Pure Mathematics 理论数学, 2022, 12(2), 264-275
Published Online February 2022 in Hans. http://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2022.122031

信息系统中的局部邻域粗糙集及属性约简

切洛太，傅　丽*

青海民族大学数学与统计学院，青海　西宁

收稿日期：2022年1月4日；录用日期：2022年2月4日；发布日期：2022年2月11日

摘　要

信息粒度和近似方法是粗糙集理论中两个重要描述数据的方法，为了解决经典邻域粗糙集的计算效率低下和数据识别有用性不足的问题，有人提出了局部邻域粗糙集模型。在局部邻域粗糙集中，邻域半径的大小直接影响数据的有效性，因此邻域半径的取值为至关重要，但在局部邻域粗糙集中邻域半径的取值范围有点过大。为了进一步缩小邻域半径取值范围，本文借助邻域半径的取值，讨论了系统的协调性。首先给出了不协调信息系统中的局部邻域粗糙集，然后将属性约简不协调集的思想引入邻域，缩小了邻域半径的取值范围，为计算局部邻域粗糙集的邻域半径提供了有效工具，同时研究了相关的性质。

关键词

不协调信息系统，邻域半径，局部粗糙集，属性约简

Local Neighborhood Rough Sets and Attribute Reduction in Information Systems

Luotai Qie, Li Fu*

School of Mathematics and Statistics, Qinghai University for Nationalities, Xining Qinghai

Received: Jan. 4th, 2022; accepted: Feb. 4th, 2022; published: Feb. 11th, 2022

Abstract

The information granular and approximation method is two important methods to describe data in rough set theory. In order to work out the computational inefficiency and lack of cognitive data in the classical neighborhood rough set, then someone proposed the local neighborhood rough set model. For the local neighborhood rough set model, the value of neighborhood radius is very of great importance since its value has a close relation with the efficiency. However, the value range

*通讯作者。

文章引用: 切洛太, 傅丽. 信息系统中的局部邻域粗糙集及属性约简[J]. 理论数学, 2022, 12(2): 264-275.
DOI: 10.12677/pm.2022.122031

of neighborhood radius in the local neighborhood rough set is too large. Further, in order to reduce the value range of neighborhood radius, this paper studies the coordination of the system. Firstly, the local neighborhood decision rough sets of the uncoordinated information system are given, then the idea of the property reduction coordination set is added to the neighborhood, which can reduce the range of radius of the neighborhood, provide an effective tool for calculating neighborhood radius of rough set. The related properties are also studied.

Keywords

The Information System of Uncoordinated Decision, Neighborhood Radius, Local Rough Set, Attribute Reduction

1. 引言

粗糙集理论是 1982 年波兰数学家 Palawk [1]第一次提出来的，是处理不确定性和不完整性数据的数学工具，在很多领域如人工智能、数据挖掘、医疗分析、模式识别等中应用广泛。此后，很多学者提出并研究了各种粗糙集模型，如 Yao [2]等将贝叶斯决策理论引入概率粗糙集模型，提出了决策粗糙集模型和三支决策[3]，给出了概率阈值(α,β)的最佳值求解方法。经典粗糙集只能处理名义型数据，其思想是根据不可分辨关系划分等价类，确定上下近似。但实际应用中，大部分数据是数值型数据，应用这种数据时首先要对原始数据进行离散化处理，在处理过程中必定造成信息丢失，从而文[4]提出了一种基于邻域关系的邻域粗糙集模型，文[5]提出了一种基于邻域的决策理论粗糙集模型。鉴于由二元关系导出的粗糙集模型，它的上、下近似是由预先计算的信息粒所构成，其中信息粒用来近似一个目标概念，计算时这些信息粒必须遍历给定论域中的所有对象。因此这种计算方法时间复杂度高，不能满足大数据高效计算的要求，从而文[6]提出了局部粗糙集的理论框架，文[7]提出局部邻域粗糙集。本文在文[7]的基础上讨论了系统的协调性，将不协调集思想引入邻域半径，证明它缩小了邻域半径的取值范围，这为计算局部邻域决策粗糙集的邻域半径提供了有效的工具。文章结构如下，第一部分回顾局部粗糙集、邻域半径、局部邻域决策的一些基本概念。第二部分给出基于局部邻域的协调决策系统。第三部分，讨论基于局部邻域的不协调决策系统。第四部分是总结和后期工作。

2. 基础知识

定义 1 [4] [8] [9]假设信息系统$S=(U,A,f,V)$的属性集为A所有属性的值域$V=\bigcup_{a\in A}V_a$，$f:U\times A\to V$，$\forall a\in A$，$\forall x\in U$，$f(x,a)\in V_a$，$B\subseteq A$，x在属性集B上的δ-邻域定义为

$$\delta_B(x)=\{y\in U \mid dis(x,y)\le\delta,\delta>0\}.$$

其中，$dis(\bullet)$表示任意对象之间的欧氏距离。

定义 2 [4] [10]设信息系统$S=(U,A,f,V)$，$\forall x\in U$，$B\subseteq A$，x在属性集B上对X的粗糙隶属度$\mu_B(x)$定义为

$$\mu_B(x)=P(X\mid\delta_B(x))=\frac{|X\cap\delta_B(x)|}{|\delta_B(x)|}.$$

其中 $P\left(X \mid \delta_B\left(x\right)\right)$ 表示分类的条件概率，$|\cdot|$ 表示集合中元素的个数。

定义 3 [11]假设 $S=\left(U,A,f,V\right)$ 为一个信息系统，$X\subseteq U$，$B\subseteq A$，则属性集 B 下 X 关于 δ-邻域的局部粗糙集上下近似集分别定义为

$$\overline{\delta_B}\left(x\right)=\left\{x\in X \mid \delta_B\left(x\right)\cap X\neq\varnothing\right\},$$

$$\underline{\delta_B}\left(x\right)=\left\{x\in X \mid \delta_B\left(x\right)\subseteq X\right\}.$$

属性集 B 下 X 关于 δ-邻域的局部粗糙集的正域、负域和边界域分别定义为

$$POS_B\left(X\right)=\underline{\delta_B}\left(x\right)=\left\{x\in X \mid P\left(X \mid \delta_B\left(x\right)\right)=1\right\},$$

$$NEG_B\left(X\right)=U-\overline{\delta_B}\left(x\right)=\left\{x\in X \mid P\left(X \mid \delta_B\left(x\right)\right)=0\right\},$$

$$BND_X\left(X\right)=\overline{\delta_B}\left(x\right)-\underline{\delta_B}\left(x\right)=\left\{x\in X \mid 0<P\left(X \mid \delta_B\left(x\right)\right)<1\right\}.$$

定义 4 [11]设信息系统 $S=\left(U,AT=C\cup D,f,V\right)$ 的条件属性 C，决策属性为 D，$X\subseteq U$，$B\subseteq C$ 则属性集 B 下 X 关于 δ-邻域的局部决策粗糙集上下近似集分别定义为

$$\overline{\delta_B^{DT}}\left(x\right)=\left\{x\in X \mid P\left(X \mid \delta_B\left(x\right)\right)>\beta\right\},$$

$$\underline{\delta_B^{DT}}\left(x\right)=\left\{x\in X \mid P\left(X \mid \delta_B\left(x\right)\right)\geq\alpha\right\}.$$

正域、负域和边界域分别定义为

$$POS_B^{DT}\left(X\right)=\underline{\delta_B^{DT}}\left(x\right)=\left\{x\in X \mid P\left(X \mid \delta_B\left(x\right)\right)\geq\alpha\right\},$$

$$BND_B^{DT}\left(X\right)=\overline{\delta_B^{DT}}\left(x\right)-\underline{\delta_B^{DT}}\left(x\right)=\left\{x\in X \mid \beta<P\left(X \mid \delta_B\left(x\right)\right)<\alpha\right\},$$

$$NEG_B^{DT}\left(X\right)=U-\overline{\delta_B^{DT}}\left(x\right)=\left\{x\in X \mid P\left(X \mid \delta_B\left(x\right)\right)\leq\beta\right\}.$$

决策粗糙集有两种状态，所有属于 X 的对象集 X 和所有不属于 X 的对象集 X^c，用 $\Omega=\left(X,X^c\right)$ 表示。

定义 5 [12] [13]设信息系统 $S=\left(U,AT=C\cup D,f,V\right)$ 的条件属性 C，决策属性为 D，若 $\delta_A\left(X\right)\subseteq U/R_d$，$R_d=\left\{\left(x_i,x_j\right) \mid d\left(x_i\right)=d\left(x_y\right)\right\}$，则称 $S=\left(U,AT=C\cup D,f,V\right)$ 为协调决策信息系统。

定义 6 [14]设信息系统 $S=\left(U,AT=C\cup D,f,V\right)$ 的条件属性 C，决策属性为 D，若 $\delta_A\left(X\right)\not\subseteq U/R_d$，$R_d=\left\{\left(x_i,x_j\right) \mid d\left(x_i\right)=d\left(x_y\right)\right\}$，则称 $S=\left(U,AT=C\cup D,f,V\right)$ 为不协调决策信息系统。

设 $U/R_d=\left\{X_1,X_2,\cdots,X_n\right\}$，$B\subseteq C$，则 U/R_d 上的概率分布函数记为

$$\mu_B^{DT}\left(x_i\right)=\left(P\left(X_1 \mid \delta_B\left(x_i\right)\right),P\left(X_2 \mid \delta_B\left(x_i\right)\right),\cdots,P\left(X_n \mid \delta_B\left(x_i\right)\right)\right),$$

进一步记

$$m_B\left(x_i\right)=\max_{t\leq n}P\left(X_t \mid \delta_B\left(x_i\right)\right)=P\left(X_{t_0} \mid \delta_B\left(x_i\right)\right)\left(x_i\in U\right),$$

称 $m_B\left(x_i\right)$ 是不确定性命题规则“若 $y\in\delta_B\left(x_i\right)$，则 $y\in X_{t_0}$“的可信度，记为

$$\eta_B\left(x_i\right)=\left\{X_{t_0} \mid P\left(X_{t_0} \mid \delta_B\left(x_i\right)\right)=\max_{t\leq n}P\left(X_t \mid \delta_B\right)\left(x_i\right)\left(x_i\in U\right)\right\}.$$

定义 7 [14]设 $S=\left(U,AT=C\cup D,f,V\right)$ 为一个不协调决策信息系统，$B\subseteq C$。

1) 若 $\forall x_i\in U$，有 $\mu_B^{DT}\left(x_i\right)=\mu_C^{DT}\left(x_i\right)$，则称 B 是分布协调集。若 B 是分布协调集且 B 的任何真子集都不是分布协调集，则称 B 为分布约简集。

2) 若 $\forall x_i \in U$，有 $\eta_B(x_i) = \eta_C(x_i)$，则称 B 是最大分布协调集。若 B 是最大分布协调集且 B 的任何真子集都不是最大分布协调集，称 B 为最大分布约简集。

定义 8 [14]设 $S = (U, AT = C \cup D, f, V)$ 为一个不协调决策信息系统，$B \subseteq C$。

1) 若 $\underline{\delta_C^{DT}}(x) = \underline{\delta_B^{DT}}(x)$，则 B 称下近似协调集。如果下近似协调集 B 任何子集都不是它的下近似协调集，那么 B 称下近似约简集。

2) 若 $\overline{\delta_C^{DT}}(x) = \overline{\delta_B^{DT}}(x)$，则 B 称上近似协调集。如果上近似协调集 B 的任何子集都不是它的上近似协调集，那么 B 称上近似约简集。

3. 局部邻域粗糙集的协调性

这部分，首先用一个例子来说明系统的协调性，以及如何把一个不协调决策信息系统转化为协调的。

例 1 设表 1 为一个不协调决策信息系统，U 为有限论域，$U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$，条件属性集 $C = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7\}$，决策属性为 D。

Table 1. Information system
表 1. 信息系统

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	d
x_1	0.7	0.87	0.7	0.61	0.76	0.68	0.66	是
x_2	0.8	0.73	0.68	0.65	0.9	0.8	0.76	是
x_3	0.75	0.84	0.65	0.5	0.45	0.6	0.56	否
x_4	0.48	0.5	0.62	0.8	0.63	0.7	0.53	否
x_5	0.67	0.6	0.52	0.52	0.65	0.6	0.48	否
x_6	0.9	0.8	0.68	0.76	0.8	0.82	0.5	是
x_7	0.9	0.93	0.6	0.65	0.58	0.7	0.72	是
x_8	0.9	0.83	0.3	0.56	0.5	0.56	0.37	否
x_9	0.54	0.8	0.45	0.5	0.62	0.38	0.7	否

$$U/R_d = \{\{x_1, x_2, x_6, x_7\}, \{x_3, x_4, x_5, x_8, x_9\}\} = \{X_1, X_2\},$$

由定义 5 知若决策系统协调，则 $\delta_A(X) \subseteq U/R_d$，$R_d = \{(x_i, x_j) \mid d(x_i) = d(x_y)\}$，即 $\delta_A(x_i) \in U/R_d, (i = 1, 2, \cdots, 9)$。

当 $\delta_1 = 0.34$ 时，$\delta_A(x_1) = \{x_1, x_2, x_6, x_7\} \subseteq X_1$。当 $\delta_1 > 0.34$ 时，$\delta_A(x_1) \notin U/R_d$。因此当 $\delta_1 \le 0.34$ 时，$\delta_A(x_1) \in U/R_d$。按照上述方法有

$\delta_2 \le 0.5$ 时，$\delta_A(x_2) \in U/R_d$。

$\delta_3 \le 0.34$ 时，$\delta_A(x_3) \in U/R_d$。

$\delta_4 \le 0.51$ 时，$\delta_A(x_4) \in U/R_d$。

$\delta_5 \le 0.4$ 时，$\delta_A(x_5) \in U/R_d$。

$\delta_6 \le 0.49$ 时，$\delta_A(x_6) \in U/R_d$。

$\delta_7 \le 0.5$ 时，$\delta_A(x_7) \in U/R_d$ 。

$\delta_8 \le 0.34$ 时，$\delta_A(x_8) \in U/R_d$ 。

$\delta_9 \le 0.46$ 时，$\delta_A(x_9) \in U/R_d$ 。

令 $\delta = \min\{\delta_1,\delta_2,\delta_3,\delta_4,\delta_5,\delta_6,\delta_7,\delta_8,\delta_9\} = \min\{0.34,0.5,0.34,0.51,0.4,0.49,0.5,0.34,0.46\} = 0.34$ 。则有 $\delta_A(X) = \{\{x_1,x_2,x_6,x_7\},\{x_1,x_7\},\{x_3\},\{x_4,x_5\},\{x_3,x_4,x_5,x_9\},\{x_3,x_5,x_8\},\{x_3,x_5,x_9\}\} \subseteq R_d = \{\{x_1,x_2,x_6,x_7\},\{x_3,x_4,x_5,x_8,x_9\}\}$ 。也就是说，当 $\delta \le 0.34$ 时，此系统转化为协调的。

命题 1 设 $S=(U,AT=C\cup D,f,V)$ 为一个信息系统，当邻域半径的取值范围为 $[0,\delta']$ 时，这里 $\delta' = \min\left\{\delta_i \mid \delta_i = \sup\left\{\delta_j \mid \delta_{jA}(x_i) = \left\{y \mid dis_A(x_i,y) \le \delta_j, \delta_j > 0\right\} \subseteq U/R_d\right\}, \delta_i > 0\right\}, (i=0,1,\cdots,7)$，则此系统是协调的。

证明： 反证法。设 $S=(U,AT=C\cup D,f,V)$ 为协调系统，取 $\delta > \min\left\{\delta_i \mid \delta_i = \sup\left\{\delta_j \mid \delta_j A(x_i) = \left\{y \mid dis_A(x_i,y) \le \delta_j, \delta_j > 0\right\} \subseteq U/R_d\right\}, \delta_i > 0\right\}$，则总能找到一个 $i_0, (i_0 = 1,2,\cdots,7)$，使得 $\delta > \delta_{i_0} = \sup\left\{\delta_j \mid \delta_{jA}(x_{i_0}) = \left\{y \mid dis_A(x_{i_0},y) \le \delta_j > 0\right\} \subseteq U/R_d\right\}, \delta_j > 0$ 满足 $\delta_A(x_{i_0}) \notin U/R_d$，则 $\delta_A(X) \not\subseteq U/R_d$ 。由定义 6 知与假设矛盾，命题 1 得证。

设 X 为不可定义集，属性 $B \subseteq C$ 下 X 关于 δ-邻域的局部邻域决策粗糙集上下近似不相等，取 $\delta \in [0,\delta']$，这里 $\delta' = \min\left\{\delta_i \mid \delta_i = \sup\left\{\delta_j \mid \delta_{jA}(x_i) = \left\{y \mid dis_A(x_i,y) \le \delta_j, \delta_j > 0\right\} \subseteq U/R_d\right\}, \delta_i > 0\right\}, (i=0,1,\cdots,7)$，则系统变成协调的，有当 $X \subseteq U, X = X_1$ 时，由局部性得到

$$\delta_C(x_1) = \{x_1,x_2,x_6,x_7\} \subseteq X_1,$$

$$\delta_C(x_2) = \{x_1,x_2,x_6,x_7\} \subseteq X_1,$$

$$\delta_C(x_6) = \{x_1,x_2,x_6,x_7\} \subseteq X_1,$$

$$\delta_C(x_7) = \{x_1,x_7\} \subseteq X_1.$$

当 $X \subseteq U, X = X_2$ 时，由局部性得到

$$\delta_C(x_3) = \{x_3\} \subseteq X_2,$$

$$\delta_C(x_4) = \{x_4,x_5\} \subseteq X_2,$$

$$\delta_C(x_5) = \{x_3,x_4,x_5,x_9\} \subseteq X_2,$$

$$\delta_C(x_8) = \{x_3,x_5,x_8\} \subseteq X_2,$$

$$\delta_C(x_9) = \{x_3,x_5,x_9\} \subseteq X_2.$$

从而属性集 C 下 X_t 关于 δ-邻域的局部决策粗糙集上下近似集分别为

$$\overline{\delta_B^{DT}}(X_1) = \left\{x \in X_1 \mid P\left(X_1 \mid \delta_B(x)\right) = 1 > \beta\right\} = X_1 = \underline{\delta_B^{DT}}(X_1)(0 \le \beta \le \alpha \le 1),$$

$$\underline{\delta_B^{DT}}(X_2) = \left\{x \in X_2 \mid P\left(X_2 \mid \delta_B(x)\right) = 1 \ge \alpha\right\} = X_2 = \overline{\delta_B^{DT}}(X_2)(0 \le \beta \le \alpha \le 1).$$

命题 2 设 $S=(U,AT=C\cup D,f,V)$ 为一个协调决策信息系统，$X_t \in U/d, (t=1,2,\cdots,n), n=|U/R_d|$，邻域半径的取值范围为 $[0,\delta']$。则 $\overline{\delta_B^{DT}}(X_t) = \underline{\delta_B^{DT}}(X_t)$，$X_t$ 是可定义集。

证明：设 $S=(U,AT=C\cup D,f,V)$ 为一个协调决策信息系统，若 $\overline{\delta_B^{DT}}(X_{t_0})\neq\underline{\delta_B^{DT}}(X_{t_0}),(t_0=1,2,\cdots,n)$，则存在 $x_{i_0}(i_0=1,2,\cdots,7)$，$\delta_jA(x_{i_0})\not\subseteq X_{t_0}$ 且 $\delta_jA(x_{i_0})\cap X_{t_0}\neq\varnothing$，从而有 $\delta_jA(x_{i_0})\notin U/R_d$，$\delta'_A(X)\not\subseteq U/R_d$，这与假设矛盾，命题 2 得证。

性质 1 设 $S=(U,AT=C\cup D,f,V)$ 为一个协调决策信息系统，$X_1,X_2\subseteq U/R_d$，$B_1,B_2\subseteq AT$，$\delta\subseteq[0,\delta']$，$0\le\beta<\alpha\le1$。下列等式成立

1) $\underline{\delta_B^{DT}}(X)=X;\overline{\delta_B^{DT}}(X)=X$

2) $\underline{\delta_B^{DT}}(\varnothing)=\overline{\delta_B^{DT}}(\varnothing)=\varnothing,\underline{\delta_B^{DT}}(U)=\overline{\delta_B^{DT}}(U)=U$

3) $\underline{\delta_B^{DT}}(X_1\cap X_2)=\underline{\delta_B^{DT}}(X_1)\cap\underline{\delta_B^{DT}}(X_2);\overline{\delta_B^{DT}}(X_1\cap X_2)=\overline{\delta_B^{DT}}(X_1)\cap\overline{\delta_B^{DT}}(X_2)$

4) $\underline{\delta_B^{DT}}(X_1\cup X_2)=\underline{\delta_B^{DT}}(X_1)\cup\underline{\delta_B^{DT}}(X_2);\overline{\delta_B^{DT}}(X_1\cup X_2)=\overline{\delta_B^{DT}}(X_1)\cup\overline{\delta_B^{DT}}(X_2)$

证明：由命题 2 知，$\underline{\delta_B^{DT}}(X)=\overline{\delta_B^{DT}}(X)$，从而上述 4 条性质容易得正。

4. 局部邻域粗糙集的属性约简

在基于局部邻域的协调系统中，集合 X_i 是可定义的，在基于局部邻域的不协调信息系统中，X_i 是不可定义的和粗糙的，因此我们只考虑系统为不协调的情况。

性质 2 设 $S=(U,A,f,V)$ 为一个不协调信息系统，$X_1,X_2\subseteq U$，$B_1,B_2\subseteq A$。下列等式成立

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

2) $\underline{\delta_B}(\varnothing)=\overline{\delta_B}(\varnothing)=\varnothing,\underline{\delta_B}(U)=\overline{\delta_B}(U)=U$

3) $\underline{\delta_B}(X_1\cap X_2)\subseteq\underline{\delta_B}(X_1)\cap\underline{\delta_B}(X_2);\overline{\delta_B}(X_1\cap X_2)\subseteq\overline{\delta_B}(X_1)\cap\overline{\delta_B}(X_2)$

4) $\underline{\delta_B}(X_1\cup X_2)\supseteq\underline{\delta_B}(X_1)\cup\underline{\delta_B}(X_2);\overline{\delta_B}(X_1\cup X_2)\supseteq\overline{\delta_B}(X_1)\cup\overline{\delta_B}(X_2)$

5) 当 $\delta^1\le\delta^2$ 时，$\underline{\delta_B^1}(X)\supseteq\underline{\delta_B^2}(X);\overline{\delta_B^1}(X)\subseteq\overline{\delta_B^2}(X)$

证明：1) 设 $\forall x\in\underline{\delta_B}(X)$，由定义 2 有 $x\in\underline{\delta_B}(x)=\{x\in X\mid\delta_B(x)\subseteq X\}\Rightarrow x\in X$，$\underline{\delta_B}(X)\subseteq X$。

2) 设 $\forall x\in\underline{\delta_B}(\varnothing)$ 由定义 2 有 $\delta_B(x)\subseteq\varnothing\Rightarrow\delta_B(x)=\varnothing$，则 $\underline{\delta_B}(\varnothing)=\varnothing$ 同样 $x\in\delta_B(x)\cap\varnothing=\varnothing\Rightarrow\overline{\delta_B}(\varnothing)=\varnothing$。

3) 设 $X_1,X_2\subseteq U$，$\forall x\in\underline{\delta_B}(X_1\cap X_2)\Rightarrow x\in\delta_B(x)\subseteq X_1\cap X_2$，$x\in\delta_B(x)\subseteq X_1\wedge x\in\delta_B(x)\subseteq X_2\Rightarrow x\in\underline{\delta_B}(X_1)\cap\underline{\delta_B}(X_2)$，则 $\underline{\delta_B}(X_1\cap X_2)\subseteq\underline{\delta_B}(X_1)\cap\underline{\delta_B}(X_2)$。

4) 设 $X_1,X_2\subseteq U$，$\forall x\in\underline{\delta_B}(X_1\cup X_2)\Rightarrow x\in\delta_B(x)\subseteq X_1\cup X_2$，$x\in\delta_B(x)\subseteq X_1\vee x\in\delta_B(x)\subseteq X_2\Rightarrow x\in\underline{\delta_B}(X_1)\cup\underline{\delta_B}(X_2)$，则 $\underline{\delta_B}(X_1\cup X_2)\subseteq\underline{\delta_B}(X_1)\cup\underline{\delta_B}(X_2)$。另外，$\forall x\in\overline{\delta_B}(X_1\cup X_2)\Leftrightarrow x\in(X_1\cup X_2)\cap\delta_B(x)\Leftrightarrow x\in(X_1)\cap\delta_B(x)\vee x\in(X_2)\cap\delta_B(x)$
$\Leftrightarrow\overline{\delta_B}(X_1\cup X_2)=\overline{\delta_B}(X_1)\cup\overline{\delta_B}(X_2)$

5) 设 $\delta^1\le\delta^2$，$\forall x\in\underline{\delta_B^2}(X)\Rightarrow\delta_B^1(x)\subseteq\delta_B^2(x)$，$x\in X$，由定义 3 有 $x\in\underline{\delta_B^2}(X)\subseteq\underline{\delta_B^1}(X)$，$x\in\underline{\delta_B^1}(X)$，则 $\underline{\delta_B^1}(X)\supseteq\underline{\delta_B^2}(X)$。另外，$\forall x\in\overline{\delta_B^1}(X)\Rightarrow\delta_B^1(x)\subseteq\delta_B^2(x)$，由定义 3 有 $x\in\delta_B^1(x)\cap X\subseteq\delta_B^2(x)\cap X$，$x\in\overline{\delta_B^2}(X)$，则 $\overline{\delta_B^1}(X)\subseteq\overline{\delta_B^2}(X)$

性质 3 设 $S=(U,AT=C\cup D,f,V)$ 为一个不协调决策信息系统，$X_1,X_2\subseteq U/R_d$，$B_1,B_2\subseteq AT$，$0\le\beta<\alpha\le1$，下列等式成立

1) $\underline{\delta_B^{DT}}(X)\subseteq X;X\subseteq\overline{\delta_B^{DT}}(X)$

2) $\underline{\delta_B^{DT}}(\varnothing)=\overline{\delta_B^{DT}}(\varnothing)=\varnothing,\underline{\delta_B^{DT}}(U)=\overline{\delta_B^{DT}}(U)=U$

3) $\underline{\delta_B^{DT}}(X_1\cap X_2)\subseteq\underline{\delta_B^{DT}}(X_1)\cap\underline{\delta_B^{DT}}(X_2);\overline{\delta_B^{DT}}(X_1\cap X_2)\subseteq\overline{\delta_B^{DT}}(X_1)\cap\overline{\delta_B^{DT}}(X_2)$

4) $\underline{\delta_B^{DT}}(X_1\cup X_2)\supseteq\underline{\delta_B^{DT}}(X_1)\cup\underline{\delta_B^{DT}}(X_2);\overline{\delta_B^{DT}}(X_1\cup X_2)\supseteq\overline{\delta_B^{DT}}(X_1)\cup\overline{\delta_B^{DT}}(X_2)$

证明：由定义 4 的局部性，1)和 2)很容易得证。

3) 设 $X_1,X_2\subseteq U/R_d$，$\forall x\in\overline{\delta_B^{DT}}(X_1\cap X_2)\Rightarrow x\in X_1\cap X_2$，$P(X_1\cap X_2\mid\delta_B(x))>\beta\Rightarrow x\in X_1$，

$P\left(X_1 | \delta_B(x)\right) \ge P\left(X_1 \cap X_2 | \delta_B(x)\right) > \beta$，且 $x \in X_2$，
$P\left(X_2 | \delta_B(x)\right) \ge P\left(X_1 \cap X_2 | \delta_B(x)\right) > \beta \Rightarrow x \in \overline{\delta_B^{DT}}(X_1) \cap \overline{\delta_B^{DT}}(X_2)$

4) 设 $X_1, X_2 \subseteq U/R_d$，$\forall x \in \overline{\delta_B^{DT}}(X_1) \cup \overline{\delta_B^{DT}}(X_2) \Rightarrow x \in X_1$，$P\left(X_1 | \delta_B(x)\right) > \beta \vee x \in X_2$，
$P\left(X_2 | \delta_B(x)\right) > \beta \Rightarrow x \in X_1 \cup X_2$，
$P\left(X_1 \cup X_2 | \delta_B(x)\right) \ge P\left(X_1 | \delta_B(x)\right) \vee P\left(X_2 | \delta_B(x)\right) > \beta \Rightarrow x \in \overline{\delta_B^{DT}}(X_1 \cup X_2)$。

在性质 3 中性质 2 的第 5 条不成立，因此，我们给出下面的推论。

推论 1 设 $\delta^1 \le \delta^2$，当 $\underline{\delta_B^{1DT}}(X) \subseteq \underline{\delta_B^{2DT}}(X)$ 时，$\forall x \in X$ 存在着一个正整数 n，使得 δ^n 为 x 的最佳邻域半径。

证明：设 $\delta^1 \le \delta^2$，当 $\underline{\delta_B^{1DT}}(X) \subseteq \underline{\delta_B^{2DT}}(X)$。则有 $\delta^1 \le \delta^2 \le \cdots \le \delta^n$，使得 $\underline{\delta_B^{1DT}}(X) \subseteq \underline{\delta_B^{2DT}}(X) \subseteq \cdots \subseteq \underline{\delta_B^{nDT}}(X) \not\subseteq \underline{\delta_B^{(n+1)DT}}(X)$，这说明当 $\delta > \delta^n$ 时，$\delta_B(x)$ 与 X 具有的共同元素不会随着 δ 的增大而增多，反而 $p\left(X | \delta_B^n(x)\right) \ge p\left(X | \delta_B^{n+1}(x)\right) \ge \cdots$。从而正域中的元素越来越少，这不利于处理数据的效果，因此，δ^n 为 x 的最佳邻域半径。

任何信息系统按照邻域半径取值范围的不同而可分为协调的部分和不协调的部分。原邻域半径的取值为 $\delta \in \left[\min\left\{dis_B(x,y) | dis_B(x,y), x, y \in U\right\}, \max\left\{dis_B(x,y) | dis_B(x,y), x, y \in U\right\}\right]$ 从上面的命题发现，当 $\delta \in \left[\min\left\{dis_B(x,y) | dis_B(x,y), x, y \in U\right\}, \delta'\right]$ 时，可定义的，也就是说边界域为空集，从而我们需要重点研究不可定义的部分，也就是边界域为非空的尽可能把它变小。

例 2 某单位选高管时，为把德才兼备的人才提拔到领导岗位上，制定了 5 个标准。a_1——政策水平，a_2——工作作风，a_3——业务能力，a_4——口才，a_5——近十年的请假次数，d——是否适合当选高管。现有 7 个目标，即 $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$。先给 7 位候选人按照 5 项指标进行打分，打分情况如表 2~5。表 2 为信息系统，$U/R_d = \{Y_1, Y_2\} = \{\{x_1, x_2, x_6\}, \{x_3, x_4, x_5, x_7\}\}$，
$\delta' = \min\{\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7\} = \min\{0.32, 0.31, 0.44, 0.31, 0.49, 0.44, 0.32\} = 0.31$。

Table 2. Judge 1
表 2. 评委 1

	a_1	a_2	a_3	a_4	a_5	d
x_1	70	80	50	60	50	0.7
x_2	80	50	70	90	60	0.65
x_3	60	60	30	50	90	0.6
x_4	50	30	80	70	70	0.5
x_5	30	40	60	50	50	0.4
x_6	80	90	70	60	40	0.7
x_7	60	30	60	50	40	0.4

Table 3. Judge 2
表 3. 评委 2

	a_1	a_2	a_3	a_4	a_5	d
x_1	80	50	60	70	50	0.7
x_2	50	70	60	80	60	0.8

Continued

x_3	60	60	50	60	50	0.6
x_4	70	50	40	60	50	0.5
x_5	50	60	40	50	40	0.4
x_6	70	60	50	50	50	0.6
x_7	40	50	60	30	40	0.4

Table 4. Judge 3
表 4. 评委 3

	a_1	a_2	a_3	a_4	a_5	d
x_1	70	80	60	60	50	0.75
x_2	80	50	70	90	60	0.7
x_3	50	50	30	50	70	0.4
x_4	60	50	60	60	60	0.6
x_5	50	40	40	40	50	0.3
x_6	70	80	60	50	50	0.72
x_7	40	50	40	50	40	0.3

Table 5. Judge 4
表 5. 评委 4

	a_1	a_2	a_3	a_4	a_5	d
x_1	70	60	80	70	60	0.7
x_2	60	50	50	70	50	0.6
x_3	70	40	60	50	80	0.65
x_4	60	50	60	50	50	0.5
x_5	70	40	40	60	50	0.4
x_6	70	70	60	40	50	0.65
x_7	60	30	70	30	40	0.45

a_1, a_2, a_3, a_4 均为效益型，则用公式 $r_{ij} = \dfrac{a_{ij}}{\underset{i}{\max}\left(a_{ij}\right)}$，$a_5$ 为成本型，则用公式 $r_{ij} = \dfrac{\underset{i}{\min}\left(a_{ij}\right)}{a_{ij}}$ [15]从而得到规范矩阵 R_i，如表 6~9。

Table 6. R_1
表 6. R_1

	a_1	a_2	a_3	a_4	a_5	d
x_1	0.88	0.89	0.63	0.67	0.8	0.7
x_2	1	0.56	0.88	1	0.67	0.6
x_3	0.75	0.67	0.38	0.56	0.44	0.65
x_4	0.63	0.33	1	0.78	0.57	0.5
x_5	0.38	0.44	0.75	0.56	0.8	0.4
x_6	1	1	0.88	0.67	1	0.65
x_7	80	90	70	60	40	0.7

Table 7. R_2
表 7. R_2

	a_1	a_2	a_3	a_4	a_5	d
x_1	1	0.71	1	0.88	0.8	0.7
x_2	0.63	1	1	1	0.67	0.8
x_3	0.75	0.86	0.83	0.75	0.8	0.6
x_4	0.88	0.71	0.67	0.75	0.8	0.5
x_5	0.63	0.86	0.67	0.63	1	0.4
x_6	0.88	0.86	0.83	0.63	0.8	0.6

Table 8. R_3
表 8. R_3

	a_1	a_2	a_3	a_4	a_5	d
x_1	0.88	1	0.86	0.67	1	0.75
x_2	1	0.63	1	1	0.83	0.7
x_3	0.63	0.63	0.43	0.56	0.71	0.4
x_4	0.75	0.63	0.86	0.67	0.83	0.6
x_5	0.63	0.5	0.57	0.44	1	0.3
x_6	1	1	0.88	0.67	1	0.72

Table 9. R_4
表 9. R_4

	a_1	a_2	a_3	a_4	a_5	d
x_1	1	0.86	1	1	0.83	0.7
x_2	0.86	0.71	0.63	1	1	0.6
x_3	1	0.57	0.75	0.71	0.63	0.65
x_4	0.86	0.71	0.75	0.71	1	0.5
x_5	1	0.57	0.5	0.86	1	0.4
x_6	1	1	0.75	0.57	1	0.65

最后我们用平均法，$r_{j,t}=\frac{\sum r_{j,t}^{i}}{n}, i=1,\cdots,n; j=1,\cdots,7; t=1,\cdots,5$ 其中 i 代表评委，j 代表目标，t 为条件属性。比如 $r_{1,3}=\frac{\sum r_{1,3}^{i}}{4}=\frac{0.63+1+0.86+1}{4}=0.87$ 也就是说目标概念 x_1 在条件属性 a_3 下的值取所有表 R_i 中对应值的平均值。由此得到下列的表 R 如表 10。

Table 10. R
表 10. R

	a_1	a_2	a_3	a_4	a_5	d
x_1	0.94	0.87	0.87	0.81	0.86	是
x_2	0.87	0.73	0.88	1	0.79	是
x_3	0.78	0.68	0.6	0.65	0.65	否
x_4	0.78	0.6	0.82	0.73	0.8	否
x_5	0.67	0.6	0.62	0.62	0.95	否
x_6	0.94	0.97	0.83	0.61	0.95	是
x_7	0.72	0.62	0.84	0.74	0.75	否

当 $\delta\le 0.31$ 时，此系统为协调的。Y_1,Y_2 是可定义的。

当 $\delta>0.31$ 时，此系统为不协调的，Y_1,Y_2 是不可定义的。因此，设 $B=\{a_2,a_3,a_4,a_5\}\subseteq C$，$\delta=0.4$ (δ 在 $(0.31,0.54]$ 中取值)，

$$\mu_C^{DT}(x_1)=(0.6,0.4)=\mu_B^{DT}(x_1),$$

$$\mu_C^{DT}(x_2)=(0.5,0.5)=\mu_B^{DT}(x_2),$$

$$\mu_C^{DT}(x_3)=(0,1)=\mu_B^{DT}(x_3),$$

$$\mu_C^{DT}(x_4)=(0.33,0.67)=\mu_B^{DT}(x_4),$$
$$\mu_C^{DT}(x_5)=(0,1)=\mu_B^{DT}(x_5),$$
$$\mu_C^{DT}(x_6)=(1,0)=\mu_B^{DT}(x_6),$$
$$\mu_C^{DT}(x_7)=(0.33,0.67)=\mu_B^{DT}(x_7),$$

则 $\mu_C^{DT}(x_i)=\mu_B^{DT}(x_i)$，由定义 9(1)知 B 是分布协调集，它的任何子集都不是它的分布协调集。因此 B 是分布约简集。同理 B 还是最大分布约简集。由于局部性，取 $\alpha=0.6,\beta=0.45,(0\le\beta\le\alpha\le1)$ 则有

$$\underline{\delta_C^{DT}}(X)=\left(\delta_C^{DT}(X_1),\delta_C^{DT}(X_2)\right)=\left(\{x_1,x_6\},\{x_3,x_4,x_5,x_7\}\right)=\underline{\delta_B^{DT}}(X),$$
$$\overline{\delta_C^{DT}}(X)=\left(\delta_C^{DT}(X_1),\delta_C^{DT}(X_2)\right)=\left(\{x_1,x_2,x_6\},\{x_3,x_4,x_5,x_7\}\right)=\overline{\delta_B^{DT}}(X),$$

B 的任何子集都不是它的上(下)近似协调集。因此 B 是上(下)近似约简集。

命题 3 对于信息系统 $S=(U,AT=C\cup D,f,V)$，若邻域半径在 $\left(\delta',\max\limits_{i,j\le n}\left(dis_C(x_i,x_j)\right)\right]$ 内取值时，则它不协调，这里

$\delta'=\min\left\{\delta_i\mid\delta_i=\sup\left\{\delta_j\mid\delta_jA(x_i)=\{y\mid dis_A(x_i,y)\le\delta_j,\delta_j>0\}\subseteq U/R_d\right\},\delta_i>0\right\},(i=0,1,\cdots,7)$。

证明： 设 $S=(U,AT=C\cup D,f,V)$ 为不协调信息系统且邻域半径 $\delta\le\delta'$，则由命题 1 得，存在一个 $\delta_A(x_{i_0})\in U/R_d$，从而 S 为协调的，与假设矛盾。因此 $\delta\in\left(\delta',\max\limits_{i,j\le n},\left(dis_C(x_i,x_j)\right)\right]$。

5. 总结

在局部邻域粗糙集中邻域半径的取值范围过大而找到最佳邻域半径有所困难，因此本文研究了协调系统下的局部邻域粗糙集和不协调系统的局部邻域粗糙集及属性约简，借助邻域信息系统的协调性，发现任何一个邻域信息系统都可分为协调的部分和不协调的部分，用这种方法缩小了邻域半径的取值范围，为计算最佳邻域半径提供了方法。后期继续研究乐观和悲观下的邻域半径及属性约简。

基金项目

青海民族大学研究生创新项目(项目编号：07M2021005)。

参考文献

[1] Pawlak, Z. (1982) Rough Sets. *International Journal of Computer and Information Sciences*, **11**, 341-356. https://doi.org/10.1007/BF01001956

[2] Yao, Y.Y. and Wong, S.K.M. (1992) A Decision Theoretic Framework for Approximating Concepts. *International Journal of Man-Machine Studies*, **37**, 793-809. https://doi.org/10.1016/0020-7373(92)90069-W

[3] Yao, Y.Y. (2010) Three-Way Decisions with Probabilistic Rough Sets. *Information Sciences*, **180**, 341-353. https://doi.org/10.1016/j.ins.2009.09.021

[4] Hu Q.H., Yu, D.R., Liu, J.F. and Wu, C. (2008) Neighborhood Rough Set Based Heterogeneous Feature Subset Selection. *Information Sciences*, **178**, 3577-3594. https://doi.org/10.1016/j.ins.2008.05.024

[5] Li, W.W., Huang, Z.Q., Jia, X.Y. and Cai, X.Y. (2016) Neighborhood Based Decision-Theoretic Rough Set Models. *International Journal of Approximate Reasoning*, **69**, 1-17. https://doi.org/10.1016/j.ijar.2015.11.005

[6] Qian, Y.H., Liang, X.Y., Wang, Q., Liang, J., Liu, B., Skowron, A., et al. (2018) Local Rough Set: A Solution to Rough Data Analysis in Big Data. *International Journal of Approximate Reasoning*, **97**, 38-63. https://doi.org/10.1016/j.ijar.2018.01.008

[7] Wang, Q., Qain, Y.H., Liang, X.Y., Guo, Q. and Liang, J. (2018) Local Neighborhood Rough Set. *Knowledge-Based*

Systems, **153**, 53-64. https://doi.org/10.1016/j.knosys.2018.04.023

[8] Lin, T.Y. (1998) Granular Computing on Binary Relations I: Data Mining and Neighborhood Systems. *Proceedings of the International Workshop on Rough Sets and Knowledge Discovery*: *Rough Sets, Fuzzy Sets and Knowledge Discovery*, 107-121.

[9] Li, G.-H., Hao W.-J. and Li, Z.-W. (2018) Generalized Consistent Space Based on Distance and Its Attribute Reduction. *Fuzzy Systems and Mathematics*, **32**, 150-154.

[10] Ziarko, W. (1993) Variable Precision Rough Set Model. *Journal of Computer and System Sciences*, **46**, 39-59. https://doi.org/10.1016/0022-0000(93)90048-2

[11] 孙颖, 蔡天使, 张毅, 鞠恒荣, 丁卫平. 基于合理粒度的局部邻域决策粗糙计算方法[J]. 南京大学学报(自然科学), 2021, 57(2): 262-271.

[12] Greco, S., Matarazzo, B. and Slowinski, R. (2001) Rough Sets Theory for Multicriteria Decision Analysis. *European Journal of Operational Research*, **129**, 1-47. https://doi.org/10.1016/S0377-2217(00)00167-3

[13] 张文修, 仇国芳. 基于粗糙集的不确定决策[M]. 北京: 清华大学出版社, 2005.

[14] Slowinski, R., Stefanowski, J., Greco, S. and Matarazzo, B. (2000) Rough Set Based Processing of Inconsistent Information in Decision analysis. *Control and Cybernetics*, **29**, 379-404.

[15] Hwang, C.L. and Yoon, K. (1981) Multiple Attribute Decision Making: Methods and Applications. Springer-Verlag, New York. https://doi.org/10.1007/978-3-642-48318-9