DOI:10.11830/ISSN.1000-5013.201812041

粗糙集中几种粒结构的代数关系

罗来鹏, 范自柱

(华东交通大学 理学院, 江西 南昌 330013)

摘要: 针对目标概念在近似空间上存在多种粒结构的问题,通过讨论目标概念的最优近似集与 Pawlak 近似集、变精度近似集之间的代数关系,得到最优近似集与 Pawlak 下、上近似集、变精度下、上近似集的等价条件;通过分析基于最优近似、基于 Pawlak 近似、基于变精度近似的分布约简之间的关系,得到在一定条件下,最优近似分布约简为 Pawlak 近似与变精度近似的分布约简. 研究结果表明:根据目标概念与基本知识粒之间不同的近似刻画,不仅可以建立不同的粗糙集模型,还可以建立不同的分布约简.

关键词: 粒计算; 粗糙集; 属性约简; 包含度; 变精度粗糙集; 最优近似; 相似度

中图分类号: TP 18　　**文献标志码:** A　　**文章编号:** 1000-5013(2019)05-0694-07

Algebraic Relation of Several Granular Structures in Rough Sets

LUO Laipeng, FAN Zizhu

(School of Sciences, East China Jiaotong University, Nanchang 330013, China)

Abstract: To the problem of existing many kinds of granular structures on approximation space, the algebraic relations among the optimal approximation set, Pawlak lower and upper approximation set and the variable precision lower and upper approximation set of the target concept are discussed. The equivalent conditions of the optimal approximation set between Pawlak approximation set, the variable precision approximation set are obtained. The relationships among the distribution reduction based on the optimal approximation, the distribution reduction based on the Pawlak approximation and the distribution reduction based on the variable precision approximation are analyzed. Under certain conditions, the conclusions that the optimal approximate distribution reduction is Pawlak approximation and variable precision approximation distribution reduction are presented. It can be seen from these above conclusions that the different rough set models can not only be established, but also different distribution reduction can be established according to the different approximate characterization between the target concept and the basic granular.

Keywords: granular computing; rough set; attribute reduction; inclusion degree; variable precision rough set; optimal approximation; similarity degree

粒计算是人工智能领域中的一种认知理念和计算方法,它主要涉及两个基本问题:信息粒化和基于信息粒的计算. 目前,主要有如模糊集、粗糙集、商空间等粒计算模型[1-2]. 在粗糙集的粒计算模型中[3],论域上的等价关系决定了论域的划分. 利用基本知识粒表示目标概念(论域的子集)是粗糙集研究的基础. 近年来,基于粗糙集的粒计算研究引起国内外学者的广泛关注,并已成功应用于人工智能、机器学习等领域[4-6]. 文献[7-8]提出目标概念的动态表示与计算方法. Chen 等[9]将粒计算理论与代数理论相结合,提出基于代数结构的粒计算模型. Hu 等[10]讨论了逻辑运算符的粒度等价性. 王宝丽等[11]提出基于

收稿日期: 2018-12-19

通信作者: 罗来鹏(1973-),男,副教授,主要从事粗糙集与粒计算等研究. E-mail:luolp789@163.com.

基金项目: 国家自然科学基金资助项目(61472138)

粒计算的犹豫模糊多准则决策方法.吴伟志等[12]针对具有多粒度标记的不完备信息系统的知识,提出最优粒度特征.吴伟志[13]介绍目前流行的 3 类基于粗糙集的多粒度数据处理模型,并提出若干研究问题.在粗糙集的粒计算研究中,处理边界的不确定性是粗糙集研究关注的难点.最具有代表性的处理方法是通过引入一个误差因子,重新定义目标概念,如较典型的模型有变精度粗糙集模型[14]和概率粗糙集模型[15-16].在一个近似空间上,目标概念的最优近似的表示与计算一直是个未得到考虑的问题.文献[17-19]从一种新视角定义和表示粗糙集的近似集,通过引入集合相似度的评价标准,发现在目标概念的下、上近似集之间存在可定义集合,它与目标概念的相似度大于目标概念与其下、上近似集之间的相似度,并由此提出粗糙集的最优近似概念.在一个给定的近似空间,最优近似与 Pawlak 下、上近似、变精度下、上近似之间,以及由此建立的属性约简之间存在的关系是一个有待研究的问题.如谭安辉等[20]从近似集和属性约简两个角度,探讨完备信息系统与不完备信息系统中多粒度粗糙集和覆盖粗糙集的关系,得到多粒度粗糙集约简与覆盖粗糙集约简间的密切联系.基于此,本文就上述 3 种近似之间关系展开讨论分析,并在此基础上,进一步讨论这三者所建立的属性约简间的关系.

1　基本概念

定义 1[3]　设(U,R)为近似空间,U 为有限论域,R 为等价关系,由 R 决定 U 上一个划分,即为系统的基本知识粒,表示为 $U/R=\{X_1,X_2,\cdots,X_m\}$.对于任意 $X\subseteq U$,X 在近似空间的下、上近似集、边界域的定义分别为 $\underline{R}(X)=\bigcup_i X_i$,$X_i\subseteq X$ 且 $X_i\in U/R$;$\bar{R}(X)=\bigcup_i X_i$,其中,$X_i\cap X\neq\Phi$,Φ 为空集,$X_i\in U/R$;$B_n(X)=\bar{R}(X)-\underline{R}(X)$.

定义 2[15]　设(U,R)为近似空间,U 为有限论域,R 为等价关系,由 R 决定 U 上一个划分,表示为 $U/R=\{X_1,X_2,\cdots,X_m\}$.对于任意 $\lambda\in(0.5,1.0]$,$X\subseteq U$ 在近似空间上关于 λ 的下近似集、上近似集、边界域的定义分别为 $\underline{R}^{\lambda}(X)=\bigcup_i X_i$,$D(X/X_i)\geqslant\lambda$,且 $X_i\in U/R$;$\bar{R}^{\lambda}(X)=\bigcup_i X_i$,其中,$D(X/X_i)>1-\lambda$ 且 $X_i\in U/R$;$B_n^{\lambda}(X)=\bar{R}^{\lambda}(X)-\underline{R}^{\lambda}(X)$,其中,$D$ 为 U 上的包含度.对于 $X,Y\subseteq U$,通常取 $D(X/Y)=\dfrac{|X\cap Y|}{|Y|}$,$|\cdot|$为集合基数.

定义 3　设(U,R)为近似空间,U 为有限论域,R 为等价关系,由 R 决定 U 上一个划分,可表示为 $U/R=\{X_1,X_2,\cdots,X_m\}$.对于任意 $X\subseteq U$,若有集合 $X^{\mathrm{opt}}\subseteq U/R$,使得对任意 $Y\subseteq U/R$,有 $S(X,X^{\mathrm{opt}})\geqslant S(X,Y)$,那么,$X^{\mathrm{opt}}$称为 X 在近似空间上的最优近似集.其中,S 为 U 上的集合相似度[21].对于 $X,Y\subseteq U$,通常取 $S(X,Y)=\dfrac{|X\cap Y|}{|X\cup Y|}$.

定义 4[3]　设(U,R)为近似空间,U 为有限论域,R 为等价关系,由 R 决定的基本知识粒可表示为 $U/R=\{X_1,X_2,\cdots,X_m\}$.$F=\{Y_1,Y_2,\cdots,Y_n\}$为 U 上的另外一个划分,则 F 在(U,R)上的分类精度与分类质量分别定义为 $\alpha_R(F)=\dfrac{\sum_{i=1}^{n}|\underline{R}(Y_i)|}{\sum_{i=1}^{n}|\bar{R}(Y_i)|}$,$\gamma_R(F)=\dfrac{\sum_{i=1}^{n}|\underline{R}(Y_i)|}{|U|}$.

同理可知,对于 $\lambda\in(0.5,1.0]$,变精度情况下的分类精度与分类质量分别定义为 $\alpha_R^{\lambda}(F)=\dfrac{\sum_{i=1}^{n}|\underline{R}^{\lambda}(Y_i)|}{\sum_{i=1}^{n}|\bar{R}^{\lambda}(Y_i)|}$,$\gamma_R^{\lambda}(F)=\dfrac{\sum_{i=1}^{n}|\underline{R}^{\lambda}(Y_i)|}{|U|}$.

2　几种近似集之间关系

2.1　Pawlak 粗糙近似与变精度近似

设(U,R)为近似空间,$U/R=\{X_1,X_2,\cdots,X_m\}$,对于任意 $X\subseteq U$,$\lambda,\lambda_1,\lambda_2\in(0.5,1.0]$,有以下 10 点结论.

1) $\underline{R}(X)\subseteq\underline{R}^{\lambda}(X)$.

2) $\overline{R}^{\lambda}(X)\subseteq\overline{R}(X)$.

3) $B_n^{\lambda}(X)\subseteq B_n(X)$.

4) 若 $\lambda_1<\lambda_2$,则 $\underline{R}^{\lambda_2}(X)\subseteq\underline{R}^{\lambda_1}(X)$.

5) 若 $\lambda_1<\lambda_2$,则 $\overline{R}^{\lambda_1}(X)\subseteq\overline{R}^{\lambda_2}(X)$.

6) 若 $\lambda_1<\lambda_2$,则 $[\underline{R}^{\lambda_2}(X)\subseteq\underline{R}(X)]\subseteq[\underline{R}^{\lambda_1}(X)-\underline{R}(X)]$.

7) 若 $\lambda_1<\lambda_2$,则 $[\overline{R}^{\lambda_1}(X)\subseteq\underline{R}(X)]\subseteq[\overline{R}^{\lambda_2}(X)-\underline{R}(X)]$.

8) 若 $\lambda_1<\lambda_2$,则 $[\overline{R}^{\lambda_1}(X)-\underline{R}^{\lambda_1}(X)]\subseteq[\overline{R}^{\lambda_2}(X)-\underline{R}^{\lambda_2}(X)]$.

9) 分类精度 $\alpha^{\lambda}(F)\geqslant\alpha_{_R}(F)$.

10) 分类质量 $\gamma_R^{\lambda}(F)\geqslant\gamma_R(F)$.

这里只证明结论 4),5).

证明:4) 若 $x\in\underline{R}^{\lambda_2}(X)$,则 $D(X/[x]_R)\geqslant\lambda_2>\lambda_1$. 所以,$x\in\underline{R}^{\lambda_1}(X)$,即 $\underline{R}^{\lambda_2}(X)\subseteq\underline{R}^{\lambda_1}(X)$.

5) 若 $x\in\overline{R}^{\lambda_1}(X)$,则 $D(X/[x]_R)>1-\lambda_1>1-\lambda_2$. 所以,$x\in\overline{R}^{\lambda_2}(X)$,即 $\overline{R}^{\lambda_1}(X)\subseteq\overline{R}^{\lambda_2}(X)$.

根据上述结论可知:在相同近似空间上,变精度粗糙集扩大了正域与负域,从而压缩边界域,即在允许一定误差下,边界的不确定性变小.

2.2 Pawlak 粗糙近似与最优近似

设 (U,R) 为近似空间,$U/R=\{X_1,X_2,\cdots,X_m\}$,对于任意 $X\subseteq U$,令 $B_n^{\delta}(X)=\left\{X_i\,\middle|\,\dfrac{|X_i\cap X|}{|X_i-X|}\geqslant\dfrac{|\underline{R}(X)|}{|X|},X_i\subseteq B_n(X)\text{且 }X_i\in U/R\right\}$,则有以下 8 个性质.

1) 若 $\Phi\subseteq B_n^{\delta}(X)\subseteq B_n(X)$,则 $X^{\mathrm{opt}}=\underline{R}(X)\cup B_n^{\delta}(X)$.

2) 若 $B_n^{\delta}(X)=\Phi$,则 $X^{\mathrm{opt}}=\underline{R}(X)$.

3) 若 $B_n^{\delta}(X)=B_n(X)=\overline{R}(X)-\underline{R}(X)$,则 $X^{\mathrm{opt}}=\overline{R}(X)$.

4) $\underline{R}(X)\subseteq X^{\mathrm{opt}}\subseteq\overline{R}(X)$.

5) $\Phi^{\mathrm{opt}}=\Phi$,$U^{\mathrm{opt}}=U$.

6) $\underline{R}(X)^{\mathrm{opt}}=\underline{R}(X)$,$\overline{R}(X)^{\mathrm{opt}}=\overline{R}(X)$.

7) $\underline{R}(X)\cap\underline{R}(Y)\subseteq(X\cap Y)^{\mathrm{opt}}\subseteq\overline{R}(X)\cap\overline{R}(Y)$.

8) $\underline{R}(X)\cup\underline{R}(Y)\subseteq(X\cup Y)^{\mathrm{opt}}\subseteq\overline{R}(X)\cup\overline{R}(Y)$.

证明:1) 假设 $B_n^{\delta}=\bigcup\limits_{i=1}^{k}[y_i]_R$,且有 $\dfrac{|[y_1]_R\cap X|}{|[y_1]_R-X|}\leqslant\cdots\leqslant\dfrac{|[y_k]_R\cap X|}{|[y]_k-X|}$,根据不等式的性质,可知

$S(\underline{R}(X),X)\leqslant\dfrac{|\underline{R}(X)|+|[y_1]_R\cap X|}{|X|+|[y_1]_R-X|}\leqslant\cdots\leqslant\dfrac{|\underline{R}(X)|+\sum\limits_{i=1}^{k}|[y_i]_R\cap X|}{|X|+\sum\limits_{i=1}^{k}|[y_i]_R-X|}$. 于是,$S(X,\underline{R}(X))\leqslant S(X,\underline{R}(X)\cup[y_1]_R)\leqslant\cdots\leqslant S(X,\underline{R}(X)\cup[y_1]_R\cdots\cup[y_k]_R)=S(X,B_n^{\delta}(X)\cup\underline{R}(X))$,所以,$X$ 最优近似集为 $X^{\mathrm{opt}}=\underline{R}(X)\cup B_n^{\delta}(X)$.

2) $B_n^{\delta}(X)=\Phi$,对于任意 $X_i\in B_n(X)$,有 $\dfrac{|X_i\cap X|}{|X_i-X|}<\dfrac{|\underline{R}|(X)}{|X|}$,则 $S(X,\underline{R}(X))=\dfrac{|\underline{R}(X)|}{|X|}>\dfrac{|\underline{R}(X)|+|X_i\cap X|}{|X|+|X_i-X|}=\dfrac{|(\underline{R}(X)\cup X_i)\cap X|}{|(\underline{R}(X)\cup X_i)\cup X|}$,即 $S(X,\underline{R}(X))>S(X,\underline{R}(X)\cup X_i)$.

3) $B_n^{\delta}(X)=B_n(X)=\overline{R}(X)-\underline{R}(X)$,由性质 1)得 $X^{\mathrm{opt}}=\underline{R}(X)\cup B_n^{\delta}(X)=\underline{R}(X)\cup B_n(X)=\overline{R}(X)$.

关于性质 1)~3)的证明,可参考文献[22],性质 4)~6)也成立.

7) 由于 $\underline{R}(X\cap Y)\subseteq(X\cap Y)^{\mathrm{opt}}\subseteq\overline{R}(X\cap Y)$,又有 $\overline{R}(X\cap Y)\subseteq\overline{R}(X)\cap\overline{R}(Y)$,$\underline{R}(X\cap Y)\supseteq\underline{R}(X)\cap\underline{R}(Y)$. 故 $\underline{R}(X)\cap\underline{R}(Y)\subseteq(X\cap Y)^{\mathrm{opt}}\subseteq\overline{R}(X)\cap\overline{R}(Y)$.

8) 证明与性质 7)的证明类似.

综上所述,变精度粗糙集近似是从基本知识粒出发,如果一个基本知识粒的大多数(超过 50%)被

划入目标概念，那么，该基本知识粒可以当作目标概念的一部分下近似集；最优近似是从由(U,R)构成的系统空间出发，如果一个基本知识粒加入到目标概念的正域中，使得其与目标概念的相似度大于目标概念与正域的相似度，那么，该基本知识粒可看成目标概念最优近似组成部分. 上述两种近似从不同角度阐述了目标概念在近似空间上的近似性问题.

2.3　最优近似与变精度近似

设(U,R)为近似空间，$U/R=\{X_1,X_2,\cdots,X_m\}$，$X\subseteq U$，可以得到以下 4 个性质.

1) $\underline{R}^{\lambda}(X)\subseteq X^{\mathrm{opt}}$.

2) 若 $\lambda<\dfrac{|X|}{|X|+|\underline{R}(X)|}$，则 $\overline{R}^{\lambda}(X)\subseteq X^{\mathrm{opt}}$.

3) 若 $\lambda<\dfrac{|\underline{R}(X)|}{|X|+|\underline{R}(X)|}$，则 $\underline{R}^{\lambda}(X)=X^{\mathrm{opt}}$.

4) 若 X 是近似空间(U,R)上的可定义集，则 $\underline{R}(X)=\underline{R}^{\lambda}(X)=X^{\mathrm{opt}}$，$\overline{R}(X)=\overline{R}^{\lambda}(X)=X^{\mathrm{opt}}$.

证明：1) 由于 $\underline{R}(X)\subseteq\underline{R}^{\lambda}(X)$，$\underline{R}(X)\subseteq X^{\mathrm{opt}}$，又对于 $\forall X_i\in B_n(X)$，且 $X_i\in\underline{R}^{\lambda}(X)$，$D(X/X_i)\geqslant\lambda>0.5$，即$\dfrac{|X\cap X_i|}{|X_i|}>0.5$，从而可得$\dfrac{|X\cap X_i|}{|X_i-X|}>1.0\geqslant\dfrac{|\underline{R}(X)|}{|X|}$，故 $X_i\in X^{\mathrm{opt}}$. 于是，$\underline{R}^{\lambda}(X)\subseteq X^{\mathrm{opt}}$.

2) 若 $\lambda<\dfrac{|X|}{|X|+|\underline{R}(X)|}$，则 $1-\lambda>\dfrac{|\underline{R}(X)|}{|X|+|\underline{R}(X)|}$，对于 $\forall X_i\in\overline{R}^{\lambda}(X)$，有 $\dfrac{|X_i\cap X|}{|X_i|}>1-\lambda>\dfrac{|\underline{R}(X)|}{|X|+|\underline{R}(X)|}$. 根据不等式的性质，有$\dfrac{|X_i\cap X|}{|X_i|-|X_i\cap X|}>\dfrac{|\underline{R}(X)|}{|X|}$，即$\dfrac{|X_i\cap X|}{|X_i-X|}>\dfrac{|\underline{R}(X)|}{|X|}$，所以，$X_i\in X^{\mathrm{opt}}$. 于是，$\overline{R}^{\lambda}(X)\in X^{\mathrm{opt}}$.

3) 由于 $\underline{R}^{\lambda}(X)\in X^{\mathrm{opt}}$，因此，只需证明对任意 $X_i\in\overline{R}^{\lambda}(X)-\underline{R}^{\lambda}(X)$，有$\dfrac{|X\cap X_i|}{|X_i-X|}<\dfrac{|\underline{R}(X)|}{|X|}$. 事实上，由于 $\lambda<\dfrac{|\underline{R}(X)|}{|X|+|\underline{R}(X)|}$，可得$\dfrac{\lambda}{1-\lambda}<\dfrac{|\underline{R}(X)|}{|X|}$，而对于任意 $X_i\in\overline{R}^{\lambda}(X)-\underline{R}^{\lambda}(X)$，有$\dfrac{|X_i\cap X|}{|X_i|}<\lambda$，从而可得$\dfrac{|X_i\cap X|}{|X_i|-|X_i\cap X|}<\dfrac{\lambda}{1-\lambda}<\dfrac{|\underline{R}(X)|}{|X|}$，即$\dfrac{|X\cap X_i|}{|X_i-X|}<\dfrac{|\underline{R}(X)|}{|X|}$. 综上可得，$\underline{R}^{\lambda}(X)=X^{\mathrm{opt}}$.

4) 因为 X 是近似空间(U,R)上的可定义集，所以 $B_n(X)=\Phi$. 显然，$\underline{R}(X)=\underline{R}^{\lambda}(X)=X^{\mathrm{opt}}$，$\overline{R}(X)=\overline{R}^{\lambda}(X)=X^{\mathrm{opt}}$.

由上述性质可知：对于协调决策表而言，相关的一些结果都相同，因此，一些粗糙集推广模型意义并不是很大. 推广模型更多是建立在不协调决策系统上，也就是说，推广模型更多的是针对系统中边界的不确定性处理问题.

3　几种属性约简之间关系

3.1　基于近似集的分布约简

粗糙集发展至今，提出很多属性约简算法，这些算法从不同角度评价属性的重要性. 文献[23-24]就当前几种属性约简算法差异与联系进行了分析，得到多种约简之间存在包含关系，有些甚至是等价关系. 张文修等[25]在讨论变精度粗糙集模型属性约简时，得到以下 2 点结论：1) 分布约简集未必是变精度上、下近似约简；2) 变精度上、下近似约简集也未必是分布约简集. 接下来，讨论基于最优近似的分布约简与其他分布约简的关系.

决策系统 $\mathrm{DS}=(U,C\cup d)$，C 为条件属性集，d 为决策属性集，$U/d=\{D_1,D_2,\cdots,D_n\}$且 $A\subseteq C$，记 $\underline{R}_A=\{\underline{A}(D_1),\underline{A}(D_2),\cdots,\underline{A}(D_n)\}$，$\overline{R}_A=\{\overline{A}(D_1),\overline{A}(D_2),\cdots,\overline{A}(D_n)\}$，$\underline{R}_A^{\lambda}=\{\underline{A}^{\lambda}(D_1),\underline{A}^{\lambda}(D_2),\cdots,\underline{A}^{\lambda}(D_n)\}$，$\overline{R}_A^{\lambda}=\{\overline{A}^{\lambda}(D_1),\overline{A}^{\lambda}(D_2),\cdots,\overline{A}^{\lambda}(D_n)\}$，$\mu_A=\{D_1{}^{\mathrm{opt}},D_2{}^{\mathrm{opt}},\cdots,D_n{}^{\mathrm{opt}}\}$.

定义 5[19]　决策系统 $\mathrm{DS}=(U,C\cup d)$为不协调决策系统，$B\subseteq A$，则有以下 4 点定义.

1) $\underline{R}_A=\underline{R}_B$，则称 B 为下近似分布协调集. 若 B 为下近似分布协调集，且 B 的任何真子集都不是下近似分布协调集，称 B 为下近似分布约简.

2) $\overline{R}_A=\overline{R}_B$,则称 B 为上近似分布协调集.若 B 为上近似分布协调集,且 B 的任何真子集都不是上近似分布协调集,称 B 为上近似分布约简.

3) $\underline{R}_A^\lambda=\underline{R}_B^\lambda$,则称 B 为 λ 下近似分布协调集.若 B 为 λ 下近似分布协调集,且 B 的任何真子集都不是 λ 下近似分布协调集,称 B 为 λ 下近似分布约简.

4) $\overline{R}_A^\lambda=\overline{R}_B^\lambda$,则称 B 为 λ 上近似分布协调集.若 B 为 λ 上近似分布协调集,且 B 的任何真子集都不是 λ 上近似分布协调集,称 B 为 λ 上近似分布约简.

根据上述定义可知:在一个决策系统中,根据保持决策类近似集的不变原则,可得多种属性重要性的评价标准,从而建立多种属性约简算法.类似地,根据决策类的最优近似,可定义一种新的属性重要性评价标准及属性约简方法,即最优近似分布约简.

定义 6 在决策系统 $\mathrm{DS}=(U,C\cup d)$ 中,$B\subseteq A$,若 $\mu_A=\mu_B$,则称 B 为最优近似分布协调集;若 B 为最优近似分布协调集,且 B 的任何真子集都不是最优近似分布协调集,则称 B 为最优近似分布约简.

3.2 最优近似分布约简与其他分布约简

定理 1 决策系统 $\mathrm{DS}=(U,C\cup d)$,C 为条件属性集,d 为决策属性集,且 $U/C=\{C_1,C_2,\cdots,C_m\}$,$U/d=\{D_1,D_2,\cdots,D_n\}$,$B_n^\delta(D_i)=\left\{C_j\;\middle|\;\dfrac{|C_j\cap D_i|}{|C_j-D_i|}\geqslant\dfrac{|\underline{C}(D_i)|}{|D_i|},C_j\subseteq B_n(D_i)\text{且}C_j\in U/C\right\}$,可得到以下 2 个性质.

1) 对于任意 $D_i\in U/d$,若 $B_n^\delta(D_i)=\Phi$,则最优近似分布约简为下近似分布约简.

2) 对于任意 $D_i\in U/d$,若 $B_n^\delta(D_i)=B_n(D_i)$,则最优近似分布约简为上近似分布约简.

证明:1) 若对于任意 $D_i\in U/d$,若 $B_n^\delta(D_i)=\Phi$,则对于任意 $i(i\leqslant n)$,有 $D_i^{\mathrm{opt}}=\underline{C}(D_i)$,即 $\underline{R}_C=\mu_C$,最优近似分布约简与下近似分布约简等价.

2) 对于任意 $D_i\in U/d$,若 $B_n^\delta(D_i)=B_n(D_i)$,则对于任意 $i(i\leqslant n)$,有 $D_i^{\mathrm{opt}}=\overline{C}(D_i)$,即 $\overline{R}_C=\mu_C$,最优近似分布约简与上近似分布约简等价.

事实上,上述约简等价是因为构造的区分矩阵完全相同.类似地,可以得到以下 3 个定理.

定理 2 决策系统 $\mathrm{DS}=(U,C\cup d)$,C 为条件属性集,d 为决策属性集,且 $U/C=\{C_1,C_2,\cdots,C_m\}$,$U/d=\{D_1,D_2,\cdots,D_n\}$.若对于任意 $C_j\subseteq B_n^\lambda(D_i)=\overline{C}^\lambda(D_i)-\underline{C}^\lambda(D_i)(i\leqslant n)$,有 $\dfrac{|C_j\cap D_i|}{|C_j-D_i|}<\dfrac{|\underline{C}(D_i)|}{|D_i|}$,则最优近似分布约简与以参数为 λ 变精度下近似分布约简等价.

证明:对于任意 $C_j\subseteq B_n^\lambda(D_i)=\overline{C}^\lambda(D_i)-\underline{C}^\lambda(D_i)(i\leqslant n)$,有 $\dfrac{|C_j\cap D_i|}{|C_j-D_i|}<\dfrac{|\underline{C}(D_i)|}{|D_i|}$.于是,对于任意 $i(i\leqslant n)$,有 $D_i^{\mathrm{opt}}=\underline{C}^\lambda(D_i)$,即最优近似分布约简与以参数为 λ 变精度下近似分布约简等价.同理,在定理 2 中,若 $D_i^{\mathrm{opt}}=\overline{C}^\lambda(D_i)(i\leqslant n)$,最优近似分布约简与以参数为 λ 变精度上近似分布约简等价.

定理 3 决策系统 $\mathrm{DS}=(U,C\cup d)$,C 为条件属性集,d 为决策属性集,且 $U/C=\{C_1,C_2,\cdots,C_m\}$,$U/d=\{D_1,D_2,\cdots,D_n\}$.若对于任意 $C_{i,j}\in B_n^\lambda(D_i)=\overline{C}^\lambda(D_i)-\underline{C}^\lambda(D_i)(i\leqslant n)$,有 $D(D_i/C_{i,j})<0.5$,变精度下近似分布约简与下近似分布约简等价.

定理 4 决策系统 $\mathrm{DS}=(U,C\cup d)$,C 为条件属性集,d 为决策属性集,且 $U/C=\{C_1,C_2,\cdots,C_m\}$,$U/d=\{D_1,D_2,\cdots,D_n\}$.若对于任意 $C_{i,j}\in B_n^\lambda(D_i)=\overline{C}^\lambda(D_i)-\underline{C}^\lambda(D_i)(i\leqslant n)$,有 $D(D_i/C_{i,j})>1-\lambda$,变精度上近似分布约简与上近似分布约简等价.

证明:参考定理 1 的证明思路.

由上述定理可知:对于一个给定的近似空间,可以从多角度定义概念的近似集,也可以从多角度讨论决策系统的属性约简;当目标概念的近似结果等价时,对应的分布约简也就等价.

4 实例分析

采用不协调决策系统来验证定理 1,2 的正确性,定理 1 的条件属性集 $C_1=\{a,b,c,e\}$,决策属性集 $d=\{d_1\}$;定理 2 的条件属性集为 $C_2=\{a_1,a_2,a_3,a_4\}$,决策属性集为 $d=\{d_2\}$.实例数据,如表 1 所示.

1) 利用验证定理 1 系统验证最优近似分布约简与 Pawlak 下、上近似分布约简关系,有 $U/C_1=$

$\{X_1, X_2, X_3\}$，$X_1=\{x_1, x_2\}$，$X_2=\{x_3, x_4\}$，$X_3=\{x_5, x_6\}$. $U/d=\{X_4, X_5\}$，其中，$X_4=\{x_1, x_2, x_3\}$，$X_5=\{x_4, x_5, x_6\}$. $\underline{C}_1(X_4)=\{x_1, x_2\}$，$\underline{C}_1(X_5)=\{x_5, x_6\}$；$\overline{C}(_1X_4)=\{x_1, x_2, x_3, x_4\}$，$\overline{C}_1(X_5)=\{x_3, x_4, x_5, x_6\}$. 并且$\frac{|\underline{C}_1(X_4)|}{|X_4|}=\frac{2}{3}$，$\frac{|\underline{C}_1(X_5)|}{|X_5|}=\frac{2}{3}$，$\frac{|X_2\cap X_4|}{|X_2-X_4|}=\frac{|X_2\cap X_5|}{|X_2-X_5|}=1>\frac{2}{3}$. 根据定理 1，最优近似分布约简与上近似分布约简等价.

表 1　定理的不协调决策系统验证

Tab. 1　Uncoordinated decision system for verification theorem

U	定理 1 系统的验证					U	定理 2 系统的验证				
	a	b	c	e	d_1		a_1	a_2	a_3	a_4	d_2
x_1	1	1	1	1	1	y_1	1	0	0	0	1
x_2	1	1	1	1	1	y_2	0	1	1	1	2
x_3	1	1	0	0	1	y_3	0	1	0	0	2
x_4	1	1	0	0	2	y_4	0	1	1	0	2
x_5	0	0	0	0	2	y_5	0	1	0	0	1
x_6	0	0	0	0	2	y_6	0	1	0	0	1

由最优近似集的计算方法可得最优近似分布，即$(X_4^{\mathrm{opt}}, X_5^{\mathrm{opt}})=((x_1, x_2, x_3, x_4),(x_3, x_4, x_5, x_6))$，而这恰好也是上近似分布，根据约简定义，约简结果显然相同.

2）利用验证定理 2 的系统，验证最优近似分布约简与变精度下、上近似分布约简关系. 取 $\lambda=0.6$，根据系统可得，$U/C_2=\{Y_1, Y_2. Y_3, Y_4\}$，$U/d=\{Y_5, Y_6\}$. 其中，$Y_1=\{y_1\}$，$Y_2=\{y_2\}$，$Y_3=\{y_3,\ y_5,\ y_6\}$，$Y_4=\{y_4\}$，$Y_5=\{y_1,\ y_5,\ y_6\}$，$Y_6=\{y_2,\ y_3,\ y_4\}$.

又 $\underline{C}_2{}^{0.6}(Y_5)=\{y_1, y_3, y_5. y_6\}$，$\overline{C}_2(Y_5)=\{y_1, y_3, y_5. y_6\}$，$\overline{C}_2{}^{0.6}(Y_5)-\underline{C}_2{}^{0.6}(Y_5)=\Phi$. $\underline{C}_2{}^{0.6}(Y_6)=\{y_2, y_4\}$，$\overline{C}_2(Y_6)=\{y_2, y_4\}$. 所以，$\overline{C}_2{}^{0.6}(Y_6)-\underline{C}_2{}^{0.6}(Y_6)=\Phi$.

根据文献[25]的研究结果及定理 2 可知：最优近似分布约简与系统 0.6 下分布约简等价，即约简结果为$\{a_3\}$.

事实上，由于$\frac{|\underline{C}_2(Y_5)|}{|Y_5|}=1/3$，$\overline{C}_2(Y_5)-\underline{C}_2(Y_5)=\{y_3, y_5, y_6\}$，$\frac{|\underline{C}_2(Y_6)|}{|Y_6|}=2/3$，$\overline{C}_2(Y_6)-\underline{C}_2(Y_6)=\{y_3, y_5, y_6\}$. 根据最优近似计算方法，最优近似分布为$((y_1, y_3, y_5, y_6),(y_2, y_4))$. 基于最优近似分布约简的区分矩阵转为区分函数，即$(a_1\vee a_2\vee a_3\vee a_4)\wedge(a_1\vee a_2\vee a_3\vee a_4)\wedge(a_3\vee a_4)\wedge a_3=a_3$，其结果与上述结果一致.

上述两个实例虽然只验证了定理 1，2 的部分结论，但是，根据这些结论的验证过程，也可以推出定理 3，4 成立.

5　结束语

粒计算是人工智能领域的一种新理念、新方法，主要用于对不确定、不精确、不完整信息的处理，其本质的思想是通过合适粒度的选择，寻找问题近似的解决方案. 针对粗糙集构建的粒空间上目标概念的多种近似集，讨论 Pawlak 粗糙集近似、变精度近似、最优近似之间的关系，得到这三者之间存在的一些代数关系，且从这些代数关系中发现不同模型在近似集计算的本质特征. 在此基础上，进一步对上述 3 种近似所建立的分布约简之间关系进行讨论，得到一些等价条件. 从这些结论中，不仅可以看到不同近似刻画所对应的属性约简结果之间的关系，以及粗糙集模型近似计算与属性约简完整统一，还可以进一步说明近似集理论作为一种新方法在不确定信息的处理上有着重要的应用.

参考文献：

[1]　苗夺谦，张清华，钱宇华，等. 从人类智能到机器实现模型：粒计算理论与方法[J]. 智能系统学报，2016，11(6)：743-757. DOI：10.11992/tis.201612014.

[2]　梁吉业，钱宇华，李德玉，等. 大数据挖掘的粒计算理论与方法[J]. 中国科学：信息科学，2015，45(11)：1355-1369.

DOI:10.1360/N112015-00092.

[3] PAWLAK Z. Rough sets[J]. Information and Computer Science,l982,11(5):34l-356.

[4] HU Qinghua,MI Jusheng,CHEN Degang. Granular computing based machine learning in the era of big data[J]. Information Sciences,2017,378(2):242-243. DOI:10.1016/j.ins.2016.10.012.

[5] 王健,徐余法,陈国初.基于相对核的属性约简[J].华侨大学学报(自然科学版),2013,34(1):10-13. DOI:10.11830/ISSN.1000-5013.2013.01.0010.

[6] 王敬,王全凤.复杂系统风险分析的模拟方法[J].华侨大学学报(自然科学版),2010,31(3):337-339. DOI:10.11830/ISSN.1000-5013.2010.03.0337.

[7] LIU Dun,LI Tianrui,ZHANG Junbo. Incremental updating approximations in probabilistic under rough sets the variation of attributes[J]. Knowledge-Based Systems,2015,73(81):81-96. DOI:10.1016/j.knosys.2014.09.008.

[8] JING Yunge,LI Tianrui,HAMIDO F,*et al*. An incremental attribute reduction approach based on knowledge granularity with a multi-granulation view[J]. Information Sciences,2017,411(8):23-38. DOI:10.1016/j.ins.2017.05.003.

[9] CHEN Linshu,WANG Jiayang,WANG Weicheng,*et al*. A new granular computing model based on algebraic structure[J]. Chinese Journal of Electronics,2019,28(1):136-142. DOI:10.1049/cje.2018.09.006.

[10] HU Xingchen,WITOLD P,WANG Xianmin. Comparative analysis of logic operators: A perspective of statistical testing and granular computing[J]. Approximate Reasoning,2015,66(11):73-90. DOI:10.1016/j.ijar.2015.07.011.

[11] 王宝丽,梁吉业,胡运红.基于粒计算的犹豫模糊多准则决策方法[J].模式识别与人工智能,2016,29(3):252-262. DOI:10.16451/j.cnki.issn1003-6059.201603007.

[12] 吴伟志,陈颖,徐优红,等.协调的不完备多粒度标记决策系统的最优粒度选择[J].模式识别与人工智能,2016,29(2):108-115. DOI:10.16451/j.cnki.issn1003-6059.201602002.

[13] 吴伟志.多粒度粗糙集数据分析研究的回顾与展望[J].西北大学学报(自然科学版),2018,48(4):501-510. DOI:10.16152/j.cnki.xdxbzr.2018-04-003.

[14] ZIARKO W. Variable precision rough set model [J]. Journal of Computer and System Science,1993,46(1):39-59. DOI:10.1016/0022-0000(93)90048-2.

[15] ZIARKO W. Probabilistic approach to rough sets[J]. Approximate Reasoning,2008,49(2):272-284. DOI:10.1016/j.ijar.2007.06.014.

[16] YAO Yiyu. Probabilistic rough set approximations[J]. Approximate Reasoning,2008,49(2):255-271. DOI:10.1016/j.ijar.2007.05.019.

[17] 张清华,薛玉斌,胡峰,等.粗糙集近似集不确定性研究[J].电子学报,2016,44(7):1574-1580. DOI:10.3969/j.issn.0372-2112.2016.07.008.

[18] 张清华,王国胤,肖雨.粗糙集的近似集[J].软件学报,2012,23(7):1745-1759. DOI:10.3724/SP.J.1001.2012.04226.

[19] 张清华,薛玉斌,王国胤.粗糙集的最优近似集[J].软件学报,2016,27(2):295-308. DOI:10.13328/j.cnki.jos.004854.

[20] 谭安辉,李进金,吴伟志.多粒度粗糙集和覆盖粗糙集间的近似与约简关系[J].模式识别与人工智能,2016,29(8):691-697. DOI:10.16451/j.cnki.issn1003-6059.201608003.

[21] 袁修久,张文修.模糊粗糙集的包含度和相似度[J].模糊系统与数学,2005,19(1):111-115.

[22] 罗来鹏,刘二根,范自柱.最优近似粗糙集[J].河南科技大学学报(自然科学版),2018,39(3):89-93. DOI:10.15926/j.cnki.issn1672-6871.2018.03.017.

[23] 邓大勇,黄厚宽,李向军.不一致决策系统中约简之间的比较[J].电子学报,2007,35(2):252-255. DOI:10.3321/j.issn:0372-2112.2007.02.014.

[24] 杜卫锋,秦克云.不协调决策表几种约简标准及其关系分析[J].电子学报,2011,39(6):1336-1339.

[25] 张文修,梁怡,吴伟志.信息系统与知识发现[M].北京:科学出版社,2003.

(**责任编辑:** 陈志贤　　**英文审校:** 黄心中)