第 30 卷 第 2 期
Vol. 30 No. 2

控 制 与 决 策
Control and Decision

2015 年 2 月
Feb. 2015

文章编号: 1001-0920 (2015) 02-0222-07 **DOI:** 10.13195/j.kzyjc.2014.0006

基于加权粒度的多粒度粗糙集

张 明[1,2], 程 科[1], 杨习贝[1,2], 唐振民[2]

(1. 江苏科技大学 计算机科学与工程学院，江苏 镇江 212003;
2. 南京理工大学 计算机科学与技术学院，南京 210094)

摘 要: 首先, 通过分析现有多粒度粗糙集模型的不足, 提出一种基于粒度加权的多粒度粗糙集模型; 然后, 通过比较得出加权多粒度粗糙集与乐观多粒度、悲观多粒度和可变多粒度粗糙集之间的关系, 讨论加权多粒度粗糙集的性质, 并分析这几种多粒度粗糙集度量之间的关系; 最后, 通过实例分析验证了所提出加权多粒度粗糙集模型的有效性.

关键词: 粗糙集；多粒度；度量
中图分类号: TP18 **文献标志码:** A

Multigranulation rough set based on weighted granulations

ZHANG Ming[1,2], *CHENG Ke*[1], *YANG Xi-bei*[1,2], *TANG Zhen-min*[2]
(1. School of Computer Science and Engineering，Jiangsu University of Science and Technology，Zhenjiang 212003，China；2. School of Computer Science and Technology，Nanjing University of Science and Technology, Nanjing 210094，China．Correspondent：ZHANG Ming，E-mail：zm_fred@163.com)

Abstract: By analyzing the limitations of the current multigranulation rough set(MGRS), a MGRS based on weighted granulations is presented firstly. Then the four rough sets such as MGRS based on weighted granulations, optimistic MGRS, pessimistic MGRS and variable MGRS are compared to obtain their relationships, the properties of the presented MGRS are addressed, and the measurements of these four kinds of MGRS are discussed. Finally, the results of examples show the effectiveness of the MGRS based on weighted granulations .

Keywords: rough set；multigranulation；measurement

0 引 言

多粒度粗糙集数据建模方法是近几年发展起来的一种新的多视角数据分析方法. 该方法由Qian等[1-4]首先提出并用于克服Pawlak粗糙集的缺陷. 例如: 在处理多源信息系统的知识发现时算法耗时过多; 在处理具有高维特征的数据时低效; 在处理多智能agent和分布式信息系统时存在不适应性等. 随后, 多粒度粗糙集得到了很多学者的关注, 文献 [2, 5] 将多粒度粗糙集理论拓展到不完备信息系统; 文献 [6] 将其拓展到多粒度优势关系粗糙集; 文献 [7] 将其拓展到多粒度模糊关系粗糙集; 文献 [8] 将其拓展到多粒度邻域系统粗糙集; 文献 [9] 将其拓展到多粒度决策粗糙集; 文献 [5, 10] 提出了一种可变多粒度粗糙集模型, 将乐观多粒度粗糙集和悲观多粒度粗糙集统一到可变多粒度粗糙集模型中; 文献 [11-12] 研究了多粒度粗糙集的应用工作等. 这些研究工作促进了多粒度粗糙集理论和应用的发展.

经过研究可以发现: 1) 乐观多粒度粗糙集下近似仅要求存在一个粒度满足知识粒与目标概念保有包含关系, 这种决策过于宽松; 2) 悲观多粒度粗糙集下近似要求所有粒度都满足知识粒与目标概念保有包含关系, 这种决策又过于严格; 3) 可变多粒度粗糙集下近似要求满足一定数目的粒度存在知识粒与目标概念保有包含关系, 这种决策考虑了满足要求的粒度空间的数量而没有考虑其质量. 由于在实际应用中不是每个粒度都是平等的, 有的粒度比较重要, 需要赋予更大的权重, 而有的粒度作用小, 可赋予较小的权重. 例如在风险投资中, 决策者会优先考虑占主导地位的重要因素并赋予较大的权值, 参考一些次要的因素并赋予较小的权值, 这些粒度 (属性) 的权值可根

收稿日期: 2014-01-03；**修回日期:** 2014-06-18.
基金项目: 国家自然科学基金项目 (61100116)；江苏省自然科学基金项目 (BK2011492)；现代农业装备与技术实验室开放基金项目 (NZ201303).
作者简介: 张明(1978−), 男, 副教授, 博士, 从事粗糙集、粒计算等研究；唐振民 (1961−), 男, 教授, 博士生导师, 从事模式识别与智能控制系统等研究.

据专家经验获取, 也可以通过 AHP 或者 Delphi 法获得[13-14], 最后综合衡量各种情况以给出判决.

为了解决此类问题, 本文提出一种加权多粒度粗糙集, 研究其性质, 并讨论了 4 种多粒度粗糙集之间的关系. 最后通过实例数据进行了分析和说明.

1　基本概念

1.1　单粒度粗糙集

信息系统 (IS) 通常定义为一个四元组: $\mathrm{IS}=\langle U,\mathrm{AT},V,f\rangle$. 其中: 论域 U 为非空有限全体对象的集合; AT 为非空有限属性集合; $\forall a\in\mathrm{AT}$, V_a 表示属性 a 的值域; V 为全体属性的值域集合, 即 $V=V_{\mathrm{AT}}=\bigcup\limits_{a\in\mathrm{AT}}V_a$; f 为信息函数, $\forall x\in U$, $a\in\mathrm{AT}$, $f(x,a)$ 表示 x 在属性 a 上的取值.

如果 $\mathrm{AT}=C\bigcup\{d\}$, C 为条件属性集, d 为决策属性, 则 $\langle U,C\bigcup\{d\},V,f\rangle$ 又被称为决策信息系统 (DIS).

定义 1　设 $\mathrm{IS}=\langle U,\mathrm{AT},V,f\rangle$, $\forall A\subseteq\mathrm{AT}$, 则 A 上的不可分辨关系定义为

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

由式 (1) 易知, 不可分辨关系 $\mathrm{IND}(A)$ 是 U 上的一个等价关系, 根据粒计算的观点[6]: 等价关系 $\mathrm{IND}(A)$ 对应一个粒度, U 上的划分 $U/\mathrm{IND}(A)=\{[x]_A:x\in U\}$ 对应一个粒结构或粒度空间, 等价类 $[x]_A$ 被称为知识粒.

定义 2　设 $\mathrm{IS}=\langle U,\mathrm{AT},V,f\rangle$, $\forall X\subseteq U$, $A\subseteq\mathrm{AT}$, X 关于属性 A 的下、上近似集分别记为

$$\underline{A}(X)=\{x\in U:[x]_A\subseteq X\}, \tag{2}$$

$$\overline{A}(X)=\{x\in U:[x]_A\bigcap X\neq\varnothing\}. \tag{3}$$

序偶 $\langle\underline{A}(X),\overline{A}(X)\rangle$ 称为 X 关于 A 的 Pawlak 粗糙集. 由于 Pawlak 粗糙集是以单一粒度空间中的知识粒通过下、上近似的形式表示未知的目标, 又称其为单粒度粗糙集.

1.2　多粒度粗糙集

定义 3　设 $\mathrm{IS}=\langle U,\mathrm{AT},V,f\rangle$, 令 A_1, A_2, $\cdots$, A_m 为 AT 的 m 个属性子集. 对于 $\forall X\subseteq U$, 定义 X 关于属性子集 $A_1,A_2,\cdots,A_m$ 的乐观多粒度粗糙集的下、上近似分别为[1]

$$\underline{\sum_{i=1}^{m}A_i}^{O}(X)=\{x\in U:[x]_{A_1}\subseteq X\bigvee [x]_{A_2}\subseteq X\bigvee\cdots\bigvee[x]_{A_m}\subseteq X\}, \tag{4}$$

$$\overline{\sum_{i=1}^{m}A_i}^{O}(X)=\sim\underline{\sum_{i=1}^{m}A_i}^{O}(\sim X), \tag{5}$$

其中: $\sim X$ 为 X 的补集. 序偶

$$\left\langle\underline{\sum_{i=1}^{m}A_i}^{O}(X),\overline{\sum_{i=1}^{m}A_i}^{O}(X)\right\rangle$$

称为 X 关于属性子集 A_1, A_2, $\cdots$, A_m 的乐观多粒度粗糙集.

定义 4　设 $\mathrm{IS}=\langle U,\mathrm{AT},V,f\rangle$, 令 A_1, A_2, $\cdots$, A_m 为 AT 的 m 个属性子集. 对于 $\forall X\subseteq U$, 定义 X 关于属性子集 $A_1,A_2,\cdots,A_m$ 的悲观多粒度粗糙集的下、上近似分别为[3]

$$\underline{\sum_{i=1}^{m}A_i}^{P}(X)=\{x\in U:[x]_{A_1}\subseteq X\bigwedge [x]_{A_2}\subseteq X\bigwedge\cdots\bigwedge[x]_{A_m}\subseteq X\}, \tag{6}$$

$$\overline{\sum_{i=1}^{m}A_i}^{P}(X)=\sim\underline{\sum_{i=1}^{m}A_i}^{P}(\sim X). \tag{7}$$

序偶 $\left\langle\underline{\sum_{i=1}^{m}A_i}^{P}(X),\overline{\sum_{i=1}^{m}A_i}^{P}(X)\right\rangle$ 称为 X 关于属性子集 A_1, A_2, $\cdots$, A_m 的悲观多粒度粗糙集.

1.3　可变多粒度粗糙集

文献 [5, 10] 分析和比较了乐观多粒度和悲观多粒度粗糙集下近似的决策过程, 进而提出了一种可变多粒度粗糙集数据分析方法, 并将乐观多粒度和悲观多粒度粗糙集统一到可变多粒度粗糙集中.

定义 5　设 $\mathrm{IS}=\langle U,\mathrm{AT},V,f\rangle$, 令 $\boldsymbol{A}=\{A_1,A_2,\cdots,A_m\}$ 为 AT 的 m 个属性子集族, 对于 $0<\beta\leqslant 1$, $\forall X\subseteq U$, 定义 X 关于属性子集 $A_1,A_2,\cdots,A_m$ 的 β 可变多粒度粗糙集的下、上近似分别为[10]

$$\underline{\sum_{i=1}^{m}A_i}^{\beta}(X)=\{x\in U:\forall A_i\in T,[x]_{A_i}\subseteq X,\ \text{其中}T\subseteq\mathbf{A}\bigwedge|T|/m\geqslant\beta\}, \tag{8}$$

$$\overline{\sum_{i=1}^{m}A_i}^{\beta}(X)=\sim\underline{\sum_{i=1}^{m}A_i}^{\beta}(\sim X). \tag{9}$$

序偶 $\left\langle\underline{\sum_{i=1}^{m}A_i}^{\beta}(X),\overline{\sum_{i=1}^{m}A_i}^{\beta}(X)\right\rangle$ 称为 X 关于属性子集族 $\boldsymbol{A}$ 的可变多粒度粗糙集.

定理 1　设 $\mathrm{IS}=\langle U,\mathrm{AT},V,f\rangle$, 令 A_1, A_2, $\cdots$, A_m 为 AT 的 m 个属性子集. 对于 $\forall X\subseteq U$, 可变多粒度粗糙集与乐观和悲观多粒度粗糙集具有如下关系[10]:

$$\underline{\sum_{i=1}^{m}A_i}^{\frac{1}{m}}(X)=\underline{\sum_{i=1}^{m}A_i}^{O}(X),$$

$$\overline{\sum_{i=1}^{m}A_i}^{\frac{1}{m}}(X)=\overline{\sum_{i=1}^{m}A_i}^{O}(X);$$

$$\underline{\sum_{i=1}^{m} A_i}^{1}(X)=\underline{\sum_{i=1}^{m} A_i}^{P}(X),$$
$$\overline{\sum_{i=1}^{m} A_i}^{1}(X)=\overline{\sum_{i=1}^{m} A_i}^{P}(X).$$

注 1 定理 1 表明, 当 $\beta=1/m$ 时, 可变多粒度粗糙集退化为乐观多粒度粗糙集; 当 $\beta=1$ 时, 可变多粒度粗糙集退化为悲观多粒度粗糙集.

2 加权多粒度粗糙集

在多粒度粗糙集决策理论中, 若乐观多粒度下近似的决策认为 m 个粒度空间中任一粒度存在知识粒 $[x]_{Ai}$ 与目标概念 X 保有包含关系, 则 x 属于 X 的下近似; 若悲观多粒度下近似的决策认为 m 个粒度空间中的所有粒度都存在知识粒 $[x]_{Ai}$ 与目标概念 X 保有包含关系, 则 x 才属于 X 的下近似; 若可变多粒度粗糙集下近似的决策认为 m 个粒度空间中满足一定数目的粒度存在知识粒 $[x]_{Ai}$ 与目标概念 X 保有包含关系, 则 x 属于 X 的下近似.

在上述多粒度粗糙集决策过程中, 对每个粒度空间的考察是平等且无差别的, 在决策时只考虑了粒度空间的数量, 而没有考虑粒度的质量. 然而, 在实际的应用问题中不是每个粒度空间都是平等的, 有的粒空间重要, 起到决定性作用, 可以赋予较大的权重; 有的粒空间作用小, 可以赋予较小的权重. 例如在风险投资中, 占主导地位的重要因素决策者会优先考虑, 并在决策过程中赋予较大的权重, 参考一些次要的因素并赋予较小的权重 (粒度权值的分配可根据领域专家的经验获得, 也可通过 AHP 或 Delphi 方法得到), 最终综合所有因素给出判决. 为了解决此类问题, 下面给出一种基于加权粒度的多粒度粗糙集.

定义 6 设 $\mathrm{IS}=\langle U,\mathrm{AT},V,f\rangle$, $0<\beta\leqslant 1$, $\boldsymbol{A}=\{A_1,A_2,\cdots,A_m\}$ 为 AT 的 m 个属性子集族. 若由 $\boldsymbol{A}$ 导出的粒空间 $U/\mathrm{IND}(A_1), U/\mathrm{IND}(A_2),\cdots, U/\mathrm{IND}(A_m)$ 对应的粒度权重为 $\alpha=\{\alpha_{A_1},\alpha_{A_2},\cdots,\alpha_{A_m}\}$, 且 $\sum_{i=1}^{m}\alpha_{A_i}=1$, 则对于 $\forall X\subseteq U$, X 关于 $\boldsymbol{A}$ 的加权多粒度粗糙集的下、上近似定义为

$$\underline{\sum_{i=1}^{m} A_i}_{\alpha}^{\beta}(X)=\Big\{x\in U:\forall A_i\in T,[x]_{A_i}\subseteq X,\ \text{其中}\, T\subseteq\boldsymbol{A}\wedge\sum_{j=1}^{|T|}\alpha_{A_j}\geqslant\beta\Big\}, \tag{10}$$

$$\overline{\sum_{i=1}^{m} A_i}_{\alpha}^{\beta}(X)=\sim\underline{\sum_{i=1}^{m} A_i}_{\alpha}^{\beta}(\sim X). \tag{11}$$

序偶 $\Big\langle\underline{\sum_{i=1}^{m} A_i}_{\alpha}^{\beta}(X),\overline{\sum_{i=1}^{m} A_i}_{\alpha}^{\beta}(X)\Big\rangle$ 称为 X 关于属性子集族 $\boldsymbol{A}$ 的加权多粒度粗糙集. X 的加权多粒度粗糙集的正域、负域和边界域分别记为

$$\mathrm{POS}^{\alpha,\beta}_{\sum_{i=1}^{m}A_i}(X)=\underline{\sum_{i=1}^{m} A_i}_{\alpha}^{\beta}(X), \tag{12}$$

$$\mathrm{NEG}^{\alpha,\beta}_{\sum_{i=1}^{m}A_i}(X)=U-\overline{\sum_{i=1}^{m} A_i}_{\alpha}^{\beta}(X), \tag{13}$$

$$\mathrm{BND}^{\alpha,\beta}_{\sum_{i=1}^{m}A_i}(X)=\overline{\sum_{i=1}^{m} A_i}_{\alpha}^{\beta}(X)-\underline{\sum_{i=1}^{m} A_i}_{\alpha}^{\beta}(X). \tag{14}$$

定理 2 设 $\mathrm{IS}=\langle U,\mathrm{AT},V,f\rangle$, $0<\beta\leqslant 1$, $\boldsymbol{A}=\{A_1,A_2,\cdots,A_m\}$ 为 AT 的 m 个属性子集族. 由 $\boldsymbol{A}$ 导出的一簇粒空间对应分配的粒度权重为 $\alpha=\{\alpha_{A_1},\alpha_{A_2},\cdots,\alpha_{A_m}\}$ 且 $\sum_{i=1}^{m}\alpha_{A_i}=1$. 若 $\alpha_{A_1}=\alpha_{A_2}=\cdots=\alpha_{A_m}=1/m$, 则有

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

证明 由定义 6 可知, 当 $\alpha_{A_1}=\alpha_{A_2}=\cdots=\alpha_{A_m}=1/m$ 时, 有 $\sum_{j=1}^{|T|}\alpha_{A_j}=|T|/m$ 成立, 此时加权多粒度粗糙集下、上近似的定义和可变多粒度粗糙集的下、上近似定义等价. □

注 2 当 $\alpha_{A_1}=\alpha_{A_2}=\cdots=\alpha_{A_m}=1/m$ 时, 由定理 2 易知, 加权多粒度粗糙集为可变多粒度粗糙集; 再结合定理 1 可知, 当 $\beta=1/m$ 时, 加权多粒度粗糙集便退化为乐观多粒度粗糙集; 当 $\beta=1$ 时, 加权多粒度粗糙集便退化为悲观多粒度粗糙集. 由此可知, 加权多粒度粗糙集是可变多粒度粗糙集、乐观和悲观多粒度粗糙集的泛化, 可变多粒度粗糙集、乐观和悲观多粒度粗糙集都是加权多粒度粗糙集的特例.

定理 3 设 $\mathrm{IS}=\langle U,\mathrm{AT},V,f\rangle$, $0<\beta\leqslant 1$, $\boldsymbol{A}=\{A_1,A_2,\cdots,A_m\}$ 为 AT 的 m 个属性子集族. 若由 $\boldsymbol{A}$ 导出的一簇粒空间对应分配的粒度权重为 $\alpha=\{\alpha_{A_1},\alpha_{A_2},\cdots,\alpha_{A_m}\}$, 且 $\sum_{i=1}^{m}\alpha_{A_i}=1$, 则加权多粒度粗糙集与乐观和悲观多粒度粗糙集相比有如下性质:

$$\underline{\sum_{i=1}^{m} A_i}^{P}(X)\subseteq\underline{\sum_{i=1}^{m} A_i}_{\alpha}^{\beta}(X)\subseteq\underline{\sum_{i=1}^{m} A_i}^{O}(X),$$
$$\overline{\sum_{i=1}^{m} A_i}^{O}(X)\subseteq\overline{\sum_{i=1}^{m} A_i}_{\alpha}^{\beta}(X)\subseteq\overline{\sum_{i=1}^{m} A_i}^{P}(X).$$

由定义 3, 定义 4 和定义 6 易证, 故证明过程略.

注 3 定理 3 表明加权多粒度粗糙集的下近似介于悲观和乐观多粒度粗糙集的下近似之间, 而上近

似介于乐观和悲观多粒度粗糙集的上近似之间.

定理 4　设 $\text{IS}=\langle U,\text{AT},V,f\rangle$, $0<\beta\leqslant 1$, $\boldsymbol{A}=\{A_1,A_2,\cdots,A_m\}$ 为 AT 的 m 个属性子集族. 若由 $\boldsymbol{A}$ 导出的一簇粒空间对应分配的粒度权重为 $\alpha=\{\alpha_{A_1},\alpha_{A_2},\cdots,\alpha_{A_m}\}$, 且 $\sum\limits_{i=1}^{m}\alpha_{A_i}=1$, 则加权多粒度粗糙集有如下的性质:

1)

$$\underline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(X)\subseteq X\subseteq\overline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(X);$$

2)

$$\underline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(U)=\overline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(U)=U,$$

$$\underline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(\varnothing)=\overline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(\varnothing)=\varnothing;$$

3)

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

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

4)

$$X\subseteq Y\Rightarrow\underline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(X)\subseteq\underline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(Y),$$

$$\overline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(X)\subseteq\overline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(Y);$$

5)

$$\underline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}\Big(\underline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(X)\Big)=\underline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(X),$$

$$\overline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}\Big(\overline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(X)\Big)=\overline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(X);$$

6)

$$\beta_1\leqslant\beta_2\Rightarrow\begin{cases}\underline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta_1}(X)\supseteq\underline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta_2}(X),\\ \overline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta_1}(X)\subseteq\overline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta_2}(X).\end{cases}$$

证明　下面仅证明性质 1, 其他证明类似.

1) 对于任意 $x\in\underline{\sum\limits_{i=1}^{m}A_i}_{\alpha}^{\beta}(X)$, $\forall A_i\in T$, 有 $[x]_{A_i}\subseteq X$ 成立. 其中: $T\subseteq\boldsymbol{A}$, $\sum\limits_{j=1}^{|T|}\alpha_{A_j}\geqslant\beta$. 又因为 $x\in[x]_{A_i}$, 所以 $x\in X$. 因此有 $\underline{\sum\limits_{i=1}^{m}A_i}_{\alpha}^{\beta}(X)\subseteq X$.

2) 设 $x\notin\overline{\sum\limits_{i=1}^{m}A_i}_{\alpha}^{\beta}(X)$, 有 $x\in\underline{\sum\limits_{i=1}^{m}A_i}_{\alpha}^{\beta}(\sim X)$ 成立. 其中: $T\subseteq\boldsymbol{A}$, $\sum\limits_{j=1}^{|T|}\alpha_{A_j}\geqslant\beta$. 那么对于任意 $A_i\in T$, 有 $[x]_{A_i}\subseteq(\sim X)$. 因此 $x\in(\sim X)$, 即 $x\notin X$, 故有 $X\subseteq\overline{\sum\limits_{i=1}^{m}A_i}_{\alpha}^{\beta}(X)$ 成立.

综上可知性质 1 成立. □

注 4　定理 4 说明了加权多粒度粗糙集的基本性质. 其中: 性质 1 表明加权多粒度粗糙集的下近似、上近似和目标概念之间的包含关系; 性质 2～性质 5 表明加权多粒度粗糙集的下近似、上近似运算满足同一律、互补率、单调性和幂等率; 性质 6 表明加权多粒度粗糙集的下近似、上近似运算关于阈值 β 是单调的.

3　多粒度粗糙集度量

从粗糙集近似的角度来看, 知识的不确定性是由边界域的存在而造成的, 边界域越大粗糙程度越大. 本节从粗糙精度、属性依赖度和目标概念的可精确近似程度等方面来度量加权多粒度粗糙集.

定义 7　设 $\text{IS}=\langle U,\text{AT},V,f\rangle$, $0<\beta\leqslant 1$, $\boldsymbol{A}=\{A_1,A_2,\cdots,A_m\}$ 为 AT 的 m 个属性子集族. 由 $\boldsymbol{A}$ 导出的一簇粒空间对应分配的粒度权重为 $\alpha=\{\alpha_{A_1},\alpha_{A_2},\cdots,\alpha_{A_m}\}$, 且 $\sum\limits_{i=1}^{m}\alpha_{A_i}=1$. 对于任意 $X\subseteq U(X\neq\varnothing)$, X 关于属性子集族 $\boldsymbol{A}$ 的乐观和悲观多粒度、可变多粒度和加权多粒度粗糙集的粗糙精度因子定义为

$$\alpha_1\Big(\sum_{i=1}^{m}A_i,X\Big)=\frac{\Big|\underline{\sum_{i=1}^{m}A_i}^{O}(X)\Big|}{\Big|\overline{\sum_{i=1}^{m}A_i}^{O}(X)\Big|},\tag{15}$$

$$\alpha_2\Big(\sum_{i=1}^{m}A_i,X\Big)=\frac{\Big|\underline{\sum_{i=1}^{m}A_i}^{P}(X)\Big|}{\Big|\overline{\sum_{i=1}^{m}A_i}^{P}(X)\Big|},\tag{16}$$

$$\alpha_3\Big(\sum_{i=1}^{m}A_i,X\Big)=\frac{\Big|\underline{\sum_{i=1}^{m}A_i}^{\beta}(X)\Big|}{\Big|\overline{\sum_{i=1}^{m}A_i}^{\beta}(X)\Big|},\tag{17}$$

$$\alpha_4\Big(\sum_{i=1}^{m}A_i,X\Big)=\frac{\left|\underline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(X)\right|}{\left|\overline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(X)\right|}. \tag{18}$$

定义 8 设 DIS $=\langle U,C\bigcup\{d\},V,f\rangle, 0<\beta\leqslant 1$, $\boldsymbol{A}=\{A_1,A_2,\cdots,A_m\}$ 为 C 的 m 个条件属性子集族. 由 $\boldsymbol{A}$ 导出的一簇粒空间对应分配的粒度权重为 $\alpha=\{\alpha_{A_1},\alpha_{A_2},\cdots,\alpha_{A_m}\}$, 且 $\sum\limits_{i=1}^{m}\alpha_{A_i}=1$. $\boldsymbol{D}=\{D_1, D_2,\cdots,D_r\}$ 为由决策属性 d 在论域 U 上导出的划分, 则决策类 $\boldsymbol{D}$ 由条件属性子集族 $\boldsymbol{A}$ 确定的粗糙分类能力, 即对 $\boldsymbol{A}$ 的属性依赖程度记为 γ, 且乐观和悲观多粒度粗糙集、可变多粒度粗糙集和加权多粒度粗糙集的依赖度因子分别记为

$$\gamma_1\Big(\sum_{i=1}^{m}A_i,\boldsymbol{D}\Big)=\frac{\sum\limits_{j=1}^{r}\Big\{\Big|\underline{\sum_{i=1}^{m}A_i}^{O}(D_j)\Big|:D_j\in\boldsymbol{D}\Big\}}{|U|}, \tag{19}$$

$$\gamma_2\Big(\sum_{i=1}^{m}A_i,\boldsymbol{D}\Big)=\frac{\sum\limits_{j=1}^{r}\Big\{\Big|\underline{\sum_{i=1}^{m}A_i}^{P}(D_j)\Big|:D_j\in\boldsymbol{D}\Big\}}{|U|}, \tag{20}$$

$$\gamma_3\Big(\sum_{i=1}^{m}A_i,\boldsymbol{D}\Big)=\frac{\sum\limits_{j=1}^{r}\Big\{\Big|\underline{\sum_{i=1}^{m}A_i}^{\beta}(D_j)\Big|:D_j\in\boldsymbol{D}\Big\}}{|U|}, \tag{21}$$

$$\gamma_4\Big(\sum_{i=1}^{m}A_i,\boldsymbol{D}\Big)=\frac{\sum\limits_{j=1}^{r}\Big\{\Big|\underline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(D_j)\Big|:D_j\in\boldsymbol{D}\Big\}}{|U|}. \tag{22}$$

定义 9 设 IS $=\langle U,\mathrm{AT},V,f\rangle$, $0<\beta\leqslant 1$, $\boldsymbol{A}=\{A_1,A_2,\cdots,A_m\}$ 为 AT 的 m 个属性子集族. 由 $\boldsymbol{A}$ 导出的一簇粒空间对应分配的粒度权重为 $\alpha=\{\alpha_{A_1},\alpha_{A_2},\cdots,\alpha_{A_m}\}$, 且 $\sum\limits_{i=1}^{m}\alpha_{A_i}=1$. 对于任意 $X\subseteq U(X\neq\varnothing)$, X 在属性子集族 $\boldsymbol{A}$ 下可被精确近似的程度记为 π, 且乐观和悲观多粒度、可变多粒度和加权多粒度粗糙集的精确近似的程度因子分别记为

$$\pi_1\Big(\sum_{i=1}^{m}A_i,X\Big)=\Big|\underline{\sum_{i=1}^{m}A_i}^{O}(X)\Big|\Big/|X|, \tag{23}$$

$$\pi_2\Big(\sum_{i=1}^{m}A_i,X\Big)=\Big|\underline{\sum_{i=1}^{m}A_i}^{P}(X)\Big|\Big/|X|, \tag{24}$$

$$\pi_3\Big(\sum_{i=1}^{m}A_i,X\Big)=\Big|\underline{\sum_{i=1}^{m}A_i}^{\beta}(X)\Big|\Big/|X|, \tag{25}$$

$$\pi_4\Big(\sum_{i=1}^{m}A_i,X\Big)=\Big|\underline{\sum_{i=1}^{m}A_i}_{\alpha}^{\beta}(X)\Big|\Big/|X|. \tag{26}$$

定理 5 设 DIS $=\langle U,C\bigcup\{d\},V,f\rangle, 0<\beta\leqslant 1$, $\boldsymbol{A}=\{A_1,A_2,\cdots,A_m\}$ 为 C 的 m 个条件属性子集族. 由 $\boldsymbol{A}$ 导出的一簇粒空间对应分配的粒度权重为 $\alpha=\{\alpha_{A_1},\alpha_{A_2},\cdots,\alpha_{A_m}\}$, 且 $\sum\limits_{i=1}^{m}\alpha_{A_i}=1$. 对于任意 $X\subseteq U(X\neq\varnothing)$, X 关于条件属性子集族 $\boldsymbol{A}$ 的加权多粒度粗糙集的粗糙精度因子、依赖度因子和可精确近似程度与乐观和悲观多粒度糙集对应的度量因子有如下的关系成立:

$$\alpha_2\Big(\sum_{i=1}^{m}A_i,X\Big)\leqslant\alpha_4\Big(\sum_{i=1}^{m}A_i,X\Big)\leqslant\alpha_1\Big(\sum_{i=1}^{m}A_i,X\Big), \tag{27}$$

$$\alpha_2\Big(\sum_{i=1}^{m}A_i,X\Big)\leqslant\alpha_3\Big(\sum_{i=1}^{m}A_i,X\Big)\leqslant\alpha_1\Big(\sum_{i=1}^{m}A_i,X\Big), \tag{28}$$

$$\gamma_2\Big(\sum_{i=1}^{m}A_i,X\Big)\leqslant\gamma_4\Big(\sum_{i=1}^{m}A_i,X\Big)\leqslant\gamma_1\Big(\sum_{i=1}^{m}A_i,X\Big), \tag{29}$$

$$\gamma_2\Big(\sum_{i=1}^{m}A_i,X\Big)\leqslant\gamma_3\Big(\sum_{i=1}^{m}A_i,X\Big)\leqslant\gamma_1\Big(\sum_{i=1}^{m}A_i,X\Big), \tag{30}$$

$$\pi_2\Big(\sum_{i=1}^{m}A_i,X\Big)\leqslant\pi_4\Big(\sum_{i=1}^{m}A_i,X\Big)\leqslant\pi_1\Big(\sum_{i=1}^{m}A_i,X\Big), \tag{31}$$

$$\pi_2\Big(\sum_{i=1}^{m}A_i,X\Big)\leqslant\pi_3\Big(\sum_{i=1}^{m}A_i,X\Big)\leqslant\pi_1\Big(\sum_{i=1}^{m}A_i,X\Big). \tag{32}$$

证明 由定理 4 和定义 7 易知

$$\alpha_2\Big(\sum_{i=1}^{m}A_i,X\Big)\leqslant\alpha_4\Big(\sum_{i=1}^{m}A_i,X\Big)\leqslant\alpha_1\Big(\sum_{i=1}^{m}A_i,X\Big)$$

成立, 所以式 (27) 成立. 同理可以证明其他关系也成立. □

注 5 定理 6 表明: 加权多粒度粗糙集的 3 种度量因子介于悲观多粒度粗糙集和乐观多粒度粗糙集的 3 种度量因子之间; 可变多粒度粗糙集的 3 种度量因子也介于悲观多粒度粗糙集和乐观多粒度粗糙集的 3 种度量因子之间; 加权多粒度粗糙集的 3 种度量因子和可变多粒度粗糙集的 3 种度量因子之间没有严格的大小关系.

4 实例分析

表 1 为一个关于风险投资的决策信息系统实例. 下面分别采用加权多粒度粗糙集决策分析方法、乐观和悲观多粒度粗糙集和可变多粒度粗糙集的决

策分析方法进行分析和比较. 其中: 论域 $U=\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$ 表示待决策的风险投资项目; $C=\{$Locus, Population density, Transportation, Investment$\}=\{a_1, a_2, a_3, a_4\}$ 为条件属性集, 条件属性的取值表示不同领域专家对风险投资项目所处的位置、人口密度、交通情况和投资额度给出的评价; d 为决策属性, 其取值表示该项目是否可以得到风险投资等.

表 1　风险投资决策信息系统表

U	a_1	a_2	a_3	a_4	d
x_1	Common	Big	Good	High	Yes
x_2	Bad	Small	Bad	Low	No
x_3	Bad	Big	Good	High	Yes
x_4	Bad	Medium	Common	Low	No
x_5	Good	Medium	Common	High	Yes
x_6	Bad	Medium	Common	Medium	No
x_7	Bad	Big	Bad	Medium	Yes
x_8	Bad	Small	Common	Medium	No
x_9	Bad	Big	Good	Medium	Yes

设决策信息系统的条件属性子集族 $\boldsymbol{A}=\{A_1, A_2, A_3, A_4\}=\{\{a_1\},\{a_2\},\{a_3\},\{a_4\}\}$, 根据专家的经验对位置、人口密度、交通情况和投资额度 4 种属性对应的粒度权值分配 $\alpha=\{0.1, 0.4, 0.3, 0.2\}$, 阈值 $\beta=0.6$.

Step 1: 根据条件属性子集族 $\boldsymbol{A}$, 计算每个领域专家的知识粒度空间如下:

$$U/\mathrm{IND}(A_1)=\{\{x_1\},\{x_2, x_3, x_4, x_6, x_7, x_8, x_9\},\{x_5\}\},$$
$$U/\mathrm{IND}(A_2)=\{\{x_1, x_3, x_7, x_9\},\{x_2, x_8\},\{x_4, x_5, x_6\}\},$$
$$U/\mathrm{IND}(A_3)=\{\{x_1, x_3, x_9\},\{x_2, x_7\},\{x_4, x_5, x_6, x_8\}\},$$
$$U/\mathrm{IND}(A_4)=\{\{x_1, x_3, x_5\},\{x_2, x_4\},\{x_6, x_7, x_8, x_9\}\}.$$

Step 2: 由决策属性 d 计算决策类如下:

$$U/\mathrm{IND}(\{d\})=\{D_1, D_2\}=$$
$$\{\{x_2, x_4, x_6, x_8\},\{x_1, x_3, x_5, x_7, x_9\}\}.$$

Step 3: 根据定义 3 ∼ 定义 6, 分别可计算出决策类的乐观和悲观多粒度粗糙集, 可变多粒度粗糙集和加权多粒度粗糙集的下近似和上近似如下:

1) 决策类 D_1 的下近似为

$$\underline{\sum_{i=1}^{4} A_i}^{O}(D_1)=\{x_2, x_4, x_8\},$$
$$\underline{\sum_{i=1}^{4} A_i}^{P}(D_1)=\varnothing,\ \underline{\sum_{i=1}^{4} A_i}^{\beta}(D_1)=\varnothing,$$
$$\underline{\sum_{i=1}^{4} A_i}^{\beta}_{\alpha}(D_1)=\{x_2\};$$

2) 决策类 D_2 的下近似为

$$\underline{\sum_{i=1}^{4} A_i}^{O}(D_2)=\{x_1, x_3, x_5, x_7, x_9\},$$
$$\underline{\sum_{i=1}^{4} A_i}^{p}(D_2)=\{x_1\},\ \underline{\sum_{i=1}^{4} A_i}^{\beta}(D_2)=\{x_1, x_3\},$$
$$\underline{\sum_{i=1}^{4} A_i}^{\beta}_{\alpha}(D_2)=\{x_1, x_3, x_9\};$$

3) 同理, 可计算决策类 D_1 的上近似为

$$\overline{\sum_{i=1}^{4} A_i}^{O}(D_1)=\{x_2, x_4, x_6, x_8\},$$
$$\overline{\sum_{i=1}^{4} A_i}^{P}(D_1)=\{x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\},$$
$$\overline{\sum_{i=1}^{4} A_i}^{\beta}(D_1)=\{x_2, x_4, x_5, x_6, x_7, x_8, x_9\},$$
$$\overline{\sum_{i=1}^{4} A_i}^{\beta}_{\alpha}(D_1)=\{x_2, x_4, x_5, x_6, x_7, x_8\};$$

4) 决策类 D_2 的上近似为

$$\overline{\sum_{i=1}^{4} A_i}^{O}(D_2)=\{x_1, x_3, x_4, x_6, x_7, x_9\},$$
$$\overline{\sum_{i=1}^{4} A_i}^{P}(D_2)=\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\},$$
$$\overline{\sum_{i=1}^{4} A_i}^{\beta}(D_2)=\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\},$$
$$\overline{\sum_{i=1}^{4} A_i}^{\beta}_{\alpha}(D_2)=\{x_1, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}.$$

由计算结果 1) 和 2) 易知下式成立:

$$\underline{\sum_{i=1}^{4} A_i}^{P}(D_1)\subseteq\underline{\sum_{i=1}^{4} A_i}^{\beta}_{\alpha}(D_1)\subseteq\underline{\sum_{i=1}^{4} A_i}^{O}(D_1),$$
$$\underline{\sum_{i=1}^{4} A_i}^{P}(D_2)\subseteq\underline{\sum_{i=1}^{4} A_i}^{\beta}_{\alpha}(D_2)\subseteq\underline{\sum_{i=1}^{4} A_i}^{O}(D_2).$$

由计算结果 3) 和 4) 易知下式成立:

$$\overline{\sum_{i=1}^{4} A_i}^{O}(D_1)\subseteq\overline{\sum_{i=1}^{4} A_i}^{\beta}_{\alpha}(D_1)\subseteq\overline{\sum_{i=1}^{4} A_i}^{P}(D_1),$$
$$\overline{\sum_{i=1}^{4} A_i}^{O}(D_2)\subseteq\overline{\sum_{i=1}^{4} A_i}^{\beta}_{\alpha}(D_2)\subseteq\overline{\sum_{i=1}^{4} A_i}^{P}(D_2).$$

这个结果验证了定理 4 的正确性.

注 6　针对风险投资项目 x_9, 在专家 A_1 的知识空间中有知识粒 $[x_9]_{A_1}=\{x_2, x_3, x_4, x_6, x_7, x_8, x_9\}\nsubseteq D_2$ 成立; 专家 A_2 有知识粒 $[x_9]_{A_2}=\{x_1, x_3, x_7, x_9\}\subseteq D_2$ 成立; 专家 A_3 有知识粒 $[x_9]_{A_3}=$

$\{x_1, x_3, x_9\} \subseteq D_2$ 成立; 专家 A_4 有知识粒 $[x_9]_{A_4} = \{x_6, x_7, x_8, x_9\} \nsubseteq D_2$ 成立. 在这4个专家的知识粒度空间中有2个粒度空间满足知识粒与目标概念之间的包含关系. 由定义可知: 1) 乐观多粒度下近似决策要求存在一个粒度满足知识粒与目标概念保有包含关系, 因此 x_9 是 D_2 的乐观多粒度下近似; 2) 悲观多粒度下近似决策要求所有粒度都满足知识粒与目标概念保有包含关系, 因此 x_9 不是 D_2 的悲观多粒度下近似; 3) 可变多粒度下近似决策要求满足一定数目的粒度空间 $(0.5 < \beta = 0.6)$ 存在知识粒与目标概念保有包含关系, 因此 x_9 也不是 D_2 的可变粒度下近似; 4) 加权多粒度粗糙集的下近似决策认为满足一定数目的加权粒度空间 $((\alpha_{A_2} + \alpha_{A_3}) = 0.7 > (\beta = 0.6))$ 存在知识粒与目标概念保有包含关系, 因此 x_9 是 D_2 的加权多粒度下近似.

由上述分析可知, 本例中的风险投资项目 x_9 虽然在4个粒度下只有两个粒度给予好评, 但这两个粒度起的作用比较大(人口密度权重为0.4, 交通情况权值为0.3), 综合起来权值达到了 $0.7 > \beta = 0.6$, 按照决策规则本文认为该项目是可获得风险投资的. 因此, 加权多粒度粗糙集的数据分析方法更具有实际意义, 它在决策过程中不仅考虑了满足包含关系的粒度空间的数量, 而且还考虑了满足条件的粒度空间的质量.

5 结 论

将粗糙集理论用于知识获取、决策支持系统是粗糙集理论和应用研究的热点和难点问题. 本文对现有多粒度粗糙集模型进行改进, 提出了一种加权多粒度粗糙集决策分析方法, 丰富和发展了多粒度粗糙集理论. 该方法可应用于多领域专家的决策分析等应用问题. 接下来将进一步研究加权多粒度粗糙集中决策规则获取和约简、粒度权值的分配, 以及加权多粒度粗糙集的应用等问题.

参考文献(References)

[1] Qian Y H, Liang J Y , Yao Y Y, et al. MGRS: A multi-granulation rough set[J]. Information Sciences, 2010, 180(6): 949-970.

[2] Qian Y H, Liang J Y, Dang C Y. Incomplete multigranulation rough set[J]. IEEE Trans on Systems, Man, and Cybernetics: Part A, 2010, 40(2): 420-431.

[3] Qian Y H, Liang J Y, Wei W. Pessimistic rough decision[C]. The 2nd Int Workshop on Rough Sets Theory. Zhoushan, 2010: 440-449.

[4] 桑妍丽, 钱宇华. 一种悲观多粒度粗糙集中的粒度约简算法[J]. 模式识别与人工智能, 2012, 25(3): 361-366.
(San Y L, Qian Y H. A granular space reduction approach to pessimistic multi-granulation rough set[J]. Pattern Recognition and Artificial Intelligence, 2012, 25(3): 361-366.)

[5] Zhang M, Xu W Y, Yang X B, et al. Incomplete variable multigranulation rough sets decision[J]. Applied Mathematics & Information Sciences, 2014, 8(3): 1159-1166.

[6] Xu W H, Sun W X, Zhang X Y, et al. Multiple granulation rough set approach to ordered information systems[J]. Int J of General Systemsl, 2012, 41(5): 475-501.

[7] Yang X B, Song X N, Dou H L, et al. Multi-granulation rough set: From crisp to fuzzy case[J]. Annals of Fuzzy Mathematics and Informatics, 2011, 1(1): 55-70.

[8] Lin G P, Qian Y H, Li J J. NMGRS: Neighborhood-based multigranulation rough sets[J]. Int J of Approximate Reasoning, 2012, 53(7): 1080-1093.

[9] Qian Y H , Zhang H, Sang Y L, et al. Multigranulation decision-theoretic rough sets[J]. Int J of Approximate Reasoning, 2014, 55(1): 225-237.

[10] 张明, 唐振民, 徐维艳, 等. 可变多粒度粗糙集模型[J]. 模式识别与人工智能, 2012, 25(4): 709-720.
(Zhang M, Tang Z M, Xu W Y, et al. Variable multigranulation rough set model[J]. Pattern Recognition and Artificial Intelligence, 2012, 25(4): 709-720.)

[11] Wu W Z, Leung Y. Theory and applications of granular labelled partitions in multi-scale decision tables[J]. Information Sciences, 2011, 181(18): 3878-3897.

[12] She Y H, He X L. On the structure of the multigranulation rough set model[J]. Knowledge-Based Systems, 2012, 36(1): 81-92.

[13] 田军, 张朋柱, 王刊良, 等. 基于德尔菲法的专家意见集成模型研究[J]. 系统工程理论与实践, 2004, 24(1): 55-62.
(Tian J, Zhang P Z, Wang K L, et al. The integraiong model of expert's opinion based on Delphi method[J]. Systems Engineering - Theory & Practice, 2004, 24(1): 55-62.)

[14] 吕跃进, 程宏涛, 覃菊莹. 基于判断可信度的层次分析排序方法[J]. 控制与决策, 2012, 27(5): 787-791.
(Lü Y J, Cheng H T, Qin J Y. Ranking method for AHP based judgement credibility[J]. Control and Decision, 2012, 27(5): 787-791.)

(责任编辑: 滕 蓉)