区间值信息系统中基于邻域系统的粗糙集模型*

王天擎[1], 赵良辉[1], 谢 军[2]
(1. 五邑大学 经济管理学院, 广东 江门 529020; 2. 江苏大学 计算机科学与通信工程学院, 江苏 镇江 212013)

摘 要: Pawlak 经典粗糙集理论主要针对离散值,对复杂现实世界中的区间值却不能有效支持。在区间值信息系统中,基于灰格运算和 Hausdorff 距离提出了一种区间值的邻域关系。在该邻域关系基础之上,依次提出了基于邻域关系、最大相容类和邻域系统三种灰色粗集模型,提高了近似空间的精确度;同时讨论了三种灰色粗集模型之间的上、下近似空间,并用实例进行分析及验证。

关键词: 粗糙集; 区间值; 邻域系统; 覆盖; 最大相容类; Hausdorff 距离
中图分类号: TP18;TP301.6 **文献标志码:** A **文章编号:** 1001-3695(2012)11-4242-04
doi:10.3969/j.issn.1001-3695.2012.11.061

Neighborhood systems based rough sets in interval-valued information system

WANG Tian-qing[1], ZHAO Liang-hui[1], XIE Jun[2]
(1. *School of Economics & Managemen, Wuyi University, Jiangmen Guangdong* 529020, *China*; 2. *School of Computer Science & Telecommunication Engineering, Jiangsu University, Zhenjiang Jiangsu* 212013, *China*)

Abstract: Pawlak's rough set theory offers a formal theoretical framework to deal with categorical data, but this model is not applicable to interval data which widely exist in real world applications. By using grey lattice operation and Hausdorff distance, this paper provided a new neighborhood relationship in an interval-valued information system. Furthermore, it proposed three forms of rough set models based on neighborhood relationship, maximal consistent blocks and neighborhood system, which was to improve the accuracy of approximations. Moreover, it employed three numerical examples to substantiate the conceptual arguments.

Key words: rough set; interval data; neighborhood system; covering; maximal consistent block; Hausdorff distance

波兰科学家 Pawlak[1,2] 开创了 Rough 集理论的先河,为处理不精确、不确定、不完备信息提供了新的数学工具。在经典粗糙集理论中,论域上的等价关系起着至关重要的作用;但在现实中,论域上的二元关系经常不是等价的,此时经典粗糙集模型的应用将受到限制。将等价关系放宽为相容关系、相似关系、限制相容关系或自反二元关系后,一般构成的是论域的一个覆盖而不再是等价关系确定的划分,即将 Pawlak 粗糙集理论推广为覆盖广义粗糙集理论[3~5]。在不完备信息系统中,Leung 等人在文献[6]中提出了基于粗糙近似空间的最大相容类,以保证其模块中的任意两个对象均满足相容关系,Guan 等人[7]并进一步将最大相容类分别引入集值和连续值系统中。邻域和邻域系统是 Lin 等人以拓扑空间为基础于 1988 年提出的,它可以通过空间中点的邻域来粒化论域,将邻域理解为基本信息粒子,用来描述空间中的其他概念。目前,基于邻域和邻域系统下的粒计算模型的研究逐渐成为热点[8~10]。

经典不完备粗糙集主要是对分类数据或离散值进行操作,然而在现实生活中的信息系统是非常复杂的,很多测量值都表示为在一定范围内的连续区间值。在经典粗糙集方法中,类似于粗糙近似空间和决策规则生成,对区间值的处理都没有被定义过。灰色系统理论是由我国数学家邓聚龙最先提出来的[11],其中涵括了灰色分类、灰色控制、灰色决策、灰色预测、灰色建模和灰色关系分析等内容,主要用来处理不确定性命题,区间值是其中的一个核心概念。类似于粗糙近似空间和决策规则生成对区间值的处理正处于研究起步阶段[12~15],因此,对于含区间值信息系统的研究有着重要的理论意义和实用价值。

本文以含区间值的灰色信息系统为研究对象。a)通过灰色理论中的灰格运算[13]和 Hausdorff 距离[12]提出了一种基于区间值的邻域关系;b)分析基于邻域关系近似空间分类的不足,提出基于最大相容类方法,用于扩大下近似空间的范围;c)在最大相容类方法的基础上,提出了基于邻域系统的近似空间分类方法,用于缩小上近似空间的范围,提高近似空间的精确性;d)在一给定的区间值灰色决策信息系统中进行了实例分析,以证明基于邻域关系、最大相容类和邻域关系三种粗集模型之间的近似空间关系。

1 基本概念

1.1 邻域近似空间

一个信息系统为一个二元组 IS = $\langle U, AT \rangle$, 其中 U 是一个

收稿日期: 2012-03-26; **修回日期:** 2012-05-04 **基金项目:** 江苏省高校自然基金资助项目(10KJB520004);广东省自然科学基金资助项目(8452902001001552)

作者简介: 王天擎(1974-),男,湖南长沙人,副教授,硕士,主要研究方向为粗糙集、智能信息处理及数据挖掘;赵良辉(1973-),男,湖南邵阳人,副教授,博士,主要研究方向为商务智能及生产调度;谢军(1973-),男,湖南衡阳人,副教授,硕士,主要研究方向为粗糙集理论及其应用(xie_jun888@163.com)。

被称为论域的非空有限对象集合,$U=\{x_1,x_2,\cdots,x_n\}$;AT 是非空有限条件属性集合。

定义 1　在信息系统 IS 中,对于 $\forall A\subseteq AT$,记

$$R(A)=\{(x,y)\in U^2:\forall a\in A,f(x,a)=f(y,a)\} \tag{1}$$

则 $R(A)$ 是 U 上的等价关系,分别称为由 A 决定的不可区分关系,显然 R 是自反、对称且传递的。它们产生的 U 上的划分分别记为

$$U/R_A=\{[x]_A:x\in U\} \tag{2}$$

其中:$[x]_A=\{y:(x,y)\in R(A)\}$ 是 x 关于 A 的等价类。

定义 2　设 IS 为一信息系统,其中 $A\subseteq AT$,对于 $\forall X\subseteq U$,X 基于 $R(A)$ 的下、上近似集分别记为 $\underline{R_A}(X)$、$\overline{R_A}(X)$,且

$$\underline{R_A}(X)=\{x\in U:R_A(x)\subseteq X\}=\{x\in U:[x]_A\subseteq X\}=\cup\{[x]_A:[x]_A\subseteq X\} \tag{3}$$

$$\overline{R_A}(X)=\{x\in U:R_A(x)\cap X\neq\varnothing\}=\{x\in U:[x]_A\cap X\neq\varnothing\}=\cup\{[x]_A:[x]_A\cap X\neq\varnothing\} \tag{4}$$

定义 3　设 U 是一个非空论域集合,$\delta=\{C_i|i\in I\}$ 为 U 上一个子集族。若 $\cup_{i\in I}C_i=U$,则称集合族 C 是 U 的一个覆盖,称序对 $\langle U,\delta\rangle$ 为一覆盖近似空间。

显然,论域 U 上的由等价关系所形成的划分也是一个覆盖,覆盖是划分的一种推广,Pawlak 近似空间是一种特殊的覆盖近似空间。

若 $\boldsymbol{C}$ 是论域 U 上的一个覆盖,$\forall x\in U$,则对象 x 基于邻域系统的覆盖记为

$$\mathrm{CNS}(x)=\{C:C\in\boldsymbol{C},x\in C\}$$

因此,根据覆盖 $\boldsymbol{C}$,可以导出论域 U 上的邻域系统为

$$\mathrm{CNS}(U)=\{\mathrm{CNS}(x):x\in U\}$$

定义 4　设 U 是一个非空论域集合,对 $\forall A\subseteq AT$,$\forall x\in U$,x 关于 A 的邻域记为 $\zeta_A(x)$,且

$$\zeta_A(x)=\{x_i\in U:\Delta(x,x_i)\leqslant\beta\} \tag{5}$$

其中:Δ 是一个距离函数,满足以下性质:

a)$\Delta(x,y)\geqslant 0$;

b)$\Delta(x,y)=0$,当且仅当 $x=y$;

c)$\Delta(x,y)=\Delta(y,x)$;

d)$\Delta(x,y)+\Delta(x,z)\geqslant\Delta(x,z)$。

定义 5　设 $U=\{x_1,x_2,\cdots,x_n\}$ 是一个非空论域集合,则 $N=\{\zeta(x_1),\zeta(x_2),\cdots,\zeta(x_n)\}$ 构成 U 的一个覆盖,称序对 $\langle U,N\rangle$ 为一邻域近似空间。对于 $\forall X\subseteq U$,则 X 基于 $\langle U,N\rangle$ 的下、上近似集可定义为

$$\underline{N}(X)=\{x\in U:\zeta(x)\subseteq X\} \tag{6}$$

$$\overline{N}(X)=\{x\in U:\zeta(x)\cap X\neq\varnothing\} \tag{7}$$

1.2　灰格运算与 Hausdorff 距离

定义 6　设 U 为一非空论域集合,对象 x 称为论域 U 中的一个元素,$X\subseteq\mathrm{R}$ 表示对象 x 可能取值的范围集合。设对象 x 中的两个值为$\underline{x}$、$\overline{x}$(其中 $\underline{x}=\inf X$,$\overline{x}=\sup X$),则使用形式 $\otimes x=x|_{\underline{\mu}}^{\overline{\mu}}(\overline{\mu}\geqslant\underline{\mu})$,可定义 x 为

a)当且仅当$\underline{x}\rightarrow-\infty$ 且$\overline{x}\rightarrow+\infty$ 时,称$\otimes x$ 为黑数;

b)当且仅当$\underline{x}=\overline{x}$时,称$\otimes x$ 为白数或白值,记为$\tilde{\otimes}x$;

c)当$\otimes x=[\underline{x},\overline{x}]$时,称$\otimes x$ 为灰数。

本文中的区间值是指其值为灰数,而与之相对的非区间值则由白数构成。

定义 7　设任意两个灰数$\otimes x$ 与$\otimes y$ 相等,称为灰格一致关系,其定义为

$$\otimes x=\otimes y\quad\text{当且仅当}\ \underline{x}=\underline{y},\overline{x}=\overline{y} \tag{8}$$

定义 8　令"$\rightarrow$"表示为两个灰数$\otimes x$ 与$\otimes y$ 包含,称为灰格包含关系,其定义为

$$\otimes x\rightarrow\otimes y\quad\text{当且仅当}\ \underline{x}\geqslant\underline{y},\overline{x}\leqslant\overline{y} \tag{9}$$

其中:$\otimes x=|\underline{x},\overline{x}|$ 且$\otimes x=|\underline{y},\overline{y}|$。对任何实数 $k(k\in\mathrm{R})$ 或者$\tilde{\otimes}x$ 包含在$\otimes y$,有

$$k\rightarrow\otimes y\ \text{当}\ k\geqslant\underline{y},\ k\leqslant\overline{y}$$

$$\tilde{\otimes}x\rightarrow\otimes y\ \text{当}\ \tilde{\otimes}x\geqslant\underline{y},\tilde{\otimes}x\leqslant\overline{y}$$

显然,灰格的包含关系具有自反性、反对称性和传递性。

定义 9　对$\otimes x$ 和$\otimes y$ 及两个灰色,它们之间的灰格运算关系如下:

a) $\otimes \mathrm{x}\vee\otimes \mathrm{y}=[\min(\underline{x},\underline{y}),\max(\overline{x},\overline{y})]$

$$\tilde{\otimes}x\vee\tilde{\otimes}y=[\min(\tilde{\otimes}x,\tilde{\otimes}y),\max(\tilde{\otimes}x,\tilde{\otimes}y)]$$

b) $$\otimes x\wedge\otimes y\begin{cases}[\underline{x},\overline{x}] & \text{if }\otimes x\rightarrow\otimes y\\ [\underline{y},\overline{y}] & \text{if }\otimes y\rightarrow\otimes x\\ [\underline{x},\overline{y}] & \text{if }\underline{x}\rightarrow\otimes y\text{ and }\overline{y}\rightarrow\otimes x\\ [\underline{y},\overline{x}] & \text{if }\underline{y}\rightarrow\otimes x\text{ and }\overline{x}\rightarrow\otimes y\\ \varnothing & \text{otherwise}\end{cases}$$

$$\tilde{\otimes}x\wedge\tilde{\otimes}y=\begin{cases}\tilde{\otimes}x & \text{if }\tilde{\otimes}x=\tilde{\otimes}y\\ \varnothing & \text{if }\tilde{\otimes}x\neq\tilde{\otimes}y\end{cases}$$

c) $\otimes x^c=\{x\in X^c|x<\underline{x},x<\overline{x}\}$

d) $\otimes x\oplus\otimes y=$

$$\begin{cases}\otimes x^c\wedge\otimes y^c & \text{if }\otimes x\wedge\otimes y=\varnothing\\ (\otimes x\vee\otimes y)\wedge(\otimes x\vee\otimes y)^c & \text{if }\otimes x\wedge\otimes y\neq\varnothing\end{cases}$$

定义 10　Hausdorff 距离经常用来进行图像处理及用来比较 A 和 B 两个集合,这个距离取决于分别在集合 A 和 B 中的两个对象 u 和 v 的比较。那么 Euclidean 距离和 Hausdorff 距离可定义为

$$d_H(A,B)=\max(h(A,B),h(B,A)) \tag{10}$$

其中:$h(A,B)=\sup_{u\in A}\inf_{v\in B}\|u-v\|$,有时 h 也被称为定向的 Hausdorff 距离。

定义 11　本文中若设 A 和 B 为两个区间值$\otimes x=[\underline{x},\overline{x}]$,$\otimes y=[\underline{y},\overline{y}]$,那么定义 10 中的 Hausdorff 距离 d_H 对于区间值数据的操作就可以简化为 $d_H^{\otimes}$,记为

$$d_H^{\otimes}(\otimes x,\otimes y)=\max\{|\underline{x}-\underline{y}|,|\overline{x}-\overline{y}|\} \tag{11}$$

2　基于区间值的邻域粗集模型

定义 12　设区间值信息系统为一个三元组 INIS $=\langle U,AT,V\rangle$,U 是一个非空有限论域集合,AT 是非空有限条件属性集合,V 是所有条件属性的值域集合,$V=V_{AT}=\cup_{a\in AT}V_a$。其中 V_{AT}是所有条件属性的值域集合;对于 $\forall a\in AT$,$\forall x\in U$,则 x 在 a 的取值 $a(x)$ 为一个区间值,即 $a(x)=[\underline{a(x)},\overline{a(x)}]$,其中,$\overline{a(x)}=\inf(a(x))$ 且$\overline{a(x)}=\sup(a(x))$。

根据定义 6 可知,当$\underline{a(x)}=\overline{a(x)}$时,区间值 $a(x)$ 可以退化为一个实数(白化值),因此,传统的信息系统可以看做为区间值信息系统的一种特殊形式。

定义 13　设 INIS 为一个区间值信息系统,对 $\forall A\subseteq AT$,记

$$S_A=\{(x,y)\in U^2:\forall a\in A,d_H^{\otimes}(a(x),a(y))\leqslant\beta\}$$

则 S_A 称为由 A 决定 U 上的邻域关系。显然 S_A 是自反和对称

的,但非传递的,对基于不可区分关系的 Pawlak 粗糙集进行的扩充。

定义 14 设 INIS 为一个区间值信息系统,$\forall x \in U$,记 $S_A(x)$ 是 x 关于 S_A 的邻域类,有

$$SN_A(x)=\{y\in U,(x,y)\in S_A\}$$

则 $SN_A(x)$ 是 x 由属性 A 所构成不可分辨对象的最大集合。显然,它们产生的 U 上的一个覆盖记为 $U/S_{AT}=\{SN_{AT}(x):x\in U\}$。

定义 15 设 INIS 为一个区间值信息系统,对于 $A\subseteq AT$,$\forall X\subseteq U$,X 基于 $R(A)$ 的下、上近似集分别记为 $\underline{SN}_A(x)$,$\overline{SN}_A(x)$ 且 $\underline{SN}_A(X)=\{x\in U:S_A(x)\subseteq X\}$,$\overline{SN}_A(X)=\{x\in U:S_A(x)\cap X\neq\varnothing\}$。

性质 1 根据定义 15,很容易证明以下性质:

a) $\underline{SN}_A(X)=\{x\in X:S_A(x)\subseteq X\}\neq\cup\{S_A(x):S_A(x)\subseteq X\}$;

b) $\overline{SN}_A(X)=\cup\{S_A(x):x\in X\}\neq\cup\{S_A(x):S_A(x)\cap X\neq\varnothing\}$。

例 1 表 1 为一个区间值决策信息系统,其中 $U=\{x_1,x_2,x_3,x_4,x_5,x_6\}$,$AT=\{a_1,a_2,a_3,a_4\}$ 为条件属性集合,d 为序决策属性;决策系统作出的决策通常应该是清晰明确的,在表 1 中决策属性的值为 $V_d=\{1,2,3\}$。令 $U/\mathrm{IND}(d)=\{X_1,X_2,X_3\}=\{\{x_1,x_2,x_3,x_5\},\{x_4\},\{x_6\}\}$。

表1 区间值决策信息系统表

U	a_1	a_2	a_3	a_4	d
x_1	[2.2, 3.0]	[1.1, 1.4]	[2.9, 3.5]	[1.2, 1.6]	2
x_2	[1.6, 1.9]	[2.1, 2.5]	[2.2, 4.1]	[1.8, 2.0]	2
x_3	[2.0, 3.1]	[2.1, 2.9]	[3.2, 3.9]	[2.1, 2.6]	2
x_4	[1.8, 2.4]	[2.1, 2.7]	[3.3, 4.2]	[1.9, 2.4]	3
x_5	[1.3, 2.0]	[2.2, 2.8]	[2.7, 4.3]	[1.1, 1.5]	2
x_6	[1.5, 1.8]	[1.2, 1.9]	[1.5, 3.0]	[1.3, 1.6]	1

本例中,设 $\beta=1.0$,则根据定义 13 和 14 可得

$$SN_{AT}(x_1)=\{x_1\},\ SN_{AT}(x_2)=\{x_2,x_5\},\ SN_{AT}(x_3)=\{x_3,x_4\}$$
$$SN_{AT}(x_4)=\{x_3,x_4,x_5\},\ SN_{AT}(x_5)=\{x_2,x_4,x_5\},\ SN_{AT}(x_6)=\{x_6\}$$

再根据定义 15 可得

$$\underline{SN}_A(X_1)=\{x_1,x_2\},\underline{SN}_A(X_2)=\varnothing,\underline{SN}_A(X_3)=\{x_6\}$$
$$\overline{SN}_A(X_1)=\{x_1,x_2,x_3,x_4,x_5\},\overline{SN}_A(X_2)=\{x_3,x_4,x_5\},\overline{SN}_A(X_3)=\{x_6\}$$

从上例中可以发现邻域关系还存在以下一些不足:a)不同的邻域类可以存在包含关系,如 $SN_{AT}(x_2)$ 和 $SN_{AT}(x_5)$,有 $SN_{AT}(x_2)\subseteq(SN_{AT}(x_5))$;b)在同一邻域类中,并不是所有两个对象都满足邻域关系,如在 $SN_{AT}(x_5)$ 中,x_3 和 x_5 不满足邻域关系 S_{AT}。为了克服上面两种不足,本文在区间值决策信息系统中提出最大相容类方法。

3 基于区间值的邻域系统粗集模型

3.1 邻域关系的最大相容类粗集模型

区间值决策信息系统中一个最大相容类应该具备以下两个条件:a)在一个最大相容类中的任意两个对象都应该满足邻域关系;b)在一个最大相容类中若再加入其他任何一个对象,则该对象必定与最大相容类中某一个对象不满足邻域关系。

定义 16 设 INIS 为一个区间值信息系统,对 $\forall A\subseteq AT$,记

$$TS(A)=\{K\subseteq U:K^2\subseteq S_A\wedge(\forall x\notin K\rightarrow(K\cup\{x\})^2\not\subseteq S_A\}$$

则 $TS(A)$ 称为由 A 决定所有最大相容类的集合。同时,称 $TS_x(A)$ 表示为包含对象 $x\in U$ 的所有最大相容类集合。

性质 2 设 INIS 为一个区间值信息系统,对 $\forall A\subseteq AT$,则 x 由 A 决定的任意相容类可以表示为包含对象 x 的最大相容类的并集,即有

$$S_A(x)=\cup\{K\in TS(A):K\not\subseteq S_A(x)\}=\cup\{K:K\in TS_x(A)\}$$

证明 显然有 $S_A(x)\supseteq\cup\{K\in TS(A):K\subseteq S_A(x)\}$,则只需要证明 $S_A(x)\subseteq\cup\{K\in TS(A);K\subseteq S_A(x)\}$。对 $\forall y\in S_A(x)$,$\{x,y\}$ 是相容的,如果 $\{x,y\}$ 是一个最大相容类,则性质 2 是正确的;如果不是,则 $\{x,y\}$ 不是最大相容类,因此,肯定存在一个最大相容类 $K\in TS(A)$ 使得 $\{x,y\}\subseteq K$。根据 K 的一致性可得 $K\subseteq S_A(x)$,则有 $y\subseteq\cup\{K\in TS(A):K\subseteq S_A(x)\}$,进一步可得 $S_A(x)\subseteq\cup\{K\in TS(A);K\subseteq S_A(x)\}$。证毕。

定义 17 设 INIS 为一个区间值信息系统,对 $\forall A\subseteq AT$ 和 $\forall X\subseteq U$,X 基于最大相容类的下、上近似集分别记为 $\underline{TS}_A(X)$ 和 $\overline{TS}_A(X)$ 且 $\underline{TS}_A(X)=\cup\{K\in TS(A):K\subseteq X\}$;$\overline{TS}_A(X)=\cup\{K\in TS(A):K\cap X\neq\varnothing\}$。

性质 3 设 INIS 为一个区间值信息系统,对 $\forall A\subseteq AT$ 和 $\forall X\subseteq U$,则有

a) $\underline{TS}_A(X)\subseteq X\subseteq\overline{TS}_A(X)$;

b) 若 $A\subseteq B$,则有 $\overline{TS}_B(X)\subseteq\overline{TS}_A(X)$,但不满足 $\underline{TS}_A(X)\subseteq\underline{TS}_B(X)$。

定理 1 设 INIS 为一个区间值信息系统,对 $\forall A\subseteq AT$ 和 $\forall X\subseteq U$,则有

$$\underline{SN}_A(X)\subseteq\underline{TS}_A(X)\subseteq X\subseteq\overline{TS}_A(X)=\overline{SN}_A(X)$$

证明 a) 证明 $\underline{SN}_A(X)\subseteq\underline{TS}_A(X)$。假设 $x\in\underline{SN}_A(X)$,根据 $\underline{SN}_A(X)$ 的定义有 $S_A(x)\subseteq X$。再根据性质 2,有 $S_A(x)=\cup\{K:K\in TS_x(A)\}\subseteq X$,则对于 $\forall K\subseteq TS_x(A)$,$K\subseteq X$,因此可得 $x\in\underline{TS}_A(X)$。

b) 证明 $\overline{TS}_A(X)=\overline{SN}_A(X)$。根据性质 1 有 $\overline{SN}_A(X)=\{x\in U:S_A(x)\cap X\neq\varnothing\}=\cup\{S_A(x):x\in X\}$。根据性质 2,对 $\forall x\in U$,即可得 $S_A(x)=\cup\{K:K\in TS_x(A)\}$。因此有 $\overline{SN}_A(X)=\cup\{S_A(x):x\in X\}=\cup\{K\in TS(A):K\cap X\neq\varnothing\}=\overline{TS}_A(X)$。证毕。

通过与基于邻域关系的粗糙集模型比较,根据定理 1 可知,基于最大相容类的粗糙集模型扩大了其下近似空间的范围,从而提高了近似空间的精确性。

例 2 通过在例 1 中所计算得到的邻域 $SN_{AT}(U)$,根据定义 16 和 17,则可计算出最大相容类集合和基于最大相容类的粗糙近似集,如下所示:

$TS(AT)=\{K_1=\{x_1\},K_2=\{x_2,x_5\},K_3=\{x_3,x_4\},K_4=\{x_4,x_5\},K_5=\{x_6\}\}$

$\underline{TS}_{AT}(X_1)=\{x_1,x_2,x_5\},\underline{TS}_{AT}(X_2)=\varnothing,\underline{TS}_{AT}(X_3)=\{x_6\}$

$\overline{TS}_{AT}(X_1)=\{x_1,x_2,x_3,x_4,x_5\},\overline{TS}_{AT}(X_2)=\{x_3,x_4,x_5\}$,$\overline{TS}_{AT}(X_3)=\{x_6\}$

以上结果也验证了定理 1 的正确性。

3.2 基于邻域系统的粗集模型

定义 18 设 INIS 为一个区间值信息系统,对 $\forall A\subseteq AT$,因为 $TS(A)$ 是 U 上的一个覆盖,则 $\forall x\in U$,基于 x 的邻域系统可定义为

$$\mathrm{TNS}_A(x)=\{K:K\in TS(A):x\in K\}$$

那么,进一步基于 U 的邻域系统则可定义为

$$\mathrm{TNS}_A(U)=\{\mathrm{TNS}_A(x):x\in U\}$$

定义 19 设 U 为一非空论域集合,$\mathrm{TNS}(U)$ 为 U 上一个给定邻域系统,对 $\forall X\subseteq U$,X 基于邻域系统的下、上近似集分别

记为$\underline{\mathrm{TNS}}_{AT}(X)$和$\overline{\mathrm{TNS}}_{AT}(X)$且

$\underline{\mathrm{TNS}}_{AT}(X)=\{x\in U: \exists TN(x)\in \mathrm{TNS}(x)\ \text{s.t.}\ TN(x)\neq\varnothing \wedge TN(x)\subseteq X\}$;

$\overline{\mathrm{TNS}}_{AT}(X)=\{x\in U: \forall T(x)\in TM(x)\ \text{s.t.}\ T(x)\cap\varnothing\}$

定理 2　根据$\underline{\mathrm{TNS}}(X)$和$\overline{\mathrm{TNS}}(X)$可得以下性质:

a) $\underline{\mathrm{TNS}}(\varnothing)=\overline{\mathrm{TNS}}_{AT}(\varnothing)=\varnothing$;

b) $\underline{\mathrm{TNS}}(X\cap Y)\subseteq\underline{\mathrm{TNS}}(X)\cap\underline{\mathrm{TNS}}(Y)$;

c) $\overline{\mathrm{TNS}}(X\cap Y)\subseteq\underline{\mathrm{TNS}}(X)\cap\underline{\mathrm{TNS}}(Y)$;

d) $\underline{\mathrm{TNS}}(X\cup Y)\supseteq\underline{\mathrm{TNS}}(X)\cup\underline{\mathrm{TNS}}(Y)$;

e) $\overline{\mathrm{TNS}}(X\cup Y)\supseteq\overline{\mathrm{TNS}}(X)\cup\overline{\mathrm{TNS}}(Y)$;

f) $X\subseteq Y\Rightarrow\underline{\mathrm{TNS}}(X)\subseteq\underline{\mathrm{TNS}}(Y),\overline{\mathrm{TNS}}(X)\subseteq\overline{\mathrm{TNS}}(Y)$。

证明

a)根据定义 18 可以直接得证。

b)根据定义 19,可得

$q\in\underline{\mathrm{TNS}}(X\cap Y)\Rightarrow\exists TN(x)\in\mathrm{TNS}(x)\ \text{s.t.}\ TN(x)\neq\varnothing\wedge TN(x)\subseteq X\cap Y\Rightarrow TN(x)\neq\varnothing,\ TN(x)\subseteq X, TN(x)\subseteq Y\Rightarrow q\in\underline{\mathrm{TNS}}(X), q\in\underline{\mathrm{TNS}}(Y)\Rightarrow q\in\underline{\mathrm{TNS}}(X)\cap\underline{\mathrm{TNS}}(Y)$。

c)性质 c)~e)的证明类似于 b),证略。

d)对$\forall x\in\underline{\mathrm{TNS}}(X)$,必然有 $TN(x)\in\mathrm{TNS}(x)$,其中,$TN(x)\neq\varnothing$ and $TN(x)\subseteq X$。因为 $X\subseteq Y$,则可得 $TN(x)\in Y$,即$x\in\underline{\mathrm{TNS}}(Y)$。类似地,不难证明$\overline{\mathrm{TNS}}(X)\subseteq\overline{\mathrm{TNS}}(Y)$。

remark 1　根据下、上近似空间集对$[\underline{\mathrm{TNS}}(X),\overline{\mathrm{TNS}}(X)]$,以下性质不一定满足:

a) $\underline{\mathrm{TNS}}(U)=U,\overline{\mathrm{TNS}}(U)=U$;

b) $\underline{\mathrm{TNS}}(X)\subseteq X$;

c) $X\subseteq\overline{\mathrm{TNS}}(X)$;

d) $\overline{\mathrm{TNS}}(X\cap Y)\supseteq\overline{\mathrm{TNS}}(X)\cap\underline{\mathrm{TNS}}(Y)$;

e) $\overline{\mathrm{TNS}}(X\cup Y)\subseteq\overline{\mathrm{TNS}}(X)\cup\overline{\mathrm{TNS}}(Y)$;

f) $\underline{\mathrm{TNS}}(\underline{\mathrm{TNS}}(X))=\underline{\mathrm{TNS}}(X),\overline{\mathrm{TNS}}(\overline{\mathrm{TNS}}(X))=\overline{\mathrm{TNS}}(X)$;

g) $\underline{\mathrm{TNS}}(X)=\neg(\overline{\mathrm{TNS}}(\neg X)),\overline{\mathrm{TNS}}(X)=\neg(\underline{\mathrm{TNS}}(\neg X))$;

h) $\underline{\mathrm{TNS}}(\neg X)=\neg\overline{\mathrm{TNS}}(X)$。

其中:$\neg X$ 表示 X 的补集。

定理 3　设 INIS 为一个区间值信息系统,对$\forall A\subseteq AT$,$\forall X\subseteq U$,则有

$\underline{\mathrm{TNS}}_{AT}(X)=\underline{\mathrm{TS}}_{A}(X)\subseteq X\subseteq\overline{\mathrm{TNS}}_{AT}(X)\subseteq\overline{TS}_{A}(X)$

证明

$$x\in\underline{TS}_A(X)\Leftrightarrow x\in Y\wedge Y\subseteq X(Y\in TS(A))\Leftrightarrow$$
$$Y=TN_A(x)\in TNS_A(x)\wedge Y\subseteq X(Y\neq\varnothing)\Leftrightarrow$$
$$x\in\underline{TNS}_A(X)$$
$$x\in\overline{\mathrm{TNS}}_A(X)\Rightarrow\forall TN_A(x)\in\mathrm{TNS}_A(x),TN_A(x)\cap X\neq\varnothing\Rightarrow$$
$$x\in Y\wedge X\cap Y\neq\varnothing(Y\in TS(A))\Rightarrow$$
$$x\in\overline{TS}_A(X)$$

证毕。

例 3　根据定义 18 及例 2 中的结果,可得到基于表 1 的邻域系统为

$$\mathrm{TNS}_{AT}(x_1)=\{K_1\}=\{\{x_1\}\},\mathrm{TNS}_{AT}(x_2)=\{K_2\}=\{\{x_2,x_5\}\},\mathrm{TNS}_{AT}(x_3)=\{K_4\}=\{\{x_3,x_4\}\},\mathrm{TNS}_{AT}(x_4)=\{K_3,K_4\}=\{\{x_3,x_4\},\{x_4,x_5\}\},\mathrm{TNS}_{AT}(x_5)=\{K_2,K_4\}=\{\{x_2,x_5\},\{x_4,x_5\}\},\mathrm{TNS}_{AT}(x_3)=\{K_5\}=\{\{x_6\}\}$$

根据定义 19 和以上结果,可计算出基于领域系统的粗糙近似集为

$$\underline{\mathrm{TNS}}_{AT}(X_1)=\{x_1,x_2,x_5\},\underline{\mathrm{TNS}}_{AT}(X_2)=\varnothing,$$
$$\underline{\mathrm{TNS}}_{AT}(X_3)=\{x_6\}$$
$$\overline{\mathrm{TNS}}_{AT}(X_1)=\{x_1,x_2,x_3,x_4,x_5\},$$
$$\overline{\mathrm{TNS}}_{AT}(X_2)=\{x_3,x_4\},$$
$$\overline{\mathrm{TNS}}_{AT}(X_3)=\{x_6\}$$

以上结果也验证了定理 3 的正确性。

4　结束语

粗糙集理论在邻域系统和包含区间值信息中的扩充已成为当前的热点问题。本文通过灰格运算和 Hausdorff 距离提出了一种基于区间值的邻域关系,并提出了基于邻域关系的粗集模型;在此基础上提出了基于最大相容类和邻域关系的两种粗集模型,分别用于扩大下近似和缩小上近似空间;最后进行了实例分析以验证三种灰色粗集模型之间的关系。笔者下一步的研究方向将是对区间值灰色决策系统中的基于邻域系统的知识约简。

参考文献:

[1] PAWLAK Z,SKOWRON A. Rudiments of rough sets[J]. Information Sciences,2007,177(1):3-27.

[2] PAWLAK Z,SKOWRON A. Rough sets:some extensions[J]. Information Sciences,2007,177(1):28-40.

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

[4] STEFANOWSKI J,TSOUKIAS A. Incomplete information tables and rough classification[J]. Computational Intelligence,2001,17(3):545-566.

[5] ZHUW,WANG Fei-yue. Reduction and axiomization of covering generalized roughsets[J]. Information Sciences,2003,152(1):217-230.

[6] LEUNG Y,LI De-yu. Maximal consistent block technique for rule acquisition in incomplete information systems[J]. Information Sciences,2003,115(1):85-106.

[7] GUAN Yan-yong,WANG Hong-kai,WANG Yun,*et al*. Attribute reduction and optimal decision rules acquisition for continuous valued information systems[J]. Information Sciences,2009,179(17):2974-2984.

[8] LIN T Y. Neighborhood systems:mathematical models of information granulations[C]//Proc of IEEE International Conference on Systems,Man & Cybernetics. 2003:3188-3193.

[9] YANG Xi-bei,LIN T Y. Knowledge operations in neighborhood system[C]//Proc of IEEE International Conference on Granular Computing. 2010:822-825.

[10] HU Qing-hua. Neighborhood rough set based heterogeneous feature subset selection[J]. Information Sciences,2010,178(18):3577-3594.

[11] DENG Ju-long. The primary methods of grey system theory[M]. Wuhan:Huazhong University of Science and Technology Press,2005.

[12] De CARVALHO F A T,De SOUZA R M C R,CHAVENT M,*et al*. Adaptive Hausdorff distances and dynamic clustering of symbolic interval data[J]. Pattern Recognition Letters,2006,27(3):167-179.

[13] YAMAGUCHI D. A grey-based rough approximation model for interval data processing[J]. Information Sciences,2007,117(21):4727-4744.

[14] WU Qiang,LIU Zong-tian. Real formal concept analysis based on grey-rough set theory[J]. Knowledge-based Systems,2009,22(1):38-45.

[15] YANG Ying-jie,JOHN R. Grey sets and greyness[J]. Information Sciences,2012,185(1):249-264.