# 不完备区间值信息系统中邻域粗集模型

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

**摘 要** 目前对未知区间值的研究还处于起步阶段。以包含复杂的遗漏型未知区间值不完备信息系统为研究对象，提出了一种基于灰格运算和 Hausdorff 距离的新的邻域关系。在此基础上，依次提出了邻域关系、最大相容类和邻域系统3种灰色粗集模型。进一步讨论了3种灰色粗集模型之间的上、下近似空间，以提高近似空间的精确度，并用实例进行了分析及验证。

**关键词** 粗糙集，不完备信息系统，区间值，邻域系统，最大相容类，Hausdorff 距离

**中图法分类号** TP18 **文献标识码** A

## Neighborhood Based Rough Sets in Incomplete Interval-valued Information System

WANG Tian-qing[1] XIE Jun[2]
(School of Economics and Management, Wuyi University, Jiangmen 529020, China)[1]
(School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, China)[2]

**Abstract** The study and use of the unknown interval value are still in its infancy. The complex incomplete interval-valued information system in which all unknown values are looked as lost was deeply investigated, and then by using grey lattice operation and Hausdorff distance, we provided a new neighborhood relationship in an interval-valued information system. Furthermore, three forms of rough set models were proposed based on neighborhood relationship, maximal consistent blocks and neighborhood system to improve the accuracy of approximations. Moreover, three numerical examples were employed to substantiate the conceptual arguments.

**Keywords** Rough set, Incomplete information system, Interval data, Neighborhood system, Maximal consistent block, Hausdorff distance

## 1 引言

波兰科学家 Z. 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。但经典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。灰色系统理论是由我国数学家邓聚龙最先提出来的[3]，其中涵括了灰色分类、灰色决策、灰色预测和灰色关系分析等内容，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热点[4-7]，在经典粗糙集方法中对区间值的处理都没有被定义过，因此，对于含区间值信息系统的研究有着重要的理论意义与实用价值。

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，使得经典粗糙集模型的应用受到很多限制。因此，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[8-10]。在不完备信息系统中，Leung 在文献[11]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 等人进一步将最大相容类分别引入集值和连续值系统中[12]。邻域和邻域系统是 Lin 以拓扑空间为基础提出的，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已逐渐成为热点[13-15]。

现实世界中由于数据测量的误差、对数据的理解或获取的限制等众多原因，使得所面临的信息系统往往是不完备的，而大多经典粗集理论都是基于完备系统的，因此在不完备信息系统[16,17]中的研究具有实际意义。在不完备信息系统中，其未知属性值可以有2种不同的含义：1. 所有的未知属性值仅仅是被遗漏的，但又是确实存在的；2. 所有的未知属性值被认为是丢失的，是不允许被比较的。对于包含未知区间值的

到稿日期：2012-06-01 返修日期：2012-09-10 本文受国家自然科学基金(61003288)，江苏省高校自然基金(10KJB520004)，广东省自然科学基金(8452902001001552)资助。

王天擎(1974—)，男，硕士，副教授，CCF 会员，主要研究方向为粗糙集、智能信息处理及数据挖掘，E-mail: wtq_wyu@163.com；谢 军(1973—)，男，博士生，副教授，主要研究方向为人工智能及粗糙集理论。

不完备信息系统研究也处于起步阶段，目前文献[18]中提出了一种填充式不完备区间值信息系统，但是作者发现此方法会简化信息表中的信息数据从而丢失许多信息。因此，本文以含未知区间值的不完备信息系统为研究对象。首先，通过灰色理论中的灰格运算[4]和 Hausdorff 距离[5]提出了一种基于遗漏型未知区间值的邻域关系，用以构成论域(集合)上的一个覆盖；然后，分析邻域关系近似空间分类的不足，提出基于最大相容类方法，构成论述(集合)上的一个完全覆盖，用以扩大下近似空间的范围；进一步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(邻域系统指论域中任一对象的所有邻域的最大相容类组成的族)，用以缩小上近似空间的范围；最后，在一给定的不完备区间值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 3 种粗集模型之间的近似空间关系。

## 2 基本概念

### 2.1 灰格运算与 Hausdorff 距离

**定义 1** 设 $U$ 为一非空论域集合，对象 $x$ 称为论域 $U$ 中的一个元素，$\mathbb{R}$ 为一实数集合，$X\subseteq\mathbb{R}$ 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：

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

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

(3) 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 2** 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：

$$\otimes x=\otimes y\text{，当且仅当}\underline{x}=\underline{y},\overline{x}=\overline{y}\tag{1}$$

**定义 3** 令“$\to$”表示两个灰数$\otimes x$与$\otimes y$包含，称为灰格包含关系，其定义为：

$$\otimes x\to\otimes y\text{，当且仅当 }\underline{x}\geqslant\underline{y},\overline{x}\leqslant\overline{y}\tag{2}$$

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

$k\to\otimes y$，当 $k\geqslant\underline{y},k\leqslant\overline{y}$；

$\widetilde{\otimes}x\to\otimes y$，当$\widetilde{\otimes}x\geqslant\underline{y},\widetilde{\otimes}x\leqslant\overline{y}$。

显然，灰格的包含关系具有自反性、反对称性和传递性。

**定义 4** 对$\otimes x$，$\otimes y$ 和两个灰色，它们之间的灰格运算关系如下：

1. $\otimes x\vee\otimes y=[\min(\underline{x},\underline{y}),\max(\overline{x},\overline{y})]$；

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

2. $$\otimes x\wedge\otimes y=\begin{cases}[\underline{x},\overline{x}], & \text{if }\otimes x\to\otimes y\\ [\underline{y},\overline{y}], & \text{if }\otimes y\to\otimes x\\ [\underline{x},\overline{y}], & \text{if }\underline{x}\to\otimes y\text{ and }\overline{y}\to\otimes x\\ [\underline{y},\overline{x}], & \text{if }\underline{y}\to\otimes x\text{ and }\overline{x}\to\otimes y\\ \varnothing, & \text{otherwise}\end{cases};$$

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

3. $\otimes x^c=\{x\in X^c\mid x<\underline{x},x<\overline{x}\}$；

4. $\otimes x\oplus\otimes y=$

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

**定义 5** Hausdorff 距离经常用来进行图像处理及比较 $A$ 和 $B$ 两个集合，这个距离取决于分别集合 $A$ 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：

$$d_H(A,B)=\max(h(A,B),h(B,A))\tag{3}$$

式中，$h(A,B)=\sup\limits_{u\in A}\inf\limits_{v\in B}\|u-v\|$，有时 $h$ 也被称为定向的 Hausdorff 距离。

**定义 6** 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 9 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

$$d_H^{\otimes}(\otimes x,\otimes y)=\max\{|\underline{x}-\underline{y}|,|\overline{x}-\overline{y}|\}\tag{4}$$

### 2.2 填充式不完备区间值信息系统

**定义 7** 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，$V=V_{AT}=\bigcup\limits_{a\in AT}V_a$，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，$\underline{a(x)}=\inf(a(x))$且$\overline{a(x)}=\sup(a(x))$。

根据定义 4 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作区间值信息系统的一种特殊形式。

**定义 8** 若$\otimes x$为一未知区间值，则其具有以下 3 种情况：

(1)$\otimes x=[\underline{x},*]$，表示区间值的下限值确定，上限值遗漏，记为$\otimes x^*$；

(2)$\otimes x=[*,\overline{x}]$，表示区间值的下限值遗漏，上限值确定，记为$\otimes x_*$；

(3)$\otimes x=[*,*]=$“$*$”，表示区间值的下、上限值同时遗漏，记为$\otimes x_*^*$。显然遗漏型离散值可看作遗漏型区间值的一种特殊形式。

文中将包含未知区间值的区间值信息系统简称为不完备区间值信息系统——$IINIS=\langle U,AT,V^*\rangle$。同时，设本文中的不完备区间值都是遗漏型区间值，则根据对遗漏型区间值中“$*$”的解释：其缺省的下、上限值仅仅是被遗漏的，但又是确实存在的，并可与任何离散值进行比较，文献[18]给出了如下的置换定义。

**定义 9** 在不完备区间值信息系统 $IINIS$ 中，对$\forall a\in AT$，$\forall x\in U$，则可对未知区间值进行以下 3 种方式填充：

(1)当 $a(x)=[\underline{a(x)},*]$时，因缺省上限值要优于或等于$\max\limits_{\forall a\in AT,x_i\in U}\{\overline{a(x)}\}$，故缺省值用$\max\limits_{\forall a\in AT,x_i\in U}\{\overline{a(x)}\}$对“$*$”进行填充；

(2)当 $a(x)=[*,\overline{a(x)}]$时,因其缺省下限值要劣于或等于 $\min\limits_{\forall a\in AT,x_i\in U}\{\underline{a(x)}\}$,故缺省值用 $\min\limits_{\forall a\in AT,x_i\in U}\{\underline{a(x)}\}$对“*”进行填充;

(3)当 $a(x)=[*,*]=$“*”时,同上所述,故用 $\max\limits_{\forall a\in AT,x_i\in U}\{\overline{a(x)}\}$对缺省上限值进行填充,而用 $\min\limits_{\forall a\in AT,x_i\in U}\{\underline{a(x)}\}$对缺省下限值进行填充。

那么将不完备区间值信息系统中所有未知区间值进行填充后,即可得一个填充式不完备区间值信息系统——$FIINIS=\langle U,AT,V^{\Delta}\rangle$。

**例 1** 给定一个不完备区间值决策信息系统,如表 1 所列,其中 $U=\{x_1,x_2,x_3,x_4\}$,$AT=\{a,b\}$。则根据定义 9,可得到表 1 的填充式不完备区间值决策信息系统,如表 2 所列。

表 1 不完备区间值信息系统表

| U | a | b |
|---|---|---|
| $x_1$ | [1.6,1.9] | [2.2,2.9] |
| $x_2$ | [3.0,3.5] | [4.0,5.5] |
| $x_3$ | [1.8,*] | [1.5,3.1] |
| $x_4$ | [2.3,2.0] | [*,*] |

表 2 填充式不完备区间值信息系统表

| U | a | b |
|---|---|---|
| $x_1$ | [1.6,1.9] | [2.2,2.9] |
| $x_2$ | [3.0,3.5] | [4.0,5.5] |
| $x_3$ | [1.8,3.5] | [1.5,3.1] |
| $x_4$ | [2.3,2.0] | [1.5,5.5] |

作者研究发现,在表 1 不完备区间值信息系统直接进行表 2 中的数据填充时,会简化信息表中的信息数据而丢失许多信息,从而使得结果失真。例如在表 1 的对象 $x_4$ 中,$b(x_4)$为$\otimes x^*_*$型未知区间值,根据遗漏型区间值的解释,其可以和其它所有对象中属性 $b$ 的区间值距离都为零,但在表 2 中,$b(x_4)$与其它对象中属性 $b$ 的区间值距离显然各有不同,因此其不适合基于邻域关系的处理。所以,作者在不改变信息表中原始数据的情况下,根据未知区间值的各种不同情况定义了一种新的领域关系。

## 3 基本不完备区间值的邻域粗集模型

### 3.1 邻域近似空间

**定义 10** 设 $U$ 是一个非空论域集合,$\delta=\{C_i\mid i\in I\}$为 $U$ 上一个子集族。若$\bigcup\limits_{i\in I}C_i=U$,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 **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:

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

因此,根据覆盖 **C**,可以导出论域 $U$ 上的邻域系统为:

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

**定义 11** 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

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

式中,$\Delta$ 是一个距离函数,满足以下性质:

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

(2)$\Delta(x,y)=0$,当且仅当 $x=y$;

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

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

**定义 12** 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:

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

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

### 3.2 不完备区间值信息系统中基于领域关系的粗集模型

**定义 13** 设 $IINIS$ 为一个不完备区间值信息系统,对$\forall A\subseteq AT$,记

$$SR_A=\{(x,y)\in U^2:\forall a\in A,D_H^{\otimes *}(a(x),a(y))=1\} \quad (8)$$

式中,

$$D_H^{\otimes *}(a(x),a(y))=\begin{cases}1, & \text{当 } a(x)=\otimes a(x)\wedge a(y)=\otimes a(y)\wedge d_H^{\otimes}(a(x),a(y))\leqslant\beta\\ 1, & \text{当 } a(x)=\otimes a(x)^*_*\vee a(y)=\otimes a(y)^*_*\\ 1, & \text{当}(a(x)=\otimes a(x)^*\vee a(y)=\otimes a(y)^*)\wedge|\underline{a(x)}-\underline{a(y)}|\leqslant\beta\\ 1, & \text{当}(a(x)=\otimes a(x)_*\vee a(y)=\otimes a(y)_*)\wedge|\overline{a(x)}-\overline{a(y)}|\leqslant\beta\\ 1, & \text{当 } a(x)=\otimes a(x)_*\wedge a(y)=\otimes a(y)^*\wedge(\underline{a(y)}-\overline{a(x)})\leqslant\beta\\ & (a(x)^*=\otimes a(x)\wedge a(y)=\otimes a(y)_*\text{ 类似})\\ 0, & \text{其它}\end{cases}$$

则 $SR_A$ 称为由 $A$ 决定 $U$ 上的邻域关系。显然 $SR_A$ 是自反和对称的,但非传递。

**定义 14** 设 $IINIS$ 为一不完备区间值信息系统,$\forall x\in U$,记 $SR_A(x)$是 $x$ 关于 $SR_A$ 的邻域类,有

$$SR_A(x)=\{y\in U,(x,y)\in R_A\} \quad (9)$$

则 $SR_A(x)$是 $x$ 由属性 $A$ 所构成的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 $U/R_{AT}=\{SR_{AT}(x):x\in U\}$。

**定义 15** 设 $IINIS$ 为一个不完备区间值信息系统,对于$A\subseteq AT$,$\forall X\subseteq U$,$X$ 基于 $SR(A)$的下、上近似集分别记为$\underline{SR_A}(X)$,$\overline{SR_A}(X)$且

$\underline{SR_A}(X)=\{x\in U:SR_A(x)\subseteq X\}$

$\overline{SR_A}(X)=\{x\in U:SR_A(x)\cap X\neq\emptyset\}$

**性质 1** 根据定义 12,很容易证明以下性质:

(1)$\underline{SR_A}(X)=\{x\in X:SR_A(x)\subseteq X\}\neq\bigcup\{SR_A(x):SR_A(x)\subseteq X\}$;

(2)$\overline{SR_A}(X)=\bigcup\{SR_A(x):x\in X\}\neq\bigcup\{SR_A(x):SR_A(x)\cap X\neq\emptyset\}$。

例 2 表 1 为一个不完备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。在表 1 中决策属性的值为 $V_d=\{1,2,3\}$,因为决策系统做出的决策通常是清晰明确的。令 $U/IND(d)=\{X_1,X_2,X_3\}=\{\{x_1,x_2,x_4,x_6\},\{x_3\},\{x_5\}\}$。

表 3 不完备区间值决策信息系统表

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

设$\beta$=1.0,则根据定义 13 和定义 14 可得:

$SR_{AT}(x_1)=\{x_1,x_4\}$,$SR_{AT}(x_2)=\{x_2,x_3\}$

$SR_{AT}(x_3)=\{x_2,x_3,x_4\}$,$SR_{AT}(x_4)=\{x_1,x_3,x_4\}$

$SR_{AT}(x_5)=\{x_5\}$,$SR_{AT}(x_6)=\{x_6\}$

根据定义 15 可得:

$\underline{SR_A}(X_1)=\{x_1,x_6\}$,$\underline{SR_A}(X_2)=\emptyset$,$\underline{SR_A}(X_3)=\{x_5\}$

$\overline{SR_A}(X_1)=\{x_1,x_2,x_3,x_4,x_6\}$,$\overline{SR_A}(X_2)=\{x_2,x_3,x_4\}$,$\overline{SR_A}(X_3)=\{x_5\}$。

在对上例的研究中发现,邻域关系还存在以下一些不足:(1)不同的邻域类可以存在包含关系,例如 $SR_{AT}(x_1)$和 $SR_{AT}(x_4)$,有 $SN_{AT}(x_1)\subseteq SN_{AT}(x_4)$;(2)在同一邻域类中,并不是所有两个对象都满足邻域关系,例如在 $SN_{AT}(x_4)$中,$x_2$ 和 $x_4$ 不满足邻域关系 $SR_{AT}$。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。

## 4 基于区间值的邻域系统粗集模型

### 4.1 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:(1)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;(2)在一个最大相容类中若再加入其它任何一个对象,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 $IINIS$ 为一个不完备区间值信息系统,对 $\forall A\subseteq AT$,记

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

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

**性质 2** 设 $IINIS$ 为一个不完备区间值信息系统,对 $\forall A\in AT$,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

$$SR_A(x)=\bigcup\{K\in TS(A): K\in S_A(x)\}=\bigcup\{K: K\in TS_x(A)\}$$

证明:显然有 $SR_A(x)\supseteq\bigcup\{K\in TS(A): K\subseteq SR_A(x)\}$,则只需要证明 $SR_A(x)\subseteq\bigcup\{K\in TS(A): K\subseteq SR_A(x)\}$。对 $\forall y\in SR_A(x)$,$\{x,y\}$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,我们可得 $K\subseteq SR_A(x)$,则有 $y\subseteq\bigcup\{K\in TS(A): K\subseteq SR_A(x)\}$,进一步可得 $SR_A(x)\subseteq\bigcup\{K\in TS(A): K\subseteq SR_A(x)\}$。证毕。

**定义 17** 设 $IINIS$ 为一个不完备区间值信息系统,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

$$\underline{TS_A}(X)=\bigcup\{K\in TS(A): K\subseteq X\}$$

$$\overline{TS_A}(X)=\bigcup\{K\in TS(A): K\cap X\neq\emptyset\}$$

**性质 3** 设 $IINIS$ 为一个不完备区间值信息系统,对 $\forall A\subseteq AT$ 和 $\forall X\subseteq U$,则有:

1. $\underline{TS_A}(X)\subseteq X\subseteq\overline{TS_A}(X)$;

2. 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 $IINIS$ 为一个不完备区间值信息系统,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:首先,证明 $\underline{SR_A}(X)\subseteq\underline{TS_A}(X)$。假设 $x\in\underline{SR_A}(X)$,根据$\underline{SN_A}(X)$的定义,有 $SR_A(x)\subseteq X$。再根据性质 2,有 $SR_A(x)=\bigcup\{K: K\in TS_x(A)\}\subseteq X$,则对于 $\forall K\subseteq TS_x(A)$,$K\subseteq X$,因此可得 $x\in\underline{TS_A}(X)$。

然后,再证明$\overline{TS_A}(X)=\overline{SR_A}(X)$。根据性质 1,有$\overline{SR_A}(X)=\{x\in U: SRA(x)\cap X\neq\emptyset\}=\bigcup\{SR_A(x): x\in X\}$。根据性质 2,对 $\forall x\in U$,即可得 $SR_A(x)=\bigcup\{K: K\in TS_x(A)\}$。因此有$\overline{SR_A}(X)=\bigcup\{SR_A(x): x\in X\}=\bigcup\{K\in TS(A): K\cap X\neq\emptyset\}=\overline{TS_A}(X)$。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 3 通过在例 2 中计算得到的邻域 $SR_{AT}(U)$,根据定义 16、定义 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:

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

$\underline{TS_{AT}}(X_1)=\{x_1,x_4,x_6\}$,$\underline{TS_{AT}}(X_2)=\emptyset$,$\underline{TS_{AT}}(X_3)=\{x_5\}$

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

以上结果也验证了定理 1 的正确性。

### 4.2 基于邻域系统的粗集模型

**定义 18** 设 $IINIS$ 为一个不完备区间值信息系统,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:

$$TNS_A(x)=\{K: K\in TS(A): x\in K\}$$

那么,进一步基于 $U$ 的邻域系统则可定义为:

$$TNS_A(U)=\{TNS_A(x): x\in U\}$$

**定义 19** 设 $U$ 为一非空论域集合,$TNS(U)$为 $U$ 上一个给定的邻域系统,对 $\forall X\subseteq U$,$X$ 基于邻域系统的下、上近似集分别记为$\underline{TNS_{AT}}(X)$和$\overline{TNS_{AT}}(X)$且:

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

$$\overline{TNS_{AT}}(X)=\{x\in U: \forall T(x)\in TM(x)\ \text{s.t.}\ T(x)\cap\emptyset\}$$

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

(1)$\underline{TNS}(\emptyset)=\overline{TNS_{AT}}(\emptyset)=\emptyset$;

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

(3) $\overline{TNS}(X\cap Y)\subseteq\overline{TNS}(X)\cap\overline{TNS}(Y)$；

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

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

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

证明：

(a)根据定义18,(1)可以直接得证。

(b)根据定义16,可得

$q\in\underline{TNS}(X\cap Y)$

$\Rightarrow\exists\, TN(x)\in TNS(x)\ \text{s.t.}\ TN(x)\neq\emptyset\wedge TN(x)\subseteq X\cap Y$

$\Rightarrow TN(x)\neq\emptyset, TN(x)\subseteq X, TN(x)\subseteq Y$

$\Rightarrow q\in\underline{TNS}(X), q\in\underline{TNS}(Y)$

$\Rightarrow q\in\underline{TNS}(X)\cap\underline{TNS}(Y)$

(c)性质(3)—性质(5)的证明与(b)类似,证略。

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

**定理3** 设$IINIS$为一个不完备区间值信息系统,对$\forall A\subseteq AT,\forall X\subseteq U$,则有

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

证明：

$x\in\underline{TS}_A(X)\Leftrightarrow x\in Y\wedge Y\subseteq X(Y\in TS(A))$

$\Leftrightarrow Y=TN_A(x)\in TNS_A(x)\wedge Y\subseteq X(Y\neq\emptyset)$

$\Leftrightarrow x\in\underline{TNS_A}(X)$

$x\in\overline{TNS}_A(X)\Rightarrow\forall\, TN_A(x)\in TNS_A(x), TN_A(x)\cap X\neq\emptyset$

$\Rightarrow x\in Y\wedge X\cap Y\neq\emptyset(Y\in TS(A))$

$\Rightarrow x\in\overline{TS}_A(X)$。证毕。

**例4** 根据定义18及例3中的结果,可得到表3不完备区间值决策信息系统的邻域系统为：

$TNS_{AT}(x_1)=\{K_1\}=\{\{x_1,x_4\}\}, TNS_{AT}(x_2)=\{K_2\}=\{\{x_2,x_3\}\}$

$TNS_{AT}(x_3)=\{K_4\}=\{\{x_2,x_3\},\{x_3,x_4\}\}, TNS_{AT}(x_4)=\{K_3,K_4\}=\{\{x_1,x_4\},\{x_3,x_4\}\}$

$TNS_{AT}(x_5)=\{K_2,K_4\}=\{\{x_5\}\}, TNS_{AT}(x_3)=\{K_5\}=\{\{x_6\}\}$

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：

$\underline{TNS}_{AT}(X_1)=\{x_1,x_4,x_6\},\underline{TNS}_{AT}(X_2)=\emptyset,\underline{TNS}_{AT}(X_3)=\{x_5\}$

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

以上结果也验证了定理3的正确性。

**结束语** 粗糙集理论在邻域系统和区间值中的扩充已成为当前的热点问题,但对不完备区间值信息系统的研究还处于起步阶段。本文在不改变信息系统中数据的情况下,利用灰格运算和Hausdorff距离提出了一种邻域关系以及基于邻域关系的粗集模型。接着在此基础之上,提出了基于最大相容类和邻域关系的两种粗集模型,分别用以扩大下近似和缩小上近似空间,以提高分类的精确度,并进行了实例分析以验证3种灰色粗集模型之间的关系。笔者下一步的研究方向是不完备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: 3-27

[2] Pawlak Z, Skowron A. Rough sets: some extensions [J]. Information Sciences, 2007, 177: 28-40

[3] 邓聚龙. 灰色系统理论基本方法[M]. 武汉：华中科技大学出版社, 2005

[4] Yamaguchi D. A grey-based rough approximation model for interval data processing[J]. Information Sciences, 2007, 177: 4727-4744

[5] Francisco De A T, Carvalho D, Renata M C R. Adaptive Hausdorff distances and dynamic clusteringof symbolic interval data [J]. Pattern Recognition Letters, 2006, 27: 167-179

[6] Wu Q, Liu Z T. Real formal concept analysis based on grey-rough set theory[J]. Knowledge-Based Systems, 2009, 22: 38-45

[7] Yang Y J, John R. Grey sets and greyness [J]. Information Sciences, 2012, 185: 249-264

[8] Kryszkiewicz M. Rough set approach to incomplete information systems[J]. Information Sciences, 1998, 112: 39-49

[9] Stefanowski J, Tsoukias A. Incomplete information tables and rough classification[J]. Computational Intelligence, 2001, 17: 545-566

[10] Zhu W, Wang F Y. Reduction and axiomization of covering generalized roughsets[J]. Information Sciences, 2003, 152: 217-230

[11] Leung Y, Li D Y. Maximal consistent block technique for rule acquisition in incomplete information systems[J]. Information Sciences, 2003, 115: 85-106

[12] Guan Y Y, Wang H K, Wang Y, et al. Attribute reduction and optimal decision rules acquisition for continuous valued information systems[J]. Information Sciences, 2009, 179: 2974-2984

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

[14] Yang X B, Lin T Y. Knowledge operations in neighborhood system[C] // 2010 IEEE International Conference on Granular Computing. 2010: 822-825

[15] Hu Q H. Neighborhood rough set based heterogeneous feature subset selection[J]. Information Sciences, 2010, 178: 3577-3594

[16] 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

[17] Yang X B, Zhang M. Neighborhood systems-based rough sets in incomplete information system[J]. Knowledge-Based Systems, 2011, 24: 858-867

[18] Yang X B, Yu D J, Yang J Y. Dominance-based rough set approach to incomplete interval-valued information system [J]. Data & Knowledge Engineering, 2009, 68: 1331-1347