信阳师范学院学报(自然科学版)
JOURNAL OF XINYANG NORMAL UNIVERSITY (Natural Science Edition)

多标记决策表的最优粒度选择

史进玲， 王伟

引用本文:
史进玲, 王伟. 多标记决策表的最优粒度选择[J]. 信阳师范学院学报自然科学版, 2021, 34(4): 549–554. doi: 10.3969/j.issn.1003–0972.2021.04.007
SHI Jinling, WANG Wei. Optimal Granularity Selection in Multi–label Decision Table[J]. Journal of Xinyang Normal University (Natural Science Edition), 2021, 34(4): 549–554. doi: 10.3969/j.issn.1003–0972.2021.04.007

在线阅读 View online: https://doi.org/10.3969/j.issn.1003–0972.2021.04.007

DOI:10.3969/j.issn.1003-0972.2021.04.007

文章编号:1003-0972(2021)04-0549-06

多标记决策表的最优粒度选择

史进玲[1*],王　伟[2]

(1.许昌学院 国际教育学院,河南 许昌 461000;2.河南师范大学 计算机与信息工程学院,河南 新乡 453007)

摘　要:针对多标记决策分类中的粒度选择问题,提出了基于决策表的全局最优粒度选择方法和基于对象的局部最优粒度选择方法.首先基于多个粒度层次分析了多标记决策表的粒度划分,引入了多粒度多标记决策表的粒化粗糙度度量方法;然后针对协调决策表和不协调决策表讨论了通用的决策表最优粒度选择方法;最后,针对全局最优粒度选择不能使每个对象都达到最优粒度的局限性,以及不协调决策表中有些对象关于决策标记分类的不确定性问题,讨论了对象的局部最优粒度选择方法,并结合实例验证了该方法的有效性.

关键词:多粒度;多标记决策表;粗糙度;最优粒度选择

中图分类号:TP182　　**文献标识码**:A　　**开放科学(资源服务)标识码(OSID)**:

Optimal Granularity Selection in Multi-label Decision Table

SHI Jinling[1*], WANG Wei[2]

(1. School of International Education, Xuchang University, Xuchang 461000, China;
2. College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China)

Abstract: To solve the problem of granularity selection in multi-label decision classification, a global and object-based optimal granularity selection method is proposed. Firstly, the granularity partitioning of multi-label decision tables is analyzed based on multiple granularity levels and the granulation roughness measurement method of multi-label decision tables is introduced. Then the general optimal granularity selection method is discussed for coordinated decision tables and uncoordinated decision tables. Finally, in view of the limitation that the global optimal granularity selection cannot make every object reach the optimal granularity, and the uncertainty of some objects in the uncoordinated decision table regarding the classification of decision markers, the local optimal granularity selection method of objects is discussed, and the effectiveness of the method is verified by an example.

Key words: multi-granularity; multi-label decision table; roughness; optimal granularity selection

0　引言

粒计算[1]是处理复杂问题求解、海量数据挖掘的一个有效工具.粒计算研究主要涉及粒化和粒的计算两个基本问题,粒的计算是指通过对粒、粒层和所有粒层组成的层次结构的分析来实现问题的求解.目前,粒计算已成为人工智能领域和大数据处理的重要方法[2].粗糙集理论[3]是一个典型的粒计算模型,它将数据以属性－对象值形式用信息系统表示,将知识对论域的划分看作是知识对论域的粒化,并通过上、下近似刻画知识的粗糙性.从粒计算的角度看,传统的粗糙集理论是从单粒度角度描述目标概念.但是,当面对多维度、多视角的现实问题时,人们往往需要从多个角度和多个粒度层次分析问题,QIAN 等[4]给出了多粒度粗糙集模型,讨论了由多个属性子集构成论域上的多粒度空间,目标决策可以通过多层粒度产生的信息粒进行刻画.为寻求最优粒度空间,许多学者对多粒度粗糙集的属性约简进行了研究,例如赵思雨等[5]针对决策表研究了多粒度粗糙集的属性约简方法,侯成军等[6]提出了基于局部可调节的多粒度粗糙集属性约简方法,郑文彬等[7]给出了基于矩阵的多粒度粗糙集粒度约简算法.针对多粒度粗糙集模型中乐观下近似的定义过于宽松、悲观下近似的定义过于严格的问题,XU 等[8]提出了广义多粒度粗糙集模型.梁美社等[9]结合最优粒度问题,提出了一种基于局部

收稿日期:2021-03-26;**修订日期**:2021-06-03;*.**通信联系人**,E-mail: shijinling126@126.com
基金项目:国家自然科学基金项目(51305128);河南省科技攻关重点项目(192102210055)
作者简介:史进玲(1982—),女,河南社旗人,讲师,硕士,主要从事粒计算、智能信息处理研究;王伟(1975—),男,河南新乡人,副教授,博士,主要从事智能信息处理、生物信息学研究.

广义多粒度粗糙集的多标记最优粒度选择方法.此外,WU 等[10]根据实际应用中对象在不同尺度下取不同值的特征,提出了多尺度信息系统的粗糙集模型,该模型引起了学者们的广泛关注,其中最优粒度问题已成为多尺度信息系统的研究热点[11-12].针对现实世界所获取信息系统的复杂性特征,很多学者对不同类型信息系统的粒度选择问题进行了深入研究,例如顾沈明[13-14]分别对多粒度决策系统、序决策系统分析了对象的局部最优粒度选择方法;吴伟志等[15]提出了广义不完备多粒度决策系统的最优粒度选择方法;刘凤玲[16]针对多决策的信息系统讨论了多粒度标记系统的最优粒度选择问题.另外,在现实问题中通常遇到多标记数据[17-18],在这样的数据中,一个对象可以同时与多个标记关联.例如一个病人可能同时患有多类疾病,如高血压、糖尿病、心脏病;一个图像可能具有多个语义,如海滩、山;一个基因可能与一组功能类相关,如代谢、转录和蛋白质合成.

多标记数据的分类是机器学习的一个主要研究领域.针对多标记数据决策表的多粒度选择问题,本文分析多标记决策表的多粒度表示,引入决策表的粒化粗糙度,讨论决策表的全局最优粒度和对象的局部最优粒度选择方法.

1　相关知识

定义 1[2]　一个决策表是一个二元组$(U, A\cup D)$,其中$U=\{x_1,x_2,\cdots,x_n\}$是一个非空有限对象集,称为论域;$A=\{a_1,a_2,\cdots,a_s\}$是一个非空有限条件属性集;$D=\{d_1,d_2,\cdots,d_t\}$是一个非空有限决策属性集;通常假设$A\cap D=\varnothing$,并且对$\forall a\in A\cup D$,在论域U上存在一个映射函数$a:U\to V_a$,称V_a为属性a的值域.在决策表中,当论域U中对象满足某些条件时,决策分类决定了决策、操作或控制应当如何进行.

对于任意属性子集$B\subseteq A$可以导出一个等价关系,也称为不可分辨关系,记R_B为:

$$R_B=\{(x,y)\in U\times U\mid \forall a\in B, a(x)=a(y)\}.$$

显然R_B在论域U中确定了一个划分,即$U/R_B=\{[x]_B\mid x\in U\}$;其中,任意划分块$[x]_B=\{y\in U\mid (x,y)\in R_B\}$,称为对象$x$关于$B$的等价类.

性质 1[3]　设B,C为论域U上的两个属性集合,则有:

$$U/(B\cup C)=U/B\cap U/C.$$

定义 2[3]　在决策表$(U, A\cup D)$中,任意属性子集$B\subseteq A$,$X\subseteq U$,记$U/B=\{B_1,B_2,\cdots,B_m\}$,$i=1,2,\cdots,m$.定义两个子集:

$$\underline{B}X=\bigcup\{B_i\mid B_i\in U/B, B_i\subseteq X\},$$

$$\overline{B}X=\bigcup\{B_i\mid B_i\in U/B, B_i\cap X\neq\varnothing\},$$

分别为X关于B的下近似集和上近似集.由此分别定义X关于B的正域、边界域及负域为:

$$\mathrm{POS}_B(X)=\underline{B}X, BN_B(X)=\overline{B}X-\underline{B}X,$$

$$\mathrm{NEG}_B(X)=U-\overline{B}X.$$

因此,$\underline{B}X$和$\mathrm{POS}_B(X)$表示了论域上根据属性集B确定属于X的对象集合;$\overline{B}X$表示了论域上可能属于X的对象集合;而边界域$BN_B(X)$表示在论域上根据属性集B无法确定属于X还是属于$U-X$的对象集合,$\mathrm{NEG}_B(X)$表示肯定不属于X的对象集合.

性质 2[3]　在决策表$(U, A\cup D)$中,若$\mathrm{POS}_C(D)=U$,则称该决策表是协调的;否则,称该决策表是不协调的.

在多标记分类中,L表示由多个标记属性组成的集合,假设$A\cap L=\varnothing$,论域中任意一个对象可能与多个标记相关联,假设任意对象x至少与标记集L中的一个标记相关联[17],若标记集L是由多个决策标记属性构成,称这样的决策表为多标记决策表.

2　多粒度多标记决策表

在传统的信息系统中,任意对象x关于a的取值是唯一的,即信息系统所反映的是单一尺度下的对象信息,但是,在现实信息系统中人们根据不同的尺度观测到对象有不同的数值,即任意对象x关于同一属性a在不同粒度层次下有不同的取值,例如在空间信息系统的遥感数据分析中,某一地表物根据观测距离或分辨率的不同,可能分别呈现出陆地、植被、庄稼地、玉米地等[2],称这样的信息系统为多尺度信息系统.

定义 3　称$S=(U, A, L)$是一个多标记决策表,其中$U=\{x_1,x_2,\cdots,x_n\}$是一个非空有限对象集合,称为论域;$A=\{a_1,a_2,\cdots,a_m\}$是一个非空有限属性集合.$L=\{l_1,l_2,\cdots,l_s\}$是由s个标记组成的非空有限标记集.在论域U上,对$\forall a\in A$,存在一个映射函数$a:U\to V_a$,称V_a为属性a的值域;假设对每个标记都看作是一个二元单粒度决策属性,对$\forall l\in L$,都存在一个映射函数$l:U\to V_l$,其中$V_l=\{0,1\}$为标记l的值域;对$\forall x\in U$,若对象x与标记

l 相关联，则有 $l(x)=1$；否则，$l(x)=0$.

假设对多标记决策表构造的粒度层数是 I，$\forall a\in A$ 均为多粒度属性，每个属性都有 I 个相同的粒度等级，在不同的粒度等级下每个对象在同一属性 a 上取不同的值；令 $k=1, 2, \cdots, I$，从细粒度向粗粒度逐层构造，记每个粒度层次下的属性集 $a^k=\{a_1^k, a_2^k, \cdots, a_m^k\}$. 因此，对于给定的粒度层数 k，得到第 k 层粒度决策表 $S_k=(U, a^k, L)$，其中 $a_j^k: U\to V_j^k$ 是满射函数，$a_j^k\in a^k$，V_j^k 是 a_j^k 在第 k 层粒度下的值域.

给定粒度层次 k，存在一个满射函数 $f_j^{k,k+1}: V_j^k\to V_j^{k+1}$，则有 $a_j^{k+1}=f_j^{k,k+1}\circ a_j^k$，即：

$$a_j^{k+1}(x)=f_j^{k,k+1}(a_j^k(x)), x\in U.$$

其中，$f_j^{k,k+1}$ 成为 k 层与 $k+1$ 层粒度之间的信息变换函数.

由于，在多粒度多标记决策表中，每个属性都有 I 个相同的粒度等级，则 $S=(U, a^k, L)$ 可以分解为 I 个决策表，即 $S=S_1\cup S_2\cup\cdots\cup S_I$.

在多粒度多标记决策表 $S=(U, a^k, L)$ 中，记 $R_{a^k}=\{(x, y)\in U\times U \mid a_j^k(x)=a_j^k(y), \forall a_j^k\in a^k, j=1, 2, \cdots, m\}$. R_{a^k} 定义了 k 层粒度下多标记决策表的不可分辨关系，给定集合 $B\subseteq A$，假设属性集 B 和 A 具有相同的粒度等级，在 k 层粒度下 B 对论域的划分 $U/R_{B^k}=\{[x]_{B^k} \mid x\in U\}$，其中 $[x]_{B^k}=\{y\in U \mid (x, y)\in R_{B^k}\}$ 表示对象 x 关于 B^k 的等价类. 对 $\forall y\in[x]_{B^k}$，显然在 k 层粒度下，x 与 y 关于 B^k 是不可分辨的，U/R_{B^k} 表示属性集 B 在 k 层粒度下对论域的粒度划分. 同样地，L 也导出一个等价关系，记 L 对论域的划分 $U/R_l=\{[x]_l \mid x\in U\}$.

定义 4　假设 $|\cdot|$ 表示集合中元素个数，为度量 k 层粒度下 B^k 对论域的粒度划分粗细程度，令 $G_R^k(B)=|U/R_{B^k}|$.

显然，$1\leqslant G_R^k(B)\leqslant|U|$. 若 $G_R^k(B)=|U|$，则说明在第 k 层粒度下，B 对论域的划分达到最细的粒度；反之，若 $G_R^k(B)=1$，则说明在 k 层粒度下，B 对论域的划分达到最粗的粒度.

定义 5　设 $S=(U, a^k, L)$ 是一个多粒度多标记决策表，定义 k 层粒度下论域上 A 关于 L 的粒化粗糙度为：

$$\rho_a^l(k)=(G_R^k(A\cup L)-G_R^k(A))/|U|.$$

性质 3　$0\leqslant\rho_a^l(k)\leqslant|U-1|/|U|$.

证明　由于 $A\cap L=\varnothing$，故 $A\subset A\cup L$. $G_R^k(A)=|U/R_{a^k}|$，$G_R^k(A\cup L)=|U/R_{A\cup L^k}|$，由性质 1 可得，$1\leqslant G_R^k(A)\leqslant G_R^k(A\cup L)$，因此有 $G_R^k(A\cup L)-G_R^k(A)\geqslant 0$. 由定义 4 得，若 a^k 对论域的划分得到最粗的粒度，有 $G_R^k(A)=1$，而 $a^k\cup L$ 对论域的划分达到最细的粒度，有 $G_R^k(A\cup L)=|U|$，此时 $G_R^k(A\cup L)-G_R^k(A)=|U|-1$，则有 $\rho_a^l(k)$ 取最大值 $|U-1|/|U|$，因此有 $0\leqslant\rho_a^l(k)\leqslant|U-1|/|U|$. 证毕.

例 1　设 $S=(U, a^k, L)$ 是一个多粒度多标记决策表，其中 $U=\{x_1, x_2, \cdots, x_{13}\}$，$a^k=\{a_1^k, a_2^k, \cdots, a_m^k\}$，$L=\{l_1, l_2, l_3\}$. 其中 $k=3$，$m=4$. 第 1 层粒度决策表 $S_1=(U, a^1, L)$ 如表 1 所示；第 2 层粒度决策表 $S_2=(U, a^2, L)$ 如表 2 所示；第 3 层粒度决策表 $S_3=(U, a^3, L)$ 如表 3 所示.

表 1　第 1 层粒度决策表

Tab. 1　Decision table of the first level granularity

U	a_1^1	a_2^1	a_3^1	a_4^1	l_1	l_2	l_3
x_1	3	2	1	1	1	0	0
x_2	2	1	1	2	1	0	1
x_3	3	2	4	5	0	1	0
x_4	5	4	3	2	0	0	1
x_5	1	3	3	4	1	1	0
x_6	1	2	2	1	1	0	1
x_7	2	5	3	4	0	1	0
x_8	2	4	5	3	0	0	0
x_9	3	4	5	3	0	0	1
x_{10}	1	3	2	4	1	1	0
x_{11}	2	5	4	5	1	0	1
x_{12}	1	2	3	1	0	1	0
x_{13}	4	4	3	5	0	0	1

表 2　第 2 层粒度决策表

Tab. 2　Decision table of the second level granularity

U	a_1^2	a_2^2	a_3^2	a_4^2	l_1	l_2	l_3
x_1	P	G	G	G	1	0	0
x_2	G	G	G	G	1	0	1
x_3	P	G	P	F	0	1	0
x_4	F	P	P	G	0	0	1
x_5	G	P	P	P	1	1	0
x_6	G	G	G	G	1	0	1
x_7	G	F	P	P	0	1	0
x_8	G	P	F	P	0	0	0
x_9	P	P	F	P	0	0	1
x_{10}	G	P	G	P	1	1	0
x_{11}	G	F	P	F	1	0	1
x_{12}	G	G	P	G	0	1	0
x_{13}	P	P	P	F	0	0	1

表 3　第 3 层粒度决策表

Tab. 3　Decision table of the third level granularity

U	a_1^3	a_2^3	a_3^3	a_4^3	l_1	l_2	l_3
x_1	N	Y	Y	Y	1	0	0
x_2	Y	Y	Y	Y	1	0	1
x_3	N	Y	N	N	0	1	0
x_4	N	N	N	Y	0	0	1
x_5	Y	N	N	N	1	1	0
x_6	Y	Y	Y	Y	1	0	1
x_7	Y	N	N	N	0	1	0
x_8	Y	N	N	N	0	0	0
x_9	N	N	N	N	0	0	1
x_{10}	Y	N	Y	N	1	1	0
x_{11}	Y	N	N	N	1	0	1
x_{12}	Y	Y	N	Y	0	1	0
x_{13}	N	N	N	N	0	0	1

在决策表 S 中，得到多标记决策属性集 L 对

论域的粒度划分为：

$U/L=\{\{x_1\},\{x_2,x_6,x_{11}\},\{x_3,x_7,x_{12}\},\{x_4,x_9,x_{13}\},\{x_5,x_{10}\},\{x_8\}\}$.

属性集 A 在各层粒度下对论域的划分分别为：

$U/a^1=\{\{x_1\},\{x_2\},\{x_3\},\{x_4\},\{x_5\},\{x_6\},\{x_7\},\{x_8\},\{x_9\},\{x_{10}\},\{x_{11}\},\{x_{12}\},\{x_{13}\}\}$,

$U/a^2=\{\{x_1\},\{x_2,x_6\},\{x_3\},\{x_4\},\{x_5\},\{x_7\},\{x_8\},\{x_9\},\{x_{10}\},\{x_{11})\},\{x_{12}\},\{x_{13}\}\}$,

$U/a^3=\{\{x_1\},\{x_2,x_6\},\{x_3\},\{x_4\},\{x_9,x_{13}\},\{x_5,x_7,x_8,x_{11}\},\{x_{10}\},\{x_{12}\}\}$.

计算得到：

$G_R^1(A)=13, G_R^2(A)=12, G_R^3(A)=8$.

定理 1　设 $S=(U, a^k, L)$是一个多粒度多标记决策表，$\rho_a^l(k)$为 k 层粒度下论域上 A 关于 L 的粒化粗糙度，决策表 S_k 是协调的当且仅当

$\rho_a^l(k)=0$.

证明　若 S_k 是协调的，由性质 2 可得

$\mathrm{POS}_{a^k}(L)=U$,

那么对 $\forall x\in U$，必有$[x]_{a^k}\subseteq[x]_l$，由定义 5 可知，$G_R^k(A)=G_R^k(A\cup L)$，故 $\rho_a^l(k)=0$；反过来，若 $\rho_a^l(k)=0$，则有 $G_R^k(A)=G_R^k(A\cup L)$，因此，对 $\forall x\in U$，必有$[x]_{a^k}\subseteq[x]_l$，因此有，$\mathrm{POS}_{a^k}(L)=U$，称决策表 S_k 是协调的. 证毕

定义 6　设 $S=(U, a^k, L)$是一个多粒度多标记决策表，若 $\rho_a^l(1)=0$，则决策表 $S_1=(U, a^1, L)$是协调的，同时称决策表 S 是协调的；否则，若 $\rho_a^l(1)\neq 0$，则决策表 $S_1=(U, a^1, L)$是不协调的，同时称决策表 S 也是不协调的.

性质 4　令 $1<i\leqslant I$，若 $S_i=(U, a^i, L)$是协调的，则 $S_{i-1}=(U, a^{i-1}, L)$也是协调的.

证明　由于 $S_i=(U, a^i, L)$是协调的，故 $\rho_a^l(i)=0$，可得 $G_R^i(A)=G_R^i(A\cup L)$，即 $[x]_{a^i}\subseteq[x]_l$；由于 $[x]_{a^{i-1}}\subseteq[x]_{a^i}$，因此 $G_R^{i-1}(A)=G_R^{i-1}(A\cup L)$，即 $\rho_a^l(i-1)=0$，则决策表 S_{i-1} 是协调的. 证毕.

由性质 4 可知，若决策表在较粗的粒度层次下是协调的，则它在较细的粒度层次下一定是协调的.

性质 5　设 $S=(U, a^k, L)$是一个多粒度多标记决策表，令 $1\leqslant k'\leqslant I-1$，则有 $\rho_a^l(k')\leqslant\rho_a^l(k'+1)$.

证明　对于 $\forall x\in U$，有$[x]_{a^{k'}}\subseteq[x]_{a^{k'+1}}$，假设 $\exists y,z\in U$，有$[y]_{a^{k'}}\cap[z]_{a^{k'}}=\varnothing$，且$[y]_{a^{k'}}\cup[z]_{a^{k'}}=[y]_{a^{k'+1}}$，令 $u\in U-\{y,z\}$，可得$[u]_{a^{k'}}=[u]_{a^{k'+1}}$. 若 $S_{k'}$是协调的，由于$[y]_{a^{k'}}\subseteq[y]_l$ 且$[z]_{a^{k'}}\subseteq[z]_l$，则有 $[y]_{a^{k'+1}}\cap[y]_l\neq[y]_{a^{k'+1}}$ 或$[y]_{a^{k'+1}}\cap[y]_l=[y]_{a^{k'+1}}$，由对象 y,z 的任意性及定义 5 可得 $\rho_a^l(k')<\rho_a^l(k'+1)$或 $\rho_a^l(k')=\rho_a^l(k'+1)$，即 $S_{k'+1}$ 不一定是协调的. 若 $S_{k'}$是不协调的，由于$[x]_{a^{k'}}\subseteq[x]_{a^{k'+1}}$，可得

$$\sum_{x\in[x]_{a^{k'}}}\left|\frac{[x]_{a^{k'}}\cap[x]_l}{[x]_{a^{k'}}}\right|\leqslant\sum_{x\in[x]_{a^{k'+1}}}\left|\frac{[x]_{a^{k'+1}}\cap[x]_l}{[x]_{a^{k'+1}}}\right|,$$

故有 $G_R^{k'}(A\cup L)-G_R^{k'}(A)\leqslant G_R^{k'+1}(A\cup L)-G_R^{k'+1}(A)$，因此有 $\rho_a^l(k')\leqslant\rho_a^l(k'+1)$. 证毕.

3　多标记决策表的全局最优粒度选择

由于在多标记决策表中，属性集 A 在不同粒度层次下对论域的粒度划分粗细程度往往是不同的，通常 k 越小，粒度划分越细；然而在实际应用中，基于最细的粒度划分进行决策标记分类的成本很高，因此，在多粒度层次中，寻求多标记决策表的全局最优粒度是多标记决策分类的一项重要工作.

由定义 5 和性质 5 容易得到定理 2.

定理 2　设 $S=(U, a^k, L)$是一个多粒度多标记决策表，令 $1\leqslant t\leqslant I-1$，若 $\rho_a^l(t)=\rho_a^l(1)$，且 $\rho_a^l(t+1)\neq\rho_a^l(t)$，则称该决策表在 t 层下，A 关于 L 对论域的划分是全局最优粒度划分.

多粒度多标记决策的全局最优粒度算法：

输入：多粒度多标记决策表 $S=(U, a^k, L)$；

输出：决策表的全局最优粒度层数.

Step 1　令 $k=1$，计算 U/L；

Step 2　计算 $G_R^k(A\cup L)$、$G_R^k(A)$、$G_R^{k+1}(A\cup L)$、$G_R^{k+1}(A)$；

Step 3　if $(\rho_a^l(k+1)\neq\rho_a^l(k))$ then go Step 6；

Step 4　$k=k+1$；

Step 5　if $(k\neq I)$, then go step2；

Step 6　输出最优粒度层数为 k；

例 2　在例 1 给出的多粒度多标记决策表中，根据全局最优粒度算法，计算 $G_R^1(A)=13$，$G_R^1(A\cup L)=13$，得到 $\rho_a^l(1)=0$；同理计算得 $\rho_a^l(2)=0$，$\rho_a^l(3)=3/13$，因此，得到该决策表的全局最优粒度层数为 2.

例 3　设 $S=(U, a^k, L)$ 是一个多粒度多标记决策表，$\forall x\in U$ 在 k 层粒度下关于属性 a 的取值如表 4 所示. 在该决策表中，$U=\{x_1, x_2, \cdots, x_{13}\}$，$k=3$，$m=4$，$a^k=\{a_1^k, a_2^k, \cdots, a_m^k\}$，$L=\{l_1, l_2, l_3\}$.

根据全局最优粒度算法，计算得：

$U/L=\{\{x_1,x_{11}\},\{x_2,x_3x_7,x_{10}\},\{x_4\},\{x_5,x_8x_{13}\},\{x_6,x_{12}\},\{x_9\}\}$.

在第 1 层粒度下，得到：

$U/a^1=\{\{x_1\},\{x_2\},\{x_3\},\{x_4\},\{x_5,x_{10}\},\{x_6\},\{x_7\},\{x_8\},\{x_9\},\{x_{11}\},\{x_{12},x_{13}\}\}$，

计算得 $\rho_a^l(1)=2/13$.

同理，在第 2 层粒度下，有：

$U/a^2=\{\{x_1\},\{x_2\},\{x_3,x_5,x_{10}\},\{x_4\},\{x_6\},\{x_7\},\{x_8,x_{12},x_{13}\},\{x_9\},\{x_{11}\}\}$，

计算得 $\rho_a^l(k)=2/13$.

在第 3 层粒度下，$U/a^3=\{\{x_1\},\{x_2\},\{x_3,x_5,x_{10}\},\{x_4\},\{x_6\},\{x_7,x_9\},\{x_8,x_{12},x_{13}\},\{x_{11}\}\}$，得到 $\rho_a^l(k)=3/13$.

由此得到表 4 所示决策表的最优粒度层数为 2.

表 4　多粒度多标记决策表

Tab. 4　Multi-granularity and multi-label decision table

U	a_1^1	a_1^2	a_1^3	a_2^1	a_2^2	a_2^3	a_3^1	a_3^2	a_3^3	a_4^1	a_4^2	a_4^3	l_1	l_2	l_3
x_1	5	F	N	2	G	Y	5	F	N	2	G	Y	1	0	1
x_2	1	G	Y	5	F	N	3	P	N	2	G	Y	1	0	0
x_3	2	G	Y	1	G	Y	3	P	N	4	P	N	1	0	0
x_4	3	P	N	3	P	N	2	G	Y	4	P	N	0	1	0
x_5	2	G	Y	1	G	Y	3	P	N	3	P	N	1	1	1
x_6	1	G	Y	2	G	Y	1	G	Y	3	P	N	1	1	0
x_7	3	P	N	4	P	N	5	F	N	4	P	N	1	0	0
x_8	2	G	Y	1	G	Y	2	G	Y	2	G	Y	1	1	1
x_9	4	P	N	5	F	N	4	P	N	3	P	N	0	0	1
x_{10}	2	G	Y	1	G	Y	3	P	N	3	P	N	1	0	0
x_{11}	1	G	Y	1	G	Y	5	F	N	1	G	Y	1	0	1
x_{12}	1	G	Y	1	G	Y	2	G	Y	2	G	Y	1	1	0
x_{13}	1	G	Y	1	G	Y	2	G	Y	2	G	Y	1	1	1

由以上分析可知，在例 1 的决策表中，$\rho_a^l(1)=0$，而在例 3 的决策表中 $\rho_a^l(1)=2/13$，根据定理 1，可知例 1 中的决策表是协调的，而例 3 中的决策表是不协调的. 通过比较发现，在以上全局最优粒度选择算法中，从不同粒度层次出发，以粒化粗糙度 $\rho_a^l(k)$ 为评价标准不仅能有效地从协调决策表中获取最优粒度，而且也能有效地从不协调决策表中获取最优粒度.

4　对象的最优粒度选择

由于以上所分析的全局最优粒度是基于论域 U 的，而不是基于单个对象的，即获得的是整个决策表的最优粒度，但是在实际中，决策表的全局最优粒度不一定是对象的最优粒度. 而且，在不一致多标记决策表中，存在某些对象关于属性集的分类对于决策标记集是不确定的，例如，在例 3 给出的决策表中对象 x_5 和 x_{10} 关于决策标记 l_2、l_3 是不确定的. 因此，为更好地进行标记分类，决策表的最优粒度选择应该综合考虑决策表的全局最优粒度和对象的局部最优粒度.

在现实决策中，对于不能确定标记分类的对象，人们通常设定一个期望阈值 $\partial_x(l_t)$，l_t 表示一个标记分类，若对象 x 与一个标记 l 的关联概率大于 $\partial_x(l_t)$，认为 x 属于该标记分类.

定义 7　假设 $S=(U,a^k,L)$ 是一个多粒度多标记决策表，$\forall x\in U$，定义对象 x 关于 a^k 的标记分类隶属度：

$$\mu_{a^k}^{l_t}(x)=\max\left\{\frac{[x]_{a^k}\cap l_t}{[x]_{a^k}},\forall l_t\in U/L\right\}.$$

显然，$0<\mu_{a^k}^{l_t}(x)\leqslant 1$. 假设在 k 层粒度下，若 $\mu_{a^k}^{l_t}(x)=1$，确定对象 x 属于 l_t 标记类；否则，若 $\mu_{a^k}^{l_t}(x)>\partial_x(l_t)$，认为对象 x 属于 l_t 标记类.

令 $1\leqslant s\leqslant I-1$，若 $\mu_{a^s}^{l_t}(x)=\mu_{a^1}^{l_t}(x)=1$，而 $\mu_{a^{s+1}}^{l_t}(x)\neq\mu_{a^1}^{l_t}(x)$，则称对象 x 在第 s 层下达到最优的粒度. 若 $\mu_{a^I}^{l_t}(x)=\mu_{a^1}^{l_t}(x)=1$，则称对象 x 在 I 层达到最优粒度；否则，若 $\mu_{a^1}^{l_t}(x)<\partial_x(l_t)$，而 $\mu_{a^s}^{l_t}(x)=\max\{\mu_{a^k}^{l_t}(x),1\leqslant k\leqslant I\}>\partial_x(l_t)$，且 $\mu_{a^{s+1}}^{l_t}(x)\neq\mu_{a^s}^{l_t}(x)$，则认为对象 x 在第 s 层达到最优的粒度.

例 4　在例 3 所示决策表中，假设 $\partial_x(l_t)=0.6$，分析每个对象的局部最优粒度.

由于 $\mu_{a^1}^{l_1}(x_1)=\mu_{a^2}^{l_2}(x_1)=\mu_{a^3}^{l_3}(x_1)=1$，$\mu_{a^1}^{l_1}(x_2)=\mu_{a^2}^{l_2}(x_2)=\mu_{a^3}^{l_3}(x_2)=1$，因此，$x_1$ 和 x_2 的最优粒度为第 3 层粒度.

同理，$\mu_{a^1}^{l_1}(x_3)=1$，$\mu_{a^2}^{l_2}(x_3)=\mu_{a^3}^{l_3}(x_3)=2/3$，可得 x_3 的最优粒度为第 1 层粒度.

$\mu_{a^1}^{l_1}(x_4)=\mu_{a^2}^{l_2}(x_4)=\mu_{a^3}^{l_3}(x_4)=1$，故 x_4 的最优粒度为第 3 层粒度.

由于 $\mu_{a^1}^{l_1}(x_5)=1/2<0.6$，$\mu_{a^2}^{l_2}(x_5)=\mu_{a^3}^{l_3}(x_5)=2/3>0.6$，则认为 x_5 的最优粒度为第 3 层粒度，确定对象 x_5 只和 l_1 关联，与 l_2、l_3 不关联.

由 $\mu_{a^1}^{l_1}(x_6)=\mu_{a^2}^{l_2}(x_6)=\mu_{a^3}^{l_3}(x_6)=1$，可得 x_6 的最优粒度为第 3 层粒度.

由 $\mu_{a^1}^{l_1}(x_7)=\mu_{a^2}^{l_2}(x_7)=1$，$\mu_{a^3}^{l_3}(x_7)=1/2$，可得 x_7 的最优粒度为第 2 层粒度.

由 $\mu_{a^1}^{l_1}(x_8)=1$，$\mu_{a^2}^{l_2}(x_8)=\mu_{a^3}^{l_3}(x_8)=2/3$，可得 x_8 的最优粒度为第 1 层粒度.

由 $\mu_{a^1}^{l_1}(x_9)=\mu_{a^2}^{l_2}(x_9)=1$，$\mu_{a^3}^{l_3}(x_9)=1/2$，可得 x_9 的最优粒度为第 2 层粒度.

$\mu_{a^1}^{l_1}(x_{10})=1/2<0.6$，$\mu_{a^2}^{l_2}(x_{10})=2/3>0.6$，认为 x_{10} 的最优粒度为第 3 层粒度，确定对象 x_{10} 只和 l_1 关联，与 l_2、l_3 不关联.

由 $\mu_{a^1}^{l_1}(x_{11})=\mu_{a^2}^{l_2}(x_{11})=\mu_{a^3}^{l_3}(x_{11})=1$，可得 x_{11}

的最优粒度为第 3 层粒度.

$$\mu_{a_1}^{l}(x_{12})=\mu_{a_1}^{l}(x_{13})=1/2,$$

$$\mu_{a_2}^{l}(x_{12})=\mu_{a_2}^{l}(x_{13})=2/3>0.6,$$

$$\mu_{a_3}^{l}(x_{12})=\mu_{a_3}^{l}(x_{13})=2/3>0.6,$$

认为 x_{12}、x_{13} 的最优粒度为第 3 层粒度,确定对象 x_{12}、x_{13} 与 l_1、l_2、l_3 关联.

通过以上分析,发现对象 x_3、x_8 的最优粒度为第 1 层粒度,对象 x_7、x_9 的最优粒度为第 2 层粒度,对象 x_1、x_2、x_4、x_5、x_6、x_{10}、x_{11}、x_{12}、x_{13} 的最优粒度为第 3 层粒度.

因此在多标记分类中,综合运用全局最优粒度选择方法和对象的局部最优粒度选择方法能更好地获取多标记数据的最优粒度,为后续多标记决策分类的研究奠定基础.

5　结论

在实际应用中,粒计算通常将复杂问题转换为多粒度问题,目前多粒度分析已成为粒计算求解复杂问题的一项重要工作.本文基于多粒度视角讨论了多标记决策表的多粒度表示,定义了不同粒度层次下条件属性集关于决策标记集的粒化粗糙度,分析了多粒度层次下协调决策表的评价指标.针对协调决策表和不协调决策表的最优粒度选择问题,提出了决策表的全局最优粒度选择方法,并讨论了对象的局部最优粒度选择方法,弥补了决策表的全局最优粒度不一定是单个对象的最优粒度问题.在后续的研究中将讨论多粒度多标记决策表的属性约简与多标记分类问题.

参考文献:

[1] BELLO M, NÁPOLES G, VANHOOF K, et al. Data quality measures based on granular computing for multi-label classification[J]. Information Sciences, 2021, 560: 51-67.

[2] 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): 501-512.
WU Weizhi. Reviews and prospects on the study of multi-granularity rough set data analysis[J]. Journal of Northwest University(Natural Science Edition), 2018, 48(4): 501-512.

[3] PAWLAK Z. Rough sets[J]. International Journal of Computer and Information Sciences, 1982, 11(5): 341-356.

[4] 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.

[5] 赵思雨, 钱婷, 魏玲. 基于决策表的多粒度粗糙集属性约简研究[J]. 陕西师范大学学报(自然科学版), 2019, 47(3): 73-78.
ZHAO Siyu, QIAN Ting, WEI Ling. The attribute reduction in MGRS based on a decision table[J]. Journal of Shaanxi Normal University(Natural Science Edition), 2019, 47(3): 73-78.

[6] 侯成军,米据生,梁美社. 基于局部可调节多粒度粗糙集的属性约简[J]. 计算机科学, 2020, 47(3): 87-91.
HOU Chengjun, MI Jusheng, LIANG Meishe. Attribute reduction based on local adjustable multi-granulation rough set[J]. Computer Science, 2020, 47(3): 87-91.

[7] 郑文彬, 李进金, 张燕兰, 等. 基于矩阵的多粒度粗糙集粒度约简方法[J]. 南京大学学报(自然科学), 2021, 57(1): 141-149.
ZHENG Wenbin, LI Jinjin, ZHANG Yanlan, et al. Matrix-based granulation reduction method for multi-granulation rough sets[J]. Journal of Nanjing University(Natural Sciences), 2021, 57(1): 141-149.

[8] XU W H, LI W T, ZHANG X T. Generalized multi-granulation rough sets and optimal granularity selection[J]. Granular Computing, 2017, 2(4): 271-288.

[9] 梁美社, 米据生, 侯成军, 等. 基于局部广义多粒度粗糙集的多标记最优粒度选择[J]. 模式识别与人工智能, 2019, 32(8): 718-725.
LIANG Meishe, MI Jusheng, HOU Chengjun, et al. Optimal granulation selection for multi-label data based on local generalized multi-granulation rough set[J]. Pattern Recognition and Artificial Intelligence, 2019, 32(8): 718-725.

[10] WU W Z, LEUNG Y. Theory and applications of granular labeled partitions in multi:Scale decision tables[J]. Information Sciences, 2011, 181(18): 3878-3897.

[11] 胡梦婷. 基于多尺度决策表的最优尺度选择方法研究[D]. 西安: 西安石油大学, 2020.
HU Mengting. Research on optimal scale selection method based on multi-scale decision table[D]. Xi'an: Xi'an Shiyou University, 2020.

[12] 刘凤玲, 林国平. 动态更新属性值变化时的最优粒度[J]. 小型微型计算机系统, 2020, 41(10): 2063-2067.
LIU Fengling, LIN Guoping. Updating optimal scale in multi-scale decision systems under environment of attribute value[J]. Journal of Chinese Computer Systems, 2020, 41(10): 2063-2067.

[13] 顾沈明, 万雅虹, 吴伟志, 等. 多粒度决策系统的局部最优粒度选择[J]. 南京大学学报(自然科学), 2016, 52(2): 280-288.
GU Shenming, WAN Yahong, WU Weizhi, et al. Local optimal granularity selections in multi-granular decision systems[J]. Journal of Nanjing University(Natural Science), 2016, 52(2): 280-288.

[14] 顾沈明, 张昊, 吴伟志, 等. 多标记序决策系统中基于局部最优粒度的规则获取[J]. 南京大学学报(自然科学), 2017, 53(6): 1012-1022.
GU Shenming, ZHANG Hao, WU Weizhi, et al. Rules acquisition based on local optimal granularities in multi-label ordered decision systems [J]. Journal of Nanjing University (Natural Science), 2017, 53(6): 1012-1022.

[15] 吴伟志, 杨丽, 谭安辉, 等. 广义不完备多粒度标记决策系统的粒度选择[J]. 计算机研究与发展, 2018, 55(6): 1263-1272.
WU Weizhi, YANG Li, TAN Anhui, et al. Granularity selections in generalized incomplete Multi-Granular labeled decision systems[J]. Journal of Computer Research and Development, 2018, 55(6): 1263-1272.

[16] 刘凤玲, 林国平, 余晓龙. 多决策的多粒度标记系统的最优粒度选择[J]. 重庆理工大学学报(自然科学版), 2020, 34(5): 263-270.
LIU Fengling, LIN Guoping, YU Xiaolong. Optimal granularity selection in Multi-Granular labeled systems with multiple decision[J]. Journal of Chongqing Institute of Technology, 2020, 34(5): 263-270.

[17] LI H, LI D Y, ZHAI Y H, et al. A novel attribute reduction approach for multi-label data based on rough set theory[J]. Information Sciences, 2016, 367/368: 827-847.

[18] BOUTELL M R, LUO J B, SHEN X P, et al. Learning multi-label scene classification[J]. Pattern Recognition, 2004, 37(9): 1757-1771.

责任编辑:郭红建