).
[27] Liang JY, Shi ZZ. The information entropy, rough entropy and knowledge granulation in rough set theory. Int'l Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2004,12(1):37–46. [doi: 10.1142/S0218488504002631]
[28] Miao DQ, Fan SD. The calculation of knowledge granulation and its application. Systems Engineering-Theory & Practice, 2002, 22(1):48–56 (in Chinese with English abstract).
[29] Wang GY, Miao DQ, Wu WZ, Liang JY. Uncertain knowledge representation and processing based on rough set. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2010,22(5):541–544 (in Chinese with English abstract).
[30] Miao DQ, Li DY, Yao YY, Wang GY, *et al*. Uncertainty and Granular Computing. Beijing: Science Press, 2011 (in Chinese).

附中文参考文献:

[1] 李德毅,刘常昱,杜鹢,韩旭.不确定性人工智能.软件学报,2004,15(11):1583–1594. http://www.jos.org.cn/1000-9825/15/1583.htm
[2] 刘洁,陈小平,蔡庆生,范焱.不确定信息的认知结构表示、推理和学习.软件学报,2002,13(4):649–651. http://www.jos.org.cn/1000-9825/13/649.htm
[5] 张铃,张钹.问题求解理论及应用——商空间粒度计算理论及应用.第 2 版,北京:清华大学出版社,1990.
[8] 王国胤,姚一豫,于洪.粗糙集理论与应用研究综述.计算机学报,2009,32(7):1229–1246. [doi: 10.3724/SP.J.1016.2009.01229]
[9] 刘少辉,盛秋戬,史忠植.一种新的快速计算正区域的方法.计算机研究与发展,2003,40(5):637–642.
[10] 刘勇,熊蓉,褚健.Hash 快速属性约简算法.计算机学报,2009,32(8):1493–1499.
[11] 胡峰,王国胤.基于分治法的快速确定规则获取算法.模式识别与人工智能,2010,23(3):349–356.
[12] 官礼和,王国胤,于洪.属性序下的增量式 Pawlak 约简算法.西南交通大学学报,2011,46(3):461–468.
[17] 张贤勇,莫智文.变精度粗糙集.模式识别与人工智能,2004,17(2):151–155.
[22] 王国胤.Rough 集理论与知识获取.西安:西安交通大学出版社,2001.
[23] 张文修,吴伟志.粗糙集理论介绍和研究综述.模糊系统与数学,2000,14(4):1–12.
[24] 苗夺谦,王国胤,刘清,林早阳,姚一豫.粒计算:过去、现在与展望.北京:科学出版社,2007.
[25] 王国胤,张清华.不同知识粒度下粗糙集的不确定性研究.计算机学报,2008,31(9):1588–1598.
[26] 梁吉业,钱宇华.信息系统中熵度量的粒化单调性.山西大学学报(自然科学版),2007,30(2):156–162.
[28] 苗夺谦,范世栋.知识的粒度计算及其应用.系统工程理论与实践,2002,22(1):48–56.
[29] 王国胤,苗夺谦,周志华,吴伟志,梁吉业.不确定信息的粗糙集表示和处理.重庆邮电大学学报(自然科学版),2010,22(5):541–544.
[30] 苗夺谦,李德毅,姚一豫,王国胤,等.不确定性与粒计算.北京:科学出版社,2011.

张清华(1974-),男,重庆人,博士,副教授,主要研究领域为不确定性信息处理,粗糙集与粒计算.

王国胤(1970-),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为 Rough 集理论,粒计算,数据挖掘,知识技术.

肖雨(1989-),男,硕士生,主要研究领域为知识发现,粗糙集与粒计算.