Advances in Applied Mathematics 应用数学进展, 2020, 9(8), 1298-1308
Published Online August 2020 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2020.98152

Hans汉斯

Rule Extraction of Multi-Granularity Probabilistic Rough Sets in Incomplete Information Systems

Xiaoan Shi*, Li Fu

School of Mathematics and Statistics, Qinghai Nationalities University, Xining Qinghai
Email: *1156028589@qq.com, fl0971@163.com

Received: Jul. 27th, 2020; accepted: Aug. 13th, 2020; published: Aug. 20th, 2020

Abstract

Based on incomplete multi-granularity information system, this paper mainly discusses the upper-lower approximation and approximate accuracy of probabilistic rough sets, and the rules extraction based on dominance-disadvantage relationship. Firstly, the optimistic (pessimistic) multi-granularity rough approximation in incomplete multi-granularity information systems is introduced. Secondly, it discusses the optimistic (pessimistic) multi-granularity lower-upper approximation of probabilistic rough sets based on the similar relation, dominance relation and disadvantage relation respectively, the rule of approximate accuracy variation with α, β is given. Finally, it discusses rule extraction of probabilistic rough sets based on the variation of dominance and disadvantage relationship with parameter γ.

Keywords

Probabilistic Rough Set Model, Rule Extraction, Dominance and Disadvantage Relationship, Incomplete Information Multi-Granularity System

不完备信息系统中多粒度概率粗糙集的规则提取

石孝安*，傅　丽

青海民族大学数学与统计学院，青海　西宁
Email: *1156028589@qq.com, fl0971@163.com

收稿日期：2020年7月27日；录用日期：2020年8月13日；发布日期：2020年8月20日

*通讯作者。

文章引用: 石孝安, 傅丽. 不完备信息系统中多粒度概率粗糙集的规则提取[J]. 应用数学进展, 2020, 9(8): 1298-1308.
DOI: 10.12677/aam.2020.98152

摘　要

本文基于不完备多粒度信息系统，主要讨论了概率粗糙集的上下近似和精度的变化，以及基于优劣关系的规则提取。首先，对不完备信息系统中多粒度粗糙集关于乐观(悲观)的上下近似进行了简单的介绍。其次，基于相似关系优势关系和劣势关系，对不完备信息系统中的多粒度概率粗糙集从乐观(悲观)的角度对其上下近似进行了讨论，得出了近似精度随 α,β 的变化而变化的规律。最后，讨论了基于优劣关系随参数 γ 变化的概率粗糙集的规则提取。

关键词

概率粗糙集模型，规则提取，优劣关系，不完备多粒度信息系统

Copyright © 2020 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

粗糙集是 1982 年英国数学家 Pawlak 提出的理论[1]，该理论是一种处理不确定的，不精确信息的数学工具。自提出以来，得到了很多学者的关注和研究，并已应用到了决策分析模式识别数据挖掘等领域[2] [3] [4]。

经典粗糙集模型是一种可以被要求利用信息是完全性的一种模型，却忽略了随机，文献[5]提出了多粒度概率粗糙集模型集，解决了经典的粗糙集中下近似过于严格和上近似过于宽松的问题。但是，却没有考虑到信息系统缺失属性值的情况。文献[6]，在不完备的决策信息系统中定义了粒度的属性缺失，粒度组合重要度。文献[7]，根据乐观和悲观的角度，对贝叶斯决策进行了讨论并对决策信息系统进行了约简的讨论。文献[8]，对经典的不完备信息系统进行了讨论和约简。文献[9]，对多粒度粗糙集的粒度约简算法进行了讨论。文献[10] [11]，分别从容差关系模型拓展领域等角度对不完备信息系统进行了研究。

本文是基于不完备信息系统进行的讨论，主要讨论了概率粗糙集的优劣关系及规则提取。在本文当中，第二部分介绍了相关概念；第三部分我们基于相似关系，对多粒度概率粗糙集进行了讨论；第四部分基于劣势关系和优势关系分别从乐观和悲观的角度对不完备的多粒度概率粗糙集进行了讨论，并给出了随 α,β 的变化规律与精度变化的规律。第五部分，讨论了基于优劣关系随参数 γ 变化的概率粗糙集的规则提取。

2. 相关概念

定义 1 [5]：设 $S=(U,C\cup D,V,f)$ 为一个决策信息系统，其中 U 为论域；$A=C\cup D$ 为有限属性集合；C 为条件属性集合，D 为决策属性集合 $V=\bigcup_{a\in A}V_a$，V_a 为属性 $a\in A$ 的值域；$f:U\times A\to V$ 是一个信息函数，$\forall a\in A,x\in U,f(x,a)\in V_a$。在缺失部分条件属性值的状况下(用*表示)，被称为不完备的决策信息系统；否则，S 是一个完备决策信息系统。

定义 2 [5]：设 $S=(U,C\cup D,V,f)$ 为一个不完备决策信息系统，$A=\{A_1,A_2,\cdots,A_m\}$ 是 C 的 m 个属性的子集族。$\forall X\subseteq U$，X 关于 A 的

(1) 乐观多粒度上(下)近似定义：

$$\underline{\sum_{i}^{m} A_i^o}(X)=\left\{x:T_{\Lambda_1}(x)\subseteq X\vee T_{\Lambda_2}(x)\subseteq X\vee\cdots\vee T_{\Lambda_m}(x)\subseteq X,x\in U\right\}$$

$$\overline{\sum_{i=1}^{m} A_i^o}(X)=\sim\left(\sum_{i=1}^{m} A_i^o(\sim X)\right)$$

(2) 悲观多粒度下(上)近似分别定义：

$$\underline{\sum_{i=0}^{m} A_i^p}(X)=\left\{x:T_{A_1}(x)\subseteq X\wedge T_{A_2}(x)\subseteq X\wedge\cdots\wedge T_{A_m}(x)\subseteq X,x\in U\right\}$$

$$\overline{\sum_{i=1}^{m} A_i^p}(X)=\sim\left(\underline{\sum_{i=1}^{m} A_i^p}(\sim X)\right)$$

定义 3 [8]：设 $I=(U,A,F)$ 是一个不完备信息系统，称 R_B 为其相似关系，如果满足

$$R_B=\left\{(x,y)\in U\otimes U\mid\forall a\in B,a(x)=a(y)\text{ 或 }a(x)=*\text{ 或 }a(y)=*\right\}$$

相似关系 R_B 是自反而且满足对称性，但通常不满足传递性。记作 $S_B(x)=\left\{y\in U\mid(x,y)\in R_B\right\}$，则 $S_B(x)$ 被称为 x 关于 R_B 的相似类；记 U/R_B 是构成 U 的一个划分。

定义 4 [8]：设 $I=(U,A,F)$ 是不完备信息系统

(1) 称 $R_A^{\geq}$ 为优势关系，当满足

$$R_A^{\geq}=\left\{(y,x)\in U\otimes U\mid V_a\in A,f_a(y)\geq f_a(x)\text{ 或 }f_a(x)=*\text{ 或 }f_a(y)=*\right\}$$

时，可以利用优势类关系归类：$U/R_A^{\geq}=\left\{[x]_A^{\geq}\mid x\in U\right\}$，其中：$[x]^{\geq}=\left\{y\in U\mid(y,x)\in R_A^{\geq}\right\}$，$[x]^{\leq}=\left\{y\in U\mid(x,y)\in R_A^{\leq}\right\}$；$[x]_A^{\geq}$ 表示在属性 A 上优于 x 的优势类，$[x]_A^{\leq}$ 表示在属性 A 上劣于 x 的劣势类。

(2) 称 $R_A^{\leq}$ 为劣势关系，当满足

$$R_A^{\leq}=\left\{(y,x)\in U\otimes U\mid V_a\in A,f_a(y)\leq f_a(x)\text{ 或 }f_a(x)=*\text{ 或 }f_a(y)=*\right\}$$

时，可以利用劣势类关系归类：$U/R_A^{\leq}=\left\{[x]_A^{\leq}\mid x\in U\right\}$，其中：$[x]^{\geq}=\left\{y\in U\mid(y,x)\in R_A^{\leq}\right\}$，$[x]^{\leq}=\left\{y\in U\mid(x,y)\in R_A^{\leq}\right\}$；$[x]_A^{\geq}$ 表示在属性 A 上优于 x 的优势类，$[x]_A^{\leq}$ 表示在属性 A 上劣于 x 的劣势类。

注：(1) 下文中的所有讨论基于不完备信息系统进行讨论，小标题中不在体现。

(2) 根据定义 2 中悲观多粒度的上下近似，在下面的文章中，可类似于乐观多粒度概率粗糙集的上下近似的讨论方式得出其相关定理和性质，本文不再给出。

3. 基于相似关系的多粒度概率粗糙集模型

定义 5：设 $MIS=\langle U,AT,V,f\rangle$ 为一个不完备多粒度信息系统，$\forall x\in U$，$X\subseteq U$，$0\leq\beta\leq 0.5<\alpha\leq 1$，则 X 在相似关系 R_B 上依参数 α,β 的乐观(悲观)多粒度概率粗糙集下近似 $\underline{\sum_{i=1}^{m} A_i^o}_{\alpha}(X)$ 和上近似 $\overline{\sum_{i=1}^{m} A_i^o}_{\beta}(X)$ 分别定义：

$$\underline{\sum_{i=1}^{m} A_i^o}_{\alpha}(X)=\left\{x\in U:p\left(X\mid[x]_{A_1}^{S_B}\right)\geq\alpha\vee P\left(X\mid[x]_{A_2}^{S_B}\right)\geq\alpha\vee\cdots\vee P\left(X\mid[x]_{A_m}^{S_B}\right)\geq\alpha\right\}$$

$$\overline{\sum_{i=1}^{m} A_i^o}_{\beta}(X) = \sim\underline{\sum_{i=1}^{m} A_i^o}_{1-\beta}(\sim X)$$

通过定义 5，可以得到在不完备系统中，乐观多粒度概率粗糙集的边界域集合：

$$Bnd^o_{\sum_{i=1}^{m} A_i(\alpha,\beta)} = \overline{\sum_{i=1}^{m} A_i^o}_{\beta}(X) - \underline{\sum_{i=1}^{m} A_i^o}_{\alpha}(X)$$

二元组 $\left\langle \underline{\sum_{i=1}^{m} A_i^o}_{\alpha}(X), \overline{\sum_{i=1}^{m} A_i^o}_{\beta}(X) \right\rangle$ 称作是 X 基于相似关系 R_B 关于粒度 $A_i\,(i=1,2,\cdots,m)$ 的乐观多粒度概率粗糙集。

定理 3.1： 设 $MIS=\langle U,AT,V,f\rangle$ 为一个不完备的多粒度信息系统，$\forall x\in U$，$X\subseteq U$，$0\le\beta\le 0.5<\alpha\le 1$，$AT=\{A_1,A_2,\cdots,A_m\}$ 则

$$\overline{\sum_{i=1}^{m} A_i^o}_{\beta}(X) = \left\{x\in U: p\left(X\mid[x]_{A_1}^{S_B}\right)>\beta\wedge P\left(X\mid[x]_{A_2}^{S_B}\right)>\beta\wedge\cdots\wedge P\left(X\mid[x]_{A_m}^{S_B}\right)>\beta\right\}$$

证明：

$$\begin{aligned}
&\overline{\sum_{i=1}^{m} A_i^o}_{\beta}(X) = \sim\underline{\sum_{i=1}^{m} A_i^o}_{1-\beta}(\sim X)\\
&=\left\{x\in U: p\left(\sim X\mid[x]_{A_1}^{S_B}\right)\ge 1-\beta\vee P\left(\sim X\mid[x]_{A_2}^{S_B}\right)\ge 1-\beta\vee\cdots\vee P\left(\sim X\mid[x]_{A_m}^{S_B}\right)\ge 1-\beta\right\}\\
&=\left\{x\in U: 1-p\left(X\mid[x]_{A_1}^{S_B}\right)\ge 1-\beta\vee 1-P\left(X\mid[x]_{A_2}^{S_B}\right)\ge 1-\beta\vee\cdots\vee 1-P\left(X\mid[x]_{A_m}^{S_B}\right)\ge 1-\beta\right\}\\
&=\left\{x\in U: \frac{\left|\sim X\cap[x]_{A_1}^{S_B}\right|}{\left|[x]_{A_1}^{S_B}\right|}\ge 1-\beta\vee\frac{\left|\sim X\cap[x]_{A_2}^{S_B}\right|}{\left|[x]_{A_2}^{S_B}\right|}\ge 1-\beta\vee\cdots\vee\frac{\left|\sim X\cap[x]_{A_m}^{S_B}\right|}{\left|[x]_{A_m}^{S_B}\right|}\ge 1-\beta\right\}\\
&=\left\{x\in U: 1-\frac{\left|X\cap[x]_{A_1}^{S_B}\right|}{\left|[x]_{A_1}^{S_B}\right|}\ge 1-\beta\vee 1-\frac{\left|X\cap[x]_{A_2}^{S_B}\right|}{\left|[x]_{A_2}^{S_B}\right|}\ge 1-\beta\vee\cdots\vee 1-\frac{\left|X\cap[x]_{A_m}^{S_B}\right|}{\left|[x]_{A_m}^{S_B}\right|}\ge 1-\beta\right\}\\
&=\left\{x\in U: p\left(X\mid[x]_{A_1}^{S_B}\right)>\beta\wedge P\left(X\mid[x]_{A_2}^{S_B}\right)>\beta\wedge\cdots\wedge P\left(X\mid[x]_{A_m}^{S_B}\right)>\beta\right\}
\end{aligned}$$

定义 6： 设 $MIS=\langle U,AT,V,f\rangle$ 为一个不完备的多粒度信息统，$\forall x\in U$，$X\subseteq U$，$AT=\{A_1,A_2,\cdots,A_m\}$，$0\le\beta\le 0.5<\alpha\le 1$，$X$ 在相似关系 R_B 上依参数 α,β 的乐观(悲观)多粒度概率粗糙集近似精度 $\lambda^o(\alpha,\beta)$ 分别定义为：

$$\lambda^o(\alpha,\beta)=\frac{\left|\underline{\sum_{i}^{m} A_i^{\,o}}_{\alpha}(X)\right|}{\left|\overline{\sum_{i}^{m} A_i^{\,o}}_{\beta}(X)\right|}\text{；}\quad \lambda^p(\alpha,\beta)=\frac{\left|\underline{\sum_{i}^{m} A_i^{\,p}}_{\alpha}(X)\right|}{\left|\overline{\sum_{i}^{m} A_i^{\,p}}_{\beta}(X)\right|}\text{。}$$

例 1： S 是一个不完备的多粒度信息系统，(如表 1 所示)，$A_1=\{a_1\}$，$A_2=\{a_2\}$，$A_3=\{a_3,a_4\}$，$U=\{x_1,x_2,\cdots,x_8\}$，$x_1,x_2,\cdots,x_8$ 是 8 个对象，a_1,a_2,a_3,a_4 是四个条件属性，D 是决策性，$A=\{A_1,A_2,A_3\}$ 是粒度空间。

Table 1. Incomplete information system S
表 1. 不完备信息系统 S

U	a_1	a_2	a_3	a_4	D
x_1	1	1	3	2	1
x_2	2	*	2	1	1
x_3	1	1	3	2	1
x_4	1	*	*	1	2
x_5	*	2	2	*	2
x_6	1	2	*	2	2
x_7	2	2	2	1	2
x_8	1	2	1	1	1

$$U/d=\{\{x_1,x_2,x_3,x_8\},\{x_4,x_5,x_6,x_7\}\}$$

根据定义 4 基于相似关系，当 α 变小，β 变大时，求得在不完备多粒度信息系统中概率粗糙集在不同决策类型的乐观(悲观)多粒度上(下)近似集(如表 2 所示)：

Table 2. Optimistic (pessimistic) upper and lower approximate sets of incomplete multi-granularity probability rough sets based on similarity relations
表 2. 基于相似关系的不完备多粒度概率粗糙集的乐观(悲观)上下近似集

	$\underline{\sum_{i=1}^{3}A_i^o}_{\alpha}(D_1)$	$\overline{\sum_{i=1}^{3}A_i^o}_{\beta}(D_1)$	$\underline{\sum_{i=1}^{3}A_i^o}_{\alpha}(D_2)$	$\overline{\sum_{i=1}^{3}A_i^o}_{\beta}(D_2)$	$\underline{\sum_{i=1}^{3}A_i^p}_{\alpha}(D_1)$	$\overline{\sum_{i=1}^{3}A_i^p}_{\beta}(D_1)$	$\underline{\sum_{i=1}^{3}A_i^p}_{\alpha}(D_2)$	$\overline{\sum_{i=1}^{3}A_i^p}_{\beta}(D_2)$
$\alpha=0.8$ $\beta=0.1$	ϕ	ϕ	ϕ	ϕ	ϕ	U	ϕ	U
$\alpha=0.7$ $\beta=0.1$	$\{x_1,x_2,x_3,x_4\}$	ϕ	$\{x_2,x_4,x_5,x_7\}$	ϕ	$\{x_1,x_2,x_3,x_4\}$	U	$\{x_2,x_4,x_5,x_7\}$	U
$\alpha=0.6$ $\beta=0.4$	$\{x_1,x_2,x_3,x_4,x_6\}$	ϕ	$\{x_1,x_2,x_4,x_5,x_6,x_7,x_8\}$	ϕ	$\{x_1,x_3\}$	U	$\{x_2,x_5,x_7\}$	U

得到近似精度为：

$$\lambda^o(0.8,0.1)(D_1)=0;\ \lambda^o(0.8,0.1)(D_2)=0;\ \lambda^p(0.8,0.1)(D_1)=0;\ \lambda^p(0.8,0.1)(D_2)=0;$$

$$\lambda^o(0.7,0.1)(D_1)=0;\ \lambda^o(0.7,0.1)(D_2)=0;\ \lambda^p(0.7,0.1)(D_1)=\frac{1}{2};\ \lambda^p(0.7,0.1)(D_2)=\frac{1}{2};$$

$$\lambda^o(0.6,0.4)(D_1)=0;\ \lambda^o(0.6,0.4)(D_2)=0;\ \lambda^p(0.6,0.4)(D_1)=\frac{2}{8};\ \lambda^p(0.6,0.4)(D_2)=\frac{3}{8}$$

通过例 1，可以发现在不完备信息系统中，基于相似关系对缺省的值进行分类；会出现在乐观的情况下所有的上(下)近似和精度都为空的情况，并且在完备的信息系统中近似精度所满足的性质此时也都不在满足。由此，我们在接下来的文章中基于优劣关系进行讨论。

4. 基于优势(劣势)关系的多粒度概粗糙集模型

定义 7: 设 $MIS=\langle U,AT,V,f\rangle$ 为一个不完备的多粒度信息系统，$\forall x\in U$，$X\subseteq U$，$0\le\beta\le 0.5<\alpha\le 1$，则

(1) 在优势关系 $R_{A_i}^{\ge}$ 上,依参数 α,β 的乐观(悲观)多粒度概率粗糙集下近似 $\underline{\sum_{i=1}^{m}A_i^o}_{\alpha}(X)$ 和上近似 $\overline{\sum_{i=1}^{m}A_i^o}_{\beta}(X)$ 分别定义：

$$\underline{\sum_{i=1}^{m}A_i^o}_{\alpha}(X)=\left\{x\in U:p\left(X\mid[x]_{A_1}^{\ge}\right)\ge\alpha\vee P\left(X\mid[x]_{A_2}^{\ge}\right)\ge\alpha\vee\cdots\vee P\left(X\mid[x]_{A_m}^{\ge}\right)\ge\alpha\right\}$$

$$\overline{\sum_{i=1}^{m}A_i^o}_{\beta}(X)=\sim\underline{\sum_{i=1}^{m}A_i^o}_{1-\beta}(\sim X)$$

通过定义，可以得到在不完备系统中，乐观(悲观)多粒度概率粗糙集的边界域集合如下：

$$Bnd^o_{\sum_{i=1}^{m}A_{i(\alpha,\beta)}}=\overline{\sum_{i=1}^{m}A_i^o}_{\beta}(X)-\underline{\sum_{i=1}^{m}A_i^o}_{\alpha}(X)$$

二元组 $\left\langle\underline{\sum_{i=1}^{m}A_i^o}_{\alpha}(X),\overline{\sum_{i=1}^{m}A_i^o}_{\beta}(X)\right\rangle$ 是一个被称作是 X 基于优势关系 $R_{A_i}^{\ge}$ 关于粒度 $A_i\,(i=1,2,\cdots,m)$ 的乐观多粒度概率粗糙集。

(2) 在劣势关系 $R_{A_i}^{\le}$ 上，依参数 α,β 的乐观(悲观)多粒度概率粗糙集下近似 $\underline{\sum_{i=1}^{m}A_i^o}_{\alpha}(X)$ 和上近似 $\overline{\sum_{i=1}^{m}A_i^o}_{\beta}(X)$ 分别定义：

$$\underline{\sum_{i=1}^{m}A_i^o}_{\alpha}(X)=\left\{x\in U:p\left(X\mid[x]_{A_1}^{\le}\right)\ge\alpha\vee P\left(X\mid[x]_{A_2}^{\le}\right)\ge\alpha\vee\cdots\vee P\left(X\mid[x]_{A_m}^{\le}\right)\ge\alpha\right\}$$

$$\overline{\sum_{i=1}^{m}A_i^o}_{\beta}(X)=\sim\underline{\sum_{i=1}^{m}A_i^o}_{1-\beta}(\sim X)$$

通过定义，可以得到，在不完备系统中，乐观多粒度概率粗糙集的边界域集合：

$$Bnd^o_{\sum_{i=1}^{m}A_{i(\alpha,\beta)}}=\overline{\sum_{i=1}^{m}A_i^o}_{\beta}(X)-\underline{\sum_{i=1}^{m}A_i^o}_{\alpha}(X)$$

二元组 $\left\langle\underline{\sum_{i=1}^{m}A_i^o}_{\alpha}(X),\overline{\sum_{i=1}^{m}A_i^o}_{\beta}(X)\right\rangle$ 称作是 X 基于劣势关系关于粒度 $A_i\,(i=1,2,\cdots,m)$ 的乐观多粒度概率粗糙集。

定理 4.1: 设 $MIS=\langle U,AT,V,f\rangle$ 为一个不完备的多粒度信息系统，$X\subseteq U$，$\forall x\in U$，$0\le\beta\le 0.5<\alpha\le 1$，则上近似可以表示为

$$\overline{\sum_{i=1}^{m}A_i^0}=\left\{x\in U:p\left(X\mid[x]_{A_1}^{\ge}\right)>\beta\wedge P\left(X\mid[x]_{A_2}^{\ge}\right)>\beta\wedge\cdots\wedge P\left(X\mid[x]_{A_m}^{\ge}\right)>\beta\right\}$$

证明：$\overline{\sum_{i=1}^{m} A_i^o}_{\beta}(X) = \sim\underline{\sum_{i=1}^{m} A_i^o}_{1-\beta}(\sim X)$ (其余等式证明可类似定理 3.1 可证)。

注：(1) 同理可得，悲观的上近似有如下性质

$$\overline{\sum_{i=1}^{m} A_i^p}_{\beta} = \left\{x \in U : p\left(X \mid [x]_{A_1}^{\geq}\right) > \beta \vee P\left(X \mid [x]_{A_2}^{\geq}\right) > \beta \vee \cdots \vee P\left(X \mid [x]_{A_m}^{\geq}\right) > \beta\right\}。$$

(2) 同理基于劣势关系也可得出如上定理。

定义 8：设 $MIS = \langle U, AT, V, f\rangle$ 为不完备多粒度信息系统 $\forall x \in U$，$X \subseteq U$，$0 \leq \beta \leq 0.5 < \alpha \leq 1$，因此 X 在优势关系 $R_{A_i}^{\geq}$ (劣势关系 $R_{A_i}^{\leq}$)上依参数 α, β 的乐观(悲观)多粒度概率粗糙集近似精度 $\lambda^o(\alpha,\beta)$ 分别定义为：

$$\lambda^o(\alpha,\beta) = \frac{\left|\underline{\sum_{i}^{m} A_i^o}_{\alpha}(X)\right|}{\left|\overline{\sum_{i}^{m} A_i^o}_{\beta}(X)\right|}；\quad \lambda^p(\alpha,\beta) = \frac{\left|\underline{\sum_{i}^{m} A_i^p}_{\alpha}(X)\right|}{\left|\overline{\sum_{i}^{m} A_i^p}_{\beta}(X)\right|}。$$

例 2 不完备多粒度信息系统为例 1 中给出的(如表 1)我们基于优势关系进行讨论，得到在优势关系下的归类：$U / R_{A_i}^{\geq} = \left\{[x_1]_{A_i}^{\geq}, [x_2]_{A_i}^{\geq}, \cdots, [x_8]_{A_i}^{\geq}\right\}$

根据定义 7，当 α 变小，β 变大时，基于优势关系求得在不完备多粒度信息系统中概率粗糙集在不同决策类的乐观(悲观)多粒度上(下)近似集(如表 3 所示)：

Table 3. Optimistic (pessimistic) upper and lower approximations of incomplete multi-granularity probability rough sets under dominance relations

表 3. 优势关系下的不完备多粒度概率粗糙集的乐观(悲观)上下近似

	$\underline{\sum_{i=1}^{3} A_i^o}_{\alpha}(D_1)$	$\overline{\sum_{i=1}^{3} A_i^o}_{\beta}(D_1)$	$\underline{\sum_{i=1}^{3} A_i^o}_{\alpha}(D_2)$	$\overline{\sum_{i=1}^{3} A_i^o}_{\beta}(D_2)$	$\underline{\sum_{i=1}^{3} A_i^p}_{\alpha}(D_1)$	$\overline{\sum_{i=1}^{3} A_i^p}_{\beta}(D_1)$	$\underline{\sum_{i=1}^{3} A_i^p}_{\alpha}(D_2)$	$\overline{\sum_{i=1}^{3} A_i^p}_{\beta}(D_2)$
$\alpha=0.8$ $\beta=0.1$	ϕ	ϕ	ϕ	ϕ	ϕ	U	ϕ	U
$\alpha=0.7$ $\beta=0.1$	ϕ	ϕ	ϕ	ϕ	ϕ	U	ϕ	U
$\alpha=0.6$ $\beta=0.4$	$\{x_1,x_3,x_6\}$	$\{x_1,x_3,x_6\}$	$\{x_2,x_4,x_5,x_6,x_7,x_8\}$	$\{x_2,x_4,x_5,x_7\}$	$\{x_1,x_3,x_6\}$	U	$\{x_2,x_4,x_5,x_6,x_7,x_8\}$	U

得到近似精度为：

$$\lambda^o(0.8,0.1)(D_1)=0;\ \lambda^o(0.8,0.1)(D_2)=0;\ \lambda^p(0.8,0.1)(D_1)=0;\ \lambda^p(0.8,0.1)(D_2)=0;$$

$$\lambda^o(0.7,0.1)(D_1)=0;\ \lambda^o(0.7,0.1)(D_2)=0;\ \lambda^p(0.7,0.1)(D_1)=0;\ \lambda^p(0.7,0.1)(D_2)=0;$$

$$\lambda^o(0.6,0.4)(D_1)=1;\ \lambda^o(0.6,0.4)(D_2)=\frac{3}{2};\ \lambda^p(0.6,0.4)(D_1)=\frac{3}{8};\ \lambda^p(0.6,0.4)(D_2)=\frac{6}{8}$$

根据定义 7，当 α 变小，β 变大时，基于劣势关系求得在不完备多粒度信息系统中概率粗糙集在不同决策类的乐观(悲观)多粒度上(下)近似集(如表 4 所示)：

Table 4. Optimistic (pessimistic) upper and lower approximations of incomplete multi-granularity probability rough sets under the inferiority relationship

表 4. 劣势关系下的不完备多粒度概率粗糙集的乐观(悲观)上下近似

	$\underline{\sum_{i=1}^{3} A_i^o}_{\alpha}(D_1)$	$\overline{\sum_{i=1}^{3} A_i}_{\beta}^{o}(D_1)$	$\underline{\sum_{i=1}^{3} A_i}_{\alpha}^{o}(D_2)$	$\overline{\sum_{i=1}^{3} A_i}_{\beta}^{o}(D_2)$	$\underline{\sum_{i=1}^{3} A_i^p}_{\alpha}(D_1)$	$\overline{\sum_{i=1}^{3} A_i}_{\beta}^{p}(D_1)$	$\underline{\sum_{i=1}^{3} A_i}_{\alpha}^{p}(D_2)$	$\overline{\sum_{i=1}^{3} A_i}_{\beta}^{p}(D_2)$
$\alpha=0.8$ $\beta=0.1$	ϕ	$\{x_4\}$	ϕ	$\{x_4\}$	ϕ	U	ϕ	U
$\alpha=0.7$ $\beta=0.1$	$\{x_1,x_2,x_3,x_4\}$	$\{x_4\}$	ϕ	$\{x_4\}$	$\{x_1,x_2,x_3,x_4\}$	U	ϕ	U
$\alpha=0.6$ $\beta=0.4$	$\{x_1,x_2,x_3,x_4\}$	$\{x_4\}$	$\{x_2,x_4,x_5,x_7,x_8\}$	$\{x_4,x_8\}$	$\{x_1,x_2,x_3,x_4\}$	U	$\{x_2,x_4,x_5,x_7,x_8\}$	U

基于劣势关系得到近似精度为：

$$\lambda^o(0.8,0.1)(D_1)=0;\ \lambda^o(0.8,0.1)(D_2)=0;\ \lambda^p(0.8,0.1)(D_1)=0;\ \lambda^p(0.8,0.1)(D_2)=0;$$

$$\lambda^o(0.7,0.1)(D_1)=\frac{4}{1};\ \lambda^o(0.7,0.1)(D_2)=0;\ \lambda^p(0.7,0.1)(D_1)=\frac{1}{2};\ \lambda^p(0.7,0.1)(D_2)=0;$$

$$\lambda^o(0.6,0.4)(D_1)=\frac{4}{1};\ \lambda^o(0.6,0.4)(D_2)=\frac{5}{2};\ \lambda^p(0.6,0.4)(D_1)=\frac{1}{2};\ \lambda^p(0.6,0.4)(D_2)=\frac{5}{8}$$

定理 4.2:设 $MIS=\langle U,AT,V,f\rangle$ 为不完备多粒度信息系统，$\forall x\in U$，$X\subseteq U$，$0\le\beta\le0.5<\alpha\le1$，因此基于优势关系和劣势关系的 α,β 的近似精度应有如下性质：

(1) $\lambda^p(\alpha,\beta)\le\lambda^0(\alpha,\beta)$

(2) $0\le\lambda^p(\alpha,\beta)\le1,\lambda^o(\alpha,\beta)\ge0$

注：在相似关系中上述性质一般不成立。

通过例 1 和例 2 的比较，可以发现在不完备系统中，基于优劣关系对缺省的值进行分类；这样可以避免依据相似关系分类时造成的信息偏差，和在乐观情况下所有的上(下)近似和精度都为空的情况。根据乐观和悲观，可以在不完备的多粒度概率粗糙集中通过调整 α,β 值，得到不同的近似精度，并有如下变化规律：

(1) 在相同的 α,β 值下，基于劣势关系得到了更大的近似精度。

(2) 当 α,β 值同时变小时，上下近似会同时变增大；

(3) 当 α,β 值同时变时，上下近似会同时变递减。

(4) 当 α 逐渐减小，β 逐渐增大时，可以得到较大的下近似和较小的上近似，并且相应的不完备多粒度概率粗糙集的近似精度会增大。相反，则会减小。

5. 基于优劣势关系的多粒度概率粗糙集的决策规则

定义 9 [8]：$MIS=(U,AT,F,d)$ 是一个不完备多粒度决策信息系统，d 是决策属性，d 把 U 分类为有序的：$D=\{D_t,t\in T\},T=\{1,\cdots,k\}$，对于任意 $r,s\in T$，当 $r>s$ 时，C_r 中的对象优于 C_s 中的对象，于是这些有序的决策类用向上联合与向下联合近似来表示：$D_t^{\ge}=\bigcup D_s,D_t^{\le}=\bigcup_{s\le t}D_s\ (t=1,\cdots,k)$，当 $x\in D_t^{\ge}$ 时，表示 x 至于属于 D_t 决策类；而当 $x\in D_t^{\le}$ 时，表示 x 至多属于 $D_t^{\le t}$ 决策类。

定义 10：设 $MIS=\langle U,AT,V,f\rangle$ 为一个不完备多粒度信统，$\forall x\in U$，$X\subseteq U$，$AT=\{A_1,A_2,\cdots,A_m\}$，$0\le\beta\le0.5<\alpha\le1$，$D_t^{\ge}$ 和 $D_t^{\le}$ 在优势关系 $R_A^{\ge}$ 上的下近似与上近似表示：

$$\underline{\sum_{i=1}^{m} R_A^{\geq}(X)}\left(D_t^{\geq}\right)$$
$$=\left\{x\in U \mid P\left(X \mid [x]_{A_1}^{\geq}\right)\geq\gamma\subseteq D_t^{\geq}\wedge[x]_{A_2}^{\geq}\geq\gamma\subseteq D_t^{\geq}\wedge\cdots\wedge[x]_{A_m}^{\geq}\geq\gamma\subseteq D_t^{\geq}\right\},(t=1,\cdots,k)$$

$$\overline{\sum_{i=1}^{m} R_A^{\geq}(X)}\left(D_t^{\geq}\right)=\bigcup_{x\in C_t^{\geq}}\left(P\left(X \mid [x]_{A_i}^{\geq}\right)\geq\gamma\right),(t=1,\cdots,k)$$

$$\underline{\sum_{i=1}^{m} R_A^{\geq}(X)}\left(D_t^{\leq}\right)$$
$$=\left\{x\in U \mid P\left(X \mid [x]_{A_i}^{\leq}\right)\geq\gamma\subseteq D_t^{\leq}\wedge P\left(X \mid [x]_{A_2}^{\leq}\right)\geq\gamma\subseteq D_t^{\leq}\wedge\cdots\wedge P\left(X \mid [x]_{A_m}^{\leq}\right)\geq\gamma\subseteq D_t^{\leq}\right\},(t=1,\cdots,k)$$

$$\overline{\sum_{i=1}^{m} R_A^{\geq}(X)}\left(D_t^{\leq}\right)=\bigcup_{x\in C_t^{\leq}}\left(P\left(X \mid [x]_{A_i}^{\leq}\right)\geq\gamma\right),(t=1,\cdots,k)$$

$D_t^{\geq}$ 和 $D_t^{\leq}$ 的边界为：

$$BN_A\left(D_t^{\geq}\right)=\overline{\sum_{i=1}^{m} R_A^{\geq}(X)}\left(D_t^{\geq}\right)-\underline{\sum_{i=1}^{m} R_A^{\geq}(X)}\left(D_t^{\geq}\right)$$

$$BN_A\left(D_t^{\leq}\right)=\overline{\sum_{i=1}^{m} R_A^{\geq}(X)}\left(D_t^{\leq}\right)-\underline{\sum_{i=1}^{m} R_A^{\geq}(X)}\left(D_t^{\leq}\right)$$

则可能的决策规则有：

(1) 确定的 $D_{\geq}$ 规则：若 $f_{a1}(x)\geq v_{a1}$ 与 $f_{a2}(x)\geq v_{a2}$ 与 $f_{ak}(x)\geq v_{ak}$ 与 $f_{ak+1}(x)\leq v_{ak+1}$ 与 $f_{ap}(x)\leq v_{ap}$，则 $x\in D_{\geq}$。

(2) 可能的 $D_{\geq}$ 规则：若 $f_{a1}(x)\geq v_{a1}$ 与 $f_{a2}(x)\geq v_{a2}$ 与 $f_{ak}(x)\geq v_{ak}$ 与 $f_{ak+1}(x)\leq v_{ak+1}$ 与 $f_{ap}(x)\leq v_{ap}$，则 x 可能属于 $D_t^{\geq}$。

(3) 确定的 $D_{\leq}$ 规则：若 $f_{a1}(x)\leq v_{a1}$ 与 $f_{a2}(x)\leq v_{a2}$ 与 $f_{ak}(x)\leq v_{ak}$ 与 $f_{ak+1}(x)\geq v_{ak+1}$ 与 $f_{ak+1}(x)\geq v_{ak+1}$，则 $x\in D_t^{\leq}$。

(4) 可能的 $D_{\leq}$ 规则：若 $f_{a1}(x)\leq v_{a1}$ 与 $f_{a2}(x)\leq v_{a2}$ 与 $f_{ak}(x)\leq v_{ak}$ 与 $f_{ak}(x)\leq v_{ak}$ 与 $f_{ap}(x)\geq v_{ap}$，则 x 可能属于 $D_t^{\leq}$。

例 3： s 是一个不完备的多粒度决策信息系统(如下表 5 所示)，$U=\{x_1,x_2,\cdots,x_8\}$，$x_1,x_2,\cdots,x_8$ 是 8 个对象，a_1,a_2,a_3,a_4 是四个条件属性，是决策属性，$A=\{A_1,A_2,A_3\}$ 是粒度间，$A_1=\{a_1\}$，$A_2=\{a_2\}$，$A_3=\{a_3,a_4\}$。

Table 5. Incomplete information system
表 5. 不完备信息系统

U	a_1	a_2	a_3	a_4	D
x_1	1	1	3	2	2
x_2	2	*	2	1	2
x_3	1	1	3	2	2
x_4	1	*	*	1	1
x_5	*	2	2	*	2

Continued

x_6	1	2	*	2	2
x_7	2	2	2	1	2
x_8	1	2	1	1	1

从表中得到决策类 $D=\{D_1,D_2\}$，其中：$D_1=\{x_1,x_2,x_3,x_5,x_6,x_8\}$，$D_2=\{x_4,x_7\}$。

由于只有两个决策类，得到向下联合 $D_1^{\le}=D_1$ 与向上联合 $D_2^{\ge}=D_2$，于是当 $X=\{x_1,x_2,x_3,x_5,x_6,x_8\}$，取 $\gamma=1$ 时，得到下近似，上近似和边界为

$$\overline{\sum_{i=1}^{3}R_A^{\ge}(X)}\left(D_1^{\ge}\right)=\phi,\quad \overline{\sum_{i=1}^{3}R_A^{\ge}(X)}\left(D_1^{\le}\right)=\phi$$

$$\underline{\sum_{i=1}^{3}R_A^{\ge}(X)}\left(D_2^{\ge}\right)=\{x_1,x_3,x_6\},\quad \underline{\sum_{i=1}^{3}R_A^{\ge}(X)}\left(D_2^{\le}\right)=\{x_1,x_3,x_6\}$$

$$BN_A\left(D_1^{\le}\right)=\phi,\quad BN_A\left(D_2^{\ge}\right)=\phi$$

取 $\gamma=0.7$ 时，得到下近似，上近似和边界为

$$\overline{\sum_{i=1}^{3}R_A^{\ge}(X)}\left(D_1^{\ge}\right)=\phi,\quad \overline{\sum_{i=1}^{3}R_A^{\ge}(X)}\left(D_1^{\le}\right)=\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8\}$$

$$\underline{\sum_{i=1}^{3}R_A^{\ge}(X)}\left(D_2^{\ge}\right)=\{x_1,x_3,x_6\},\quad \underline{\sum_{i=1}^{3}R_A^{\ge}(X)}\left(D_2^{\le}\right)=\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8\}$$

$$BN_A\left(D_1^{\le}\right)=\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8\},\quad BN_A\left(D_2^{\ge}\right)=\{x_2,x_4,x_5,x_7,x_8\}$$

因此，可以将不完备信息中的多粒度层整体考虑来得到规则：

$r_1:(a_1,\ge,1)\wedge(a_2,\ge,1)\wedge(a_3,\ge,3)\wedge(a_4,\ge,1)\to(d,\ge,2)$，支持这个决策规则的对象有 x_1,x_3；

$r_2:(a_1,\ge,2)\wedge(a_2,\ge,*)\wedge(a_3,\ge,2)\wedge(a_4,\ge,1)\to(d,\ge,2)$，支持这个决策规则的对象有 x_2,x_5,x_7；

$r_3:(a_1,\ge,1)\wedge(a_2,\ge,*)\wedge(a_3,\ge,*)\wedge(a_4,\ge,1)\to(d,\ge,1)$，支持这个决策规则的对象有 $x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8$；

$r_4:(a_1,\ge,*)\wedge(a_2,\ge,2)\wedge(a_3,\ge,2)\wedge(a_4,\ge,*)\to(d,\ge,2)$，支持这个决策规则的对象有 x_5,x_7；

$r_5:(a_1,\ge,1)\wedge(a_2,\ge,2)\wedge(a_3,\ge,*)\wedge(a_4,\ge,2)\to(d,\ge,2)$，支持这个决策规则的对象有 x_5,x_6；

$r_6:(a_1,\ge,1)\wedge(a_2,\ge,2)\wedge(a_3,\ge,1)\wedge(a_4,\ge,1)\to(d,\ge,1)$，支持这个决策规则的对象有 x_2,x_5,x_6,x_7,x_8。

当 $\gamma=0.7$，规则 $r_1,r_2,\cdots,r_8$ 全部可能是确定的 $D_{\ge}$ 决策规则，也有可能是 $D_{\le}$ 决策规则。

当 $\gamma=1$ 时，规则 r_1 时确定的 $D_{\ge}$ 决策规则，$r_2,\cdots,r_6$，可能是 $D_{\le}$ 决策规则，但是也有可能是确定的 $D_{\ge}$ 决策规则。

由此，可以发现当 γ 值越大，则得到的确定性规则越多，反之会越少。这样我们就可以根据决策者的不同要求来获取更为有用的规则。

6. 结论及展望

本文是基于不完备信息系统进行的研究。在本文开始，介绍了不完备系统中的多粒度粗糙集理论和乐观粗糙集近似和悲观多粒度粗糙集近似的定义；第三部分给出了不完备信息系统中基于相似关系的多

粒度概率粗糙集的相关定义及性质，讨论了其上下近似和近似精度。第四部分基于优势关系劣势关系分别讨论了多粒度概率粗糙集的上下近似和近似精度，并做了对比，通过调整参数α,β的值来得到更高的近似精度，讨论了近似精度满足的性质和变化规律。最后在第五部分提出了不完备信息系统中的多粒度概率粗糙集基于优劣关系的规则提取。

每个相关的定义都给出了具体实例验证，进一步探讨概率粗糙集模型基于优劣关系的规则提取的有效性，接下来的工作将针对这些规则的给出，尝试给出更为高效的约简及其规则提取的算法。

基金项目

青海省自然科学基金(2018-ZJ-911)；青海民族大学科研创新团队专项经费支助。

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] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001-2007, 36(6): 1237-1240.

[3] 李金海, 吕跃进. 决策系统的快速属性约简算法[J]. 电子科技大学学报, 2007, 36(6): 1237-1240.

[4] 覃丽珍, 姚炳学, 李金海. 基于信息量的完备覆盖约简算法[J]. 计算机科学, 2012, 39(10): 235-239.

[5] 陈青梅. 多粒度概率粗糙集若干问题研究[D]: [硕士学位论文]. 南宁: 广西大学, 2014.

[6] 胡善忠, 徐怡, 何明慧, 王冉. 不完备系统中的一种多粒度粗糙集粒度约简算法[J]. 微电子学与计算机, 2018, 35(6): 41-46.

[7] 桑妍丽, 钱宇华. 多粒度决策粗糙集中的粒度约简方法[J]. 计算机科学, 2017, 44(5): 199-205.

[8] 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.

[9] 胡善忠. 多粒度粗糙集的粒度约简算法研究[D]: [硕士学位论文]. 合肥: 安徽大学, 2018.

[10] 许韦, 吴陈, 杨习贝. 基于容差关系的不完备可变精度多粒度粗糙集[J]. 计算机应用研究, 2013, 30(6): 1712-1715.

[11] 翟永健. 不完备信息系统中多粒度粗糙集理论与约简研究[D]: [博士学位论文]. 南京: 南京理工大学, 2014.