计算机科学
COMPUTER SCIENCE

介粒度空间中的最优粒度选择和属性约简

李腾, 李德玉, 翟岩慧, 张少霞

引用本文

李腾, 李德玉, 翟岩慧, 张少霞. 介粒度空间中的最优粒度选择和属性约简[J]. 计算机科学, 2023, 50(10): 71-79.

LI Teng, LI Deyu, ZHAI Yanhui, ZHANG Shaoxia. Optimal Granularity Selection and Attribute Reduction in Meso-granularity Space [J]. Computer Science, 2023, 50(10): 71-79.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230500218

介粒度空间中的最优粒度选择和属性约简

李 腾[1] 李德玉[1,2] 翟岩慧[1,2] 张少霞[3]
1 山西大学计算机与信息技术学院 太原 030006
2 计算智能与中文信息处理教育部重点实验室(山西大学) 太原 030006
3 山西财经大学信息学院 太原 030006
(lt596018984@163.com)

摘 要 以往的形式概念分析采用介粒度形式背景,满足对数据跨层粒化的需求,但其既没有将寻找最优粒度和属性约简有效结合起来,又没有在多粒度的背景下高效地解决组合爆炸问题。为此,基于介粒度中粒度选择和属性约简的联系,提出了一种新的最优粒度选择方式——最优粒度约简,以同步进行粒度选择和属性约简。鉴于寻找最优粒度约简存在组合爆炸的问题,设计了逐步搜索方法,通过已搜索的信息更新粒度空间,去除大量非最优粒度约简,显著提高了搜索效率。实验结果表明了所提方法的有效性和优势。

关键词:形式概念分析;多粒度决策形式背景;最优粒度;属性约简;粒计算
中图法分类号 TP182

Optimal Granularity Selection and Attribute Reduction in Meso-granularity Space

LI Teng[1], LI Deyu[1,2], ZHAI Yanhui[1,2] and ZHANG Shaoxia[3]
1 School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China
2 Key Laboratory of Computational Intelligence and Chinese Information Processing (Shanxi University), Ministry of Education, Taiyuan 030006, China
3 College of Information, Shanxi University of Finance and Economics, Taiyuan 030006, China

Abstract The conventional formal concept analysis adopts a meso-granularity formal context to meet the requirements of cross-layer granulation of data. However, it does not effectively combine the search for optimal granularity with attribute reduction, nor does it efficiently solve the problem of combination explosion in a multi-granular context. Therefore, based on the connection between granularity selection and attribute reduction in the meso-granularity, a new optimal granularity selection method (i.e., optimal granularity reduction) is proposed to synchronize the selection of the optimal granularity and attribute reduction. In view of the combination explosion in searching for optimal granularity reduction, a stepwise search method is designed to update the granularity space with searched information, eliminating a large number of non-optimal granularity reduction and significantly improving search efficiency. Experimental results demonstrate the effectiveness and superiority of this method.

Keywords Formal concept analysis, Multi-granularity decision formal context, Optimal granularity, Attribute reduction, Granular computing

1 引言

Willie[1]通过形式化方式描述了概念组成的对象、属性和它们之间的从属结构关系等信息,提出了形式概念分析,该方法成为了重要的智能系统研究方法。随着相关研究的深入,学者们将粒计算[2-5]、三支决策[6-8]以及多粒度[9-11]等理论与形式概念分析紧密结合,并将其广泛应用于多粒度认知[12]、知识表示和发现[13-15]、不确定性信息处理[16]等前沿领域。

为了处理现实中的多粒化问题,一些学者从多维度、多层次的角度研究了形式概念分析,受粒计算“多粒度”思想的启发,并置多个单粒度形式背景,提出了多粒度标记形式背景[17]。但这样依然不能满足形式背景对跨层粒度组合的需求。为此,Li 等[18]通过对粒度属性的分层和分块,提出了介粒度形式背景和广义介粒度形式背景,丰富了粒度组合方式。

考虑多粒度后,在处理实际复杂问题时,数据中对象特征的表达存在过多冗余,这会使得相关工作变得低效,甚至产生

到稿日期:2023-05-29 返修日期:2023-07-27
基金项目:国家自然科学基金(62072294,61972238);山西省基础研究计划资助项目(202103021223303)
This work was supported by the National Natural Science Foundation of China(62072294,61972238) and Fundamental Research Program of Shanxi Province(202103021223303).
通信作者:李德玉(lidysxu@163.com)

错误结果。因此多粒度形式背景的决策代价至关重要，其中最优粒度选择和属性约简作为重要因素，已受到越来越多学者的关注。Wu 等[19] 基于信息粒协调结构关系提出了粒约简和粒规则的概念。Hao 等[20] 则讨论了基于粒标记规则的多标记背景下的最优标记选择。Li 等[21-22] 在多粒度决策形式背景中引入信息熵，讨论了最优粒度选择和属性约简。

但是，以往的研究或者只关注寻找最优粒度组合，但在多粒度的背景下不能高效地解决组合爆炸问题，或者只关注属性约简，没有将寻找最优粒度和属性约简有效地结合起来。因此，本文定义了一种新的粒度组合方式，以同时实现最优粒度选择和属性约简。在此基础上，提出了一种逐步搜索方法，该方法通过已搜索的信息识别并过滤非最优粒度约简，逐步缩小搜索范围，从而显著减少单粒度决策形式背景所需的协调性检查次数，提高了搜索效率。经过实验证明，逐步搜索最优粒度约简的方法是有效的。

2　基本知识

定义 1[1]　一个信息系统可以表示为二元组 $C=(G,A)$，其中 $G=\{g_1,g_2,\cdots,g_n\}$ 是非空有限对象集，$A=\{a_1,a_2,\cdots,a_m\}$ 是非空有限属性集，且任意对象 $x\in G$ 在每个属性 $a\in A$ 下均有唯一取值，记为 $a(x)$。在此基础上，一个决策系统亦为二元组 $(G,A\cup\{d\})$，其中 (G,A) 是信息系统，A 是条件属性集，$d\notin A$ 是决策属性，不妨记任意对象 $x\in G$ 在属性 d 下的取值为 $d(x)$。

定义 2[2]　三元组 $Z=(G,A,I)$ 是形式背景，其中 $G=\{g_1,g_2,\cdots,g_p\}$ 为对象集，每个 $g_i(i\leqslant p)$ 称为一个对象，$A=\{a_1,a_2,\cdots,a_q\}$ 为属性集，每个 $a_j(j\leqslant q)$ 称为一个属性，I 为 G 与 M 之间的二元关系，$I\subseteq G\times A$。若 $(g,a)\in I$，则表示对象 g 拥有属性 a，记为 gIa。为了使形式背景与布尔矩阵对应，将 $(g,a)\in I$ 记为 1，将 $(g,a)\notin I$ 记为 0，其中 Z 也表示为一个单粒度形式背景。

对任意的 $X\subseteq G,B\subseteq A$，Wille[1] 给出了两者之间的一对诱导算子，表达式如下：

$$X^*=\{a\mid a\in A,\forall g\in X,gIa\}$$

$$B^*=\{g\mid g\in G,\forall a\in A,gIa\}$$

其中，X^* 表示 X 中所有对象共同具有的属性集合，B^* 表示拥有 B 中所有属性的对象集合。这对诱导算子的具体性质见文献[1]。

定义 3[1]　设 $Z=(G,A,I)$ 为形式背景，$N\subseteq A$，$I_N=I\cap(G\times N)$，称 (G,N,I_N) 为 (G,A,I) 的属性子背景。为了与形式背景 (G,A,I) 的情形进行区别，子背景 (G,N,I_N) 上的概念诱导算子如下：

$$X^{*N}=\{a\in N:\forall x\in X,(x,a)\in I_N\}$$

$$B^{*N}=\{x\in G:\forall a\in B,(x,a)\in I_N\}$$

定义 4[23]　设属性不相交的两个形式背景 (G,A,I) 和 (G,D,J)，称五元组 $F=(G,A,I,D,J)$ 为决策形式背景，其中 A 为条件属性集，D 为决策属性集，F 也表示为一个单粒度决策形式背景。为了避免混淆，不妨用 $(\cdot)^{*A}$ 与 $(\cdot)^{*D}$ 表示算子 $(\cdot)^*$ 作用于条件子背景 (G,A,I) 和决策子背景 (G,D,J)。

例 1　表 1 列出了一个信息系统 $C=(G,A)$，其中 $G=\{g_1,g_2,g_3,g_4,g_5,g_6,g_7,g_8\}$，表示某高校的 8 名本科学生，$A=\{a_1,a_2,a_3\}$，$a_1$ 表示学生数学科目绩点，a_2 表示大学生英语六级成绩(CET-6)，a_3 表示学生获得奖学金的情况，d 表示全国大学生数学建模竞赛是否获奖。数字 1 表示该学生符合所在列的条件，0 表示不符合。

表 1　信息系统 $C=(G,A)$

Table 1　Information system $C=(G,A)$

G	a_1	a_2	a_3	d
g_1	4.1	550	二等奖	1
g_2	4.5	512	二等奖	1
g_3	3.6	426	二等奖	0
g_4	3.3	387	—	0
g_5	3.2	409	—	0
g_6	2.8	400	—	0
g_7	2.5	403	三等奖	0
g_8	3.8	391	三等奖	0

定义 5[19]　设 $F=(G,A,I,D,J)$ 为决策形式背景，若对任意的 $g\in G$，都有 $g^{*A*A}\subseteq g^{*D*D}$，则称 F 为协调决策形式背景。

定义 6[19]　设 $F=(G,A,I,D,J)$ 为协调决策形式背景且 $B\subseteq A$，若对任意的 $g\in G$，都有 $g^{*B*B}\subseteq g^{*D*D}$，则称 B 为 F 的协调集。进一步，若对任意 $E\subset B$，$\exists g\in G$，使得 $g^{*E*E}\not\subset g^{*D*D}$，则称 B 为 F 的约简集。

3　多粒度决策形式背景的最优粒度约简

本章基于粒度属性的分层和分块，从类属性块组合的角度在多粒度决策形式背景上定义介粒度决策形式背景，提出了一种新的粒度空间。最优粒度在粗糙集中被广泛应用[5-8]，参考多粒度形式背景的最优粒度[21] 和属性约简[22]，提出了介粒度决策形式背景的最优粒度约简。

定义 7[18]　设 $Z_1=(G,A,I)$ 和 $Z_2=(G,B,J)$ 为两个形式背景。对于 $c\in A$，若存在 $M=\{a_1,a_2,\cdots,a_m\}\subseteq B$，则满足 $\bigcup_{i=1}^{m}a_i^*=c^*$，且对于 M 中任意不同的 a_i 和 a_j，$a_i^*\cap a_j^*=\varnothing$，则称 c 可由 M 非交融合得到。若 Z_1 中任意的 $c_i\in A(i=1,2,\cdots,|A|)$ 均可由 $M_i\subseteq B$ 非交融合得到，且满足 $\bigcup M_i=B$，对于任意不同的 M_i 和 M_j，$M_i\cap M_j=\varnothing$，则记 $Z_1\leqslant Z_2$，并称 Z_1 比 Z_2 的粒度粗。

非交融合可以解释为将形式背景不相交的若干属性合并，产生新的属性。这种合并方式明显能够将一个形式背景转化为另一个粒度更粗的形式背景。

定义 8　设 $S^e=(G,A^e,I^e,D,J)(e=1,2,\cdots,n)$ 为 n 个决策形式背景。若对任意的 $1\leqslant i<n$，都有 $(G,A^i,I^i)\leqslant(G,A^{i+1},I^{i+1})$，且 $A^i\neq A^{i+1}$，则记 $S^i\prec S^{i+1}$，并称 $S^e=(G,A^e,I^e,D,J)(e=1,2,\cdots,n)$ 为多粒度决策形式背景。

显然，从 $S^n=(G,A^n,I^n,D,J)$ 出发，可以得到一系列非交融合的决策形式背景 (G,A^e,I^e,D,J)。非交融合的具体形式取决于属性融合时所采用的特定融合模式。此外，属性融合的泛化过程是不可逆的，在此过程中会导致信息丢失。

若 S^n 为协调决策形式背景，则称 $S^e=(G,A^e,I^e,D,J)(e=1,2,\cdots,n)$ 为协调多粒度决策形式背景。本文假设多粒度决策形式背景均是协调的。

例 2　表 1 所列的信息系统可以通过梯级转化为决策形式背景(G,A^1,I^1,D,J)，具体如表 2 所列。具体地，将信息系统中的属性a_1分成 2 个子属性b_1和b_2，其中b_1为数学科目绩点 3.5 分及以上，b_2为数学科目绩点 3.5 以下；将多值属性a_2分成两个子属性b_3和b_4，b_3为英语六级成绩 425 分及以上，b_4为英语六级成绩 425 分以下；将a_3分成两个子属性b_5和b_6，b_5为获奖学金，b_6为未获奖学金。$D=\{d\}$为全国大学生数学建模竞赛是否获奖。在该决策形式背景中，学生的属性值表示为 0 或 1 的布尔值。数字 1 表示该学生符合所在列的条件，0 表示不符合。

表 2　决策形式背景(G,A^1,I^1,D,J)

Table 2　Formal decision context(G,A^1,I^1,D,J)

G	b_1	b_2	b_3	b_4	b_5	b_6	d
g_1	1	0	1	0	1	0	1
g_2	1	0	1	0	1	0	1
g_3	1	0	1	0	1	0	0
g_4	0	1	0	1	0	1	0
g_5	0	1	0	1	0	1	0
g_6	0	1	0	1	0	1	0
g_7	0	1	0	1	1	0	0
g_8	1	0	0	1	1	0	0

将表 2 所列的决策形式背景进一步通过梯级转化形成(G,A^2,I^2,D,J)，具体如表 3 所列。具体地，将属性b_1分为两个子属性c_1和c_2，c_1为数学科目绩点 4.0 分及以上，c_2为数学科目绩点介于 3.5～3.9 分之间；将属性b_3也分为两个子属性c_4和c_5，c_4为英语六级成绩 500 分及以上，c_5为英语六级成绩介于 425～499 分；将属性b_5也分为两个子属性c_7和c_8，c_7为二等奖学金，c_8为三等奖学金；c_3，c_6，c_9分别与b_2，b_4，b_6保持一致。

表 3　决策形式背景(G,A^2,I^2,D,J)

Table 3　Formal decision context(G,A^2,I^2,D,J)

G	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	d
g_1	1	0	0	1	0	0	1	0	0	1
g_2	1	0	0	1	0	0	1	0	0	1
g_3	0	1	0	0	1	0	1	0	0	0
g_4	0	0	1	0	0	1	0	0	1	0
g_5	0	0	1	0	0	1	0	0	1	0
g_6	0	0	1	0	0	1	0	0	1	0
g_7	0	0	1	0	0	1	0	1	0	0
g_8	0	1	0	0	0	1	0	1	0	0

此时，b_1,b_2,b_3,b_4,b_5,b_6可分别由$M_1=\{c_1,c_2\}$，$M_2=\{c_3\}$，$M_3=\{c_4,c_5\}$，$M_4=\{c_6\}$，$M_5=\{c_7,c_8\}$，$M_6=\{c_9\}$非交融合得到，且有$\bigcup_{i=1}^{6}M_i=A^2$，对于任意不同的M_i和M_j，$M_i\cap M_j=\varnothing$，即(G,A^2,I^2)可以通过非交融合产生(G,A^1,I^1)。换言之，表 2、表 3 所列的决策形式背景可以联合产生多粒度决策形式背景$S^e=(G,A^e,I^e,D,J)(e=1,2)$。

定义 9[17]　设$Z=(G,A,I)$为形式背景，对于$M=\{a_1,a_2,\cdots,a_m\}\subseteq A$，若满足$\bigcup_{i=1}^{m}a_i^*=G$，且对于$M$中任意不同的$a_i$和$a_j$，$a_i^*\cap a_j^*=\varnothing$，则称$M$为$(G,A,I)$的一个类属性块。

上述定义指出，类属性块M中的属性所拥有的对象集可以构成论域G的划分。如果形式背景中的属性集A可以被划分成多个类属性块，则称A是可被类属性块划分的。

定义 10[17]　设$S^e=(G,A^e,I^e,D,J)(e=1,2,\cdots,n)$为多粒度决策形式背景。若$S^e$的属性集$A^e$均可划分为$s$个非空类属性块$M_1^e,M_2^e,\cdots,M_s^e$，且对任意的$1\leqslant i<n$，都有$M_t^i(t=1,2,\cdots,s)$的属性可由$M_t^{i+1}$非交融合产生，则称$S^e$为可类属性块划分的多粒度决策形式背景。

注意，后文中的多粒度决策形式背景均可被类属性块划分。

过去的研究先把信息系统转换成多个单粒度形式背景，再并置且添加决策属性，得到多粒度决策形式背景[17]，导致后续相关问题的探讨主要集中于这些单粒度形式背景的知识发现及其相互关系，不利于挖掘多粒度决策形式背景的多层次知识。文献[18]在多粒度决策形式背景中研究了介粒度决策形式背景，避免了过去研究的局限性，使多粒度决策形式背景的多层次知识能够被更全面地挖掘和理解。下面基于类属性块组合重新定义介粒度决策形式背景，并给出了相应的语义解释。

定义 11　设$S^e=(G,A^e,I^e,D,J)(e=1,2,\cdots,n)$为多粒度决策形式背景，$A^e$的非空类属性块记为$M_1^e,M_2^e,\cdots,M_s^e$。令$l_i\in\{1,2,\cdots,n\}$，表示选择第$l_i$个类属性块$M_i^{l_i}$，记$K=(l_1,l_2,\cdots,l_s)$为粒度组合，称$S^K=(G,A^K,I^K,D,J)$为$S^e$产生的介粒度决策形式背景，其中$A^K=\bigcup M_i^{l_i}$，$I^K=\bigcup((G\times M_i^{l_i})\cap I^{l_i})$。对于粒度组合$K=(l_1,l_2,\cdots,l_s)$和$K'=(l'_1,l'_2,\cdots,l'_s)$，若对于任意的$i\in\{1,2,\cdots,s\}$，都有$l'_i\leqslant l_i$，则记为$K'\leqslant K$，$S^{K'}\leqslant S^K$。进一步，若$K'\neq K$，则记为$K'\prec K$，$S^{K'}\prec S^K$。$S^K$中所有粒度级别组合称为$S^K$的粒度空间，记为$L=\{(l_1,l_2,\cdots,l_s)\mid 1\leqslant l_i\leqslant n,i=1,2,\cdots,s\}$。

显然，当$K=(e,e,\cdots,e)(e=1,2,\cdots,n)$时，$K$对应的介粒度决策形式背景为$S^e$中的多粒度决策形式背景。为了叙述方便，本文用$|\cdot|$表示集合的基数，显然有$|L|=n^s$。

定义 12　设$S^e=(G,A^e,I^e,D,J)(e=1,2,\cdots,n)$为多粒度决策形式背景，$S^K$为$S^e$产生的介粒度决策形式背景。若$S^K$为协调决策形式背景，且对于任意的$K'\prec K$，$S^{K'}$不协调，则称$S^K$为最优决策形式背景，$K$为最优粒度组合。

例 3　对于表 2 和表 3 产生的多粒度决策形式背景$S^e=(G,A^e,I^e,D,J)(e=1,2)$，记$M_1^1=\{b_1,b_2\}$，$M_2^1=\{b_3,b_4\}$，$M_3^1=\{b_5,b_6\}$，容易验证$A^1$可划分为非空类属性块$M_1^1$，$M_2^1$，$M_3^1$。类似地，$A^2$可划分为$M_1^2=\{c_1,c_2,c_3\}$，$M_2^2=\{c_4,c_5,c_6\}$，$M_3^2=\{c_7,c_8,c_9\}$。根据例 2，$M_t^1(t=1,2,3)$可由$M_t^2$非交融合产生，即$S^e=(G,A^e,I^e,D,J)(e=1,2)$为可类属性块划分的多粒度决策形式背景。

令$K=(1,2,1)$，即分别选择M_1^1，M_2^2，M_3^1类属性块，由此产生的介粒度决策形式背景$S^K=(G,A^K,I^K,D,J)$，具体如表 4 所列。

由定义 5 容易验证，粒度组合(2,2,2)，(1,2,2)，(2,1,2)，(2,2,1)，(1,1,2)，(2,1,1)对应的介粒度决策形式背景都是协调的。(1,1,2)，(1,1,1)对应的介粒度决策形式背景都是不协调的。由定义 12 可知，最优粒度组合为(1,2,1)和(2,1,1)。

表 4 介粒度决策形式背景(G,A^K,I^K,D,J)

Table 4 Meso-granularity formal decision context(G,A^K,I^K,D,J)

G	b_1	b_2	c_4	c_5	c_6	b_5	b_6	d
g_1	1	0	1	0	0	1	0	1
g_2	1	0	1	0	0	1	0	1
g_3	1	0	0	1	0	1	0	0
g_4	0	1	0	0	1	0	1	0
g_5	0	1	0	0	1	0	1	0
g_6	0	1	0	0	1	0	1	0
g_7	0	1	0	0	1	1	0	0
g_8	1	0	0	0	1	1	0	0

在寻找最优粒度组合时，移除某些类属性块也可能保持介粒度决策形式背景的协调性，这表明最优粒度选择和属性约简存在一定的联系。下文定义了一种偏序关系下的介粒度空间，可以同时进行最优粒度选择和属性约简。

定义 13 设$S^e=(G,A^e,I^e,D,J)(e=1,2,\cdots,n)$为多粒度决策形式背景。令$l_i\in\{0,1,\cdots,n\}$表示选择类属性块$M_i^{l_i}$，其中若$l_i=0$，则记$M_i^0=\varnothing$。对于粒度组合$K=(l_1,l_2,\cdots,l_s)$，称$S^K=(G,A^K,I^K,D,J)$为$S^e$产生的扩展介粒度决策形式背景，其中$A^K=\bigcup M_i^{l_i}$，$I^K=\bigcup((G\times M_i^{l_i})\cap I^{l_i})$。记$\varphi=\{(l_1,l_2,\cdots,l_s)\mid 0\leqslant l_i\leqslant n,i=1,2,\cdots,s\}$为介粒度空间。

定义 14 设$S^e=(G,A^e,I^e,D,J)(e=1,2,\cdots,n)$为多粒度决策形式背景，$\varphi$为$S^e$的介粒度空间。对于$K^*\in\varphi$，若$S^{K^*}$为协调决策形式背景，且对于$\varphi$中任意的$K'\prec K^*$，$S^{K'}$不协调，则称$S^{K^*}$为$\varphi$中的最优决策形式背景，$K^*$为$\varphi$中的最优粒度约简。

例 4 例 2 的多粒度决策形式背景$S^e=(G,A^e,I^e,D,J)(e=1,2)$采用介粒度空间后，粒度组合从 8 个增加到了 27 个，扩大了决策形式背景粒度选择的范围。由例 3 可知，最优粒度组合为$K=(1,2,1)$，移除第一个和第三个类属性块得到$K'=(0,2,0)$，容易验证S^K也是协调性的。进一步可以验证，$(0,2,0)$为最优粒度约简。

显然，定义 11 中的粒度空间$L\subset\varphi$，$|\varphi|=(n+1)^s$，并且$\langle\varphi,\geqslant\rangle$是一个有界格。对于$\Phi(S)=\{S^K\mid K\in\varphi\}$，$\Phi(S)$和$\varphi$的元素之间存在一对一的对应关系。

下面给出有界格φ中的一些记号和结论。

定义 15 设$S^e=(G,A^e,I^e,D,J)(e=1,2,\cdots,n)$为多粒度决策形式背景，$\varphi$为$S^e$的介粒度空间。对于$\varphi'\subseteq\varphi$，$K'\in\varphi'$，$K'$在$\varphi'$中的下界和上界为：

$$LB_{\varphi'}(K)=\{K\mid K'\in\varphi'\wedge K'\prec K\}$$

$$UB_{\varphi'}(K)=\{K\mid K'\in\varphi'\wedge K'\succ K\}$$

其中，$LB_{\varphi'}(K)$表示φ'中比K严格粗的粒度组合集合，$UB_{\varphi'}(K)$表示φ'中比K严格细的粒度组合集合。

对于粒度组合$K,K'\in\varphi$，若$K\leqslant K'$或者$K\geqslant K'$，则称K和K'有偏序关系。显然φ中任意两个粒度组合不一定有偏序关系。此时，对于$K=(l_1,l_2,\cdots,l_s)$，$K'=(l'_1,l'_2,\cdots,l'_s)$，$K$和$K'$没有偏序关系当且仅当存在$u,v\in\{1,2,\cdots,s\}$，满足$l_u<l'_u$，$l_v>l'_v$。

定义 16 设$S^e=(G,A^e,I^e,D,J)(e=1,2,\cdots,n)$为多粒度决策形式背景，$\varphi$为$S^e$的介粒度空间。对于$\varphi'\subseteq\varphi$，$\varphi'$中的最大元素集合记为$MAX(\varphi')=\{K\mid K\in\varphi'\wedge\exists K'(K'\in\varphi'\wedge K'\succ K)\}$，$\varphi'$中的最大协调元素集合记为$CMAX(\varphi')=\{K\mid K\in MAX(\varphi')\wedge S^K$是协调的$\}$，$\varphi'$中的最大不协调元素集合记为$IMAX(\varphi')=\{K\mid K\in MAX(\varphi')\wedge S^K$是不协调的$\}$。

命题 1 设$S^e=(G,A^e,I^e,D,J)(e=1,2,\cdots,n)$为多粒度决策形式背景，$\varphi$为$S^e$的介粒度空间。对于$K',K\in\varphi$，$K'\leqslant K$。

1)若$S^{K'}$是协调的，则S^K也是协调的。

2)若S^K是不协调的，则$S^{K'}$也是不协调的。

证明：为了方便，用$(\cdot)^{*K}$与$(\cdot)^{*K'}$分别表示$(\cdot)^{*A^K}$与$(\cdot)^{*A^{K'}}$。

1)由$S^{K'}$是协调的可知，对于任意的$g\in G$，有$g^{*K'*K'}\subseteq g^{*D*D}$。由$K'\leqslant K$可知，$(G,A^{K'},I^{K'})\leqslant(G,A^K,I^K)$，$A^{K'}$中的属性可由$A^K$非交融合产生。特别地，对于$e_K\in g^{*K}$，存在$M=\{a_1,a_2,\cdots,a_m\}\subseteq A^K$满足$e_K\in M$，且有$e_K^{*K}\subseteq\bigcup_{i=1}^{m}a_i^{*K}=e_{K'}{}^{*K'}$，即$e_{K'}$由$M$非交融合产生。由$e_K\in g^{*K}$可知，$g\in e_K^{*K}\subseteq e_{K'}^{*K'}$，因此$g\in e_{K'}^{*K'}$，即$e_{K'}\in g^{*K'}$。故有$g^{*K*K}=(\bigcup_{e_K\in g^{*K}}e_K)^{*K}=\bigcap_{e_K\in g^{*K}}e_K^{*K}\subseteq\bigcap_{e_{K'}\in g^{*K'}}e_{K'}^{*K'}=(\bigcup_{e_{K'}\in g^{*K'}}e_{K'})^{*K'}=(g^{*K'})^{*K'}=g^{*K'*K'}\subseteq g^{*D*D}$，即$g^{*K*K}\subseteq g^{*D*D}$。因此，$S^K$是协调的。

2)类似 1)可证。

4 逐步搜索最优粒度约简的方法

在多粒度决策形式背景中，寻找最优粒度约简，可以使多粒度决策形式背景相比过去的单粒度决策形式背景场景和需求的适用性均有所提高。决策形式背景从用单粒度表示到用多粒度表示，适用的场景、需求、问题范围都会有所增加。然而，其难点在于存在组合爆炸问题，判断决策形式背景协调性的复杂度呈现指数级增长，达到了$O((n+1)^s)$。为了提高搜索效率，本文采用减少协调性检查次数的方法，高效寻找所有最优粒度约简。具体来说，从介粒度空间φ出发，首先找到一个最优粒度约简，然后利用粒度组合之间的偏序关系识别并过滤与该最优粒度约简相关的非最优粒度约简，接着在剩余的粒度组合中进行下一轮搜索，直到找到所有最优粒度约简。基于上述思想，本章提出了逐步搜索最优粒度约简的方法，讨论了相关性质，设计了相应的算法。

首先给出在介粒度空间φ中找出一个最优粒度约简的算法，如算法 1 所示。

算法 1 搜索最优粒度约简

输入：多粒度决策形式背景$S^e=(G,A^e,I^e,D,J)(e=1,2,\cdots,n)$，$S^e$的介粒度空间$\varphi$，初始粒度组合$K_0=(n,n,\cdots,n)$，向量$\mathbf{K}_0$的长度 s

输出：最优粒度约简 **K**

```
K=K0
For j=1 to s do
  K[j]=0
  While K[j]<n do
    If S^K不协调 then
      K[j]=K[j]+1
    Else
      Break
    End if
  End while
End for
Return K
```

算法 1 从初始粒度组合$\mathbf{K}_0=(n,n,\cdots,n)$开始，在不破坏介粒度决策形式背景的协调性的前提下，在 φ 内依次寻找每个类属性块可以选择的最粗粒度，得到一个最优粒度约简。

算法 1 中的第 4－10 行，需要进行最多 n 次协调性检查，第 2－11 行需要遍历 s 次。因此，算法 1 协调性检查的时间复杂度为 $O(ns)$。

例 5　对于例 2 的多粒度决策形式背景$S^e=(G,A^e,I^e,D,J)(e=1,2)$，$\varphi$ 为S^e的介粒度空间。调用算法 1 得到 φ 中的一个最优粒度约简的过程如下。

初始组合为 $K=K_0=(2,2,2)$。开始搜索第一个分量，令 $K[1]=0$。由于 $K=(0,2,2)$对应的S^K是协调的，因此从$(0,2,2)$开始搜索第二个分量。令 $K[2]=0$，对于 $K=(0,0,2)$，由于S^K是不协调的，因此需要验证 $K=(0,1,2)$。由于$(0,1,2)$对应的S^K也是不协调的，因此 $K=(0,2,2)$。此时第二个分量无法继续增加，因此从$(0,2,2)$开始搜索第三个分量。类似地，$K=(0,2,0)$对应的S^K也是协调的，因此输出$K=(0,2,0)$。

容易看出，与最优粒度约简 $K=(0,2,0)$有偏序关系的粒度组合都是非最优粒度约简。例如$K_1=(0,1,0)\prec K$，容易验证S^{K_1}是不协调的，因此为非最优粒度约简；$K_2=(1,2,1)\succ K$，容易验证S^{K_2}是协调的，但由$K_2\succ K$ 和S^K是协调的可知，K_2是非最优粒度约简。

此外，在算法 1 的运行过程中，$K'=(0,1,2)$对应的$S^{K'}$是不协调，容易验证 $K''\in LB_\varphi(K')$也都是非最优粒度约简。例如，$K_3=(0,1,1)\prec K$，容易验证S^{K_3}是不协调的。

因此，为了找出多个最优粒度约简，可以利用粒度组合之间的偏序关系识别并过滤部分非最优粒度约简，从而缩小搜索空间。

命题 2　设$S^e=(G,A^e,I^e,D,J)(e=1,2,\cdots,n)$为多粒度决策形式背景，$\varphi$ 为S^e的介粒度空间，对于$K'\in\varphi$，有：

1）若 K' 是 φ 的最优粒度约简，则 $K\in(UB_\varphi(K')\cup LB_\varphi(K'))$不是 φ 中的最优粒度约简。

2）若$S^{K'}$是不协调的，则 $K\in LB_\varphi(K')$不是 φ 中的最优粒度约简。

证明：1）由K'是 φ 中的最优粒度约简可知，$S^{K'}$是协调的。当 $K\in UB_\varphi(K')$时，由$K'\prec K$ 可知，$S^{K'}$是协调的，由定义 14 可知，K 不是 φ 中的最优粒度约简；当 $K\in LB_\varphi(K')$时，由K'是 φ 的最优粒度约简和 $K\prec K'$可知，S^K是不协调的，故 K 不是 φ 中的最优粒度约简。

2）若$S^{K'}$是不协调的，由命题 1 可知，$K\in LB_\varphi(K')$对应的S^K是不协调的，故 K 不是 φ 中的最优粒度约简。

命题 2 表明，若在 φ 中找到一个最优粒度约简K'，则可以识别出一些非最优粒度约简。下面基于这一思想给出逐步搜索最优粒度约简的方法。

设$S^e=(G,A^e,I^e,D,J)(e=1,2,\cdots,n)$为多粒度决策形式背景，$\varphi$ 为S^e的介粒度空间。逐步搜索方法从 φ 开始搜索一个最优粒度约简，同时删除搜索过程中识别出的非最优粒度约简，剩余的粒度组合作为下一轮迭代搜索的初始空间，循环迭代到剩余的粒度空间为空时结束。

具体地，第 i 轮迭代的搜索空间记为$\varphi_i=BND(\varphi_{i-1})$，其中$\varphi_1=\varphi$。第 i 轮迭代得到的最优粒度约简记为K_i^*，第 i 轮迭代可识别的非最优粒度约简记为 $NEG(\varphi_i)$，第 i 轮迭代剩余的粒度空间记为 $BND(\varphi_i)$。其中，$NEG(\varphi_i)$由两部分组成，一部分是K_i^* 根据命题 2 中 1）确定的非最优粒度约简 $UB_\varphi(K_i^*)\cup LB_\varphi(K_i^*)$，另一部分是搜索过程中发现的最大不协调元素集合 $ISC(\varphi_i)$，以及根据命题 2 中 2）确定的非最优粒度约简 $\bigcup_{K\in ISC(\varphi_i)}LB_\varphi(K)$，即 $NEG(\varphi_i)=UB_\varphi(K_i^*)\cup LB_\varphi(K_i^*)\cup ISC(\varphi_i)\cup\bigcup_{K\in ISC(\varphi_i)}LB_\varphi(K)$，$BND(\varphi_i)=\varphi_i-\{K_i^*\}-NEG(\varphi_i)$。

为了节省存储空间，在迭代时无需计算 $NEG(\varphi_i)$和 $BND(\varphi_i)$，只需记录已搜索的最优粒度约简集合 OSC_i 和搜索过程中发现的最大不协调元素集合 ISC_i 即可。其中$\varphi_1=\varphi$，$OSC_0=\varnothing$，$OSC_i=OSC_{i-1}\cup\{K_i^*\}$，$ISC_0=\varnothing$，$ISC_i=MAX(ISC_{i-1}\cup ISC(\varphi_i))$。由上述可知：

$$\begin{aligned}\varphi_{i+1}&=BND(\varphi_i)=\varphi_i-\{K_i^*\}-NEG(\varphi_i)\\&=\varphi_i-\{K_i^*\}-UB_\varphi(K_i^*)\cup LB_\varphi(K_i^*)\cup ISC(\varphi_i)\cup\\&\quad\bigcup_{K\in ISC(\varphi_i)}LB_\varphi(K)\\&=\varphi-OSC_i-\bigcup_{K_1\in OSC_i}UB_\varphi(K_1)-\bigcup_{K_1\in OSC_i}LB_\varphi(K_1)-\\&\quad ISC_i-\bigcup_{K_2\in ISC_i}LB_\varphi(K_2)\end{aligned}$$

故 $K\in\varphi_{i+1}$当且仅当对于任意的$K_1\in OSC_i$，K 和K_1没有偏序关系，且对于任意的$K_2\in ISC_i$，都有 $K\not\preccurlyeq K_2$，即只需与 OSC_i和 ISC_i比较即可确定搜索空间。

下面给出在粒度空间φ_i中找出一个最优粒度约简的算法，如算法 2 所示。

算法 2　逐步搜索最优粒度约简

输入：多粒度决策形式背景$S^e=(G,A^e,I^e,D,J)(e=1,2,\cdots,n)$，粒度搜索空间$\varphi_i$，初始粒度组合$\mathbf{K}_i$，向量$\mathbf{K}_i$的长度 s

输出：最优粒度约简$\mathbf{K}_i^*$，最大不协调元素集合 $ISC(\varphi_i)(i\geqslant 1)$

```
1. Ki* = Ki, ISC(φi) = ∅
2. For j=1 to s do
3.    Ki*[j]=0
4.       While K[j]<Ki[j]do
5.       If Ki* ∉ φi or S^{Ki*} 不协调 then
6.       Ki*[j]=Ki*[j]+1
7.       Else
8.       Break
9.       End if
10.      End while
11.   K′=Ki*
12.   If K′[j]>0 then
13.      K′[j]=K′[j]-1
14.      If K′∈φi then
15.         ISC(φi)=ISC(φi)∪{K′}
16.      End if
17.   End if
18. End for
19. For K∈ISC(φi) do
20.      If K<Ki* then
21.      ISC(φi)=ISC(φi)-{K}
22.      End if
23. End for
24. Return Ki*, ISC(φi)
```

算法 2 扩展了算法 1，可以在迭代的粒度空间中搜索

最优粒度约简，同时根据命题 2 增加了需要存储的最大不协调元素集合 $ISC(\varphi_i)$，协调性检查的时间复杂度也为 $O(ns)$。其中，算法 2 中的第 12－17 行为得到最大不协调元素集合 $ISC(\varphi_i)$。为了节省存储空间，第 19－23 行删除了 $ISC(\varphi_i)$ 中与 $LB_\varphi(K_i^*)$ 中重复的元素。事实上，若 $K_2 \in ISC(\varphi_i)$ 且 $K_2 \prec K_i^*$，则 $LB_\varphi(K_2) \subset LB_\varphi(K_i^*)$。由命题 2 可知 $\varphi_{i+1} = \varphi - OSC_i - \bigcup_{K_1 \in OSC_i} UB_\varphi(K_1) - \bigcup_{K_1 \in OSC_i} LB_\varphi(K_1) - ISC_i - \bigcup_{K_2 \in ISC_i} LB_\varphi(K_2)$，$\bigcup_{K_1 \in OSC_i} LB_\varphi(K_1)$ 和 $\bigcup_{K_2 \in ISC_i} LB_\varphi(K_2)$ 中相同的元素可以被去除。

在迭代搜索过程中，第一个最优粒度约简可以通过算法 1 从最细粒度组合 $(n,n,\cdots,n)$ 开始搜索，为了更高效地搜索后续迭代中的最优粒度约简，给出下面的结论。

命题 3 设 $S^e = (G, A^e, I^e, D, J)(e = 1,2,\cdots,n)$ 为多粒度决策形式背景，φ 为 S^e 的介粒度空间。对于 $\varphi_1 = \varphi$，$\varphi_{i+1} = BND(\varphi_i)$，若 $CMAX(\varphi_{i+1}) \neq \emptyset$，则 φ_{i+1} 中存在最优粒度约简。

证明：由 $CMAX(\varphi_{i+1}) \neq \emptyset$ 可知，存在 $K_1 \in CMAX(\varphi_{i+1})$，又由 $CMAX$ 的定义可知，S^{K_1} 是协调的，下面分 3 种情况进行讨论。1)对于任意的 $K_2 \in LB_{\varphi_{i+1}}(K_1)$，都有 S^{K_2} 是不协调的，由定义 14 可知，K_1 就是最优粒度约简。2) $LB_{\varphi_{i+1}}(K_1) = \emptyset$，即不存在 $K \prec K_1$，故 K_1 就是最优粒度约简。3)存在 $K_3 \in LB_{\varphi_{i+1}}(K_1)$，$S^{K_3}$ 是协调的，此时需要继续判断 $K_4 \in LB_{\varphi_{i+1}}(K_3)$ 是否符合 1)或者 2)的情况。若不符合就用 3)的情况继续迭代，因为 φ_{i+1} 是有限集，所以最终必然存在最优粒度约简，故命题 3 成立。

由命题 3 可知，若存在 $K \in CMAX(\varphi_{i+1})$，则存在最优粒度约简，此时可以将 K 作为算法 2 的初始粒度组合进行新一轮的搜索。命题 4 给出了一种简单的搜索 $K \in CMAX(\varphi_{i+1})$ 的方法。

命题 4 设 $S^e = (G, A^e, I^e, D, J)(e = 1,2,\cdots,n)$ 为多粒度决策形式背景，φ 为 S^e 的介粒度空间。对于 $\varphi' \subseteq \varphi$，$K = (l_1, l_2, \cdots, l_s) \in \varphi'$。若对任意的 $K' = (l_1', l_2', \cdots, l_s') \in \varphi'$，都有 $\sum_{i=1}^{s} l_i \geqslant \sum_{i=1}^{s} l_i'$，则 $K \in MAX(\varphi')$。

根据命题 4，可以在每轮搜索结束后剩余的粒度空间中计算拥有最大分量和的粒度组合，并将其中协调的粒度组合作为初始粒度组合进行下一轮迭代，如例 6 所示。

例 6 对于例 5 找到的最优粒度约简 $K_1^* = (0,2,0)$，$ISC(\varphi_1) = (0,1,2)$。由命题 2 计算 $\varphi_2 = \{(2,0,2),(1,0,2),(2,0,1),(1,0,1),(2,0,0),(1,0,0),(2,1,2),(1,1,2),(2,1,1),(1,1,1),(2,1,0),(1,1,0)\}$，由命题 4 计算 $MAX(\varphi_2) = \{(2,1,2)\}$。因为 $K = (2,1,2)$ 对应的 S^K 是协调的，下一轮迭代初始粒度组合为(2,1,2)。

算法 3 给出了在介粒度空间 φ 中找到所有最优粒度约简的方法。

算法 3 逐步搜索所有最优粒度约简

输入：多粒度决策形式背景 $S^e = (G, A^e, I^e, D, J)(e = 1,2,\cdots,n)$，$S^e$ 的介粒度空间 φ

输出：φ 的所有最优粒度约简 OSC

1. OSC $= \emptyset$，ISC $= \emptyset$，i $= 1$，$\varphi_1 = \varphi$
2. While $\varphi_i \neq \emptyset$ do
3. 由命题 4 计算满足条件的 MAX(φ_i)，其中 MAX(φ_1) $= \{(n, n, \cdots, n)\}$
4. IMAX(φ_i) $= \emptyset$
5. For K $\in$ MAX(φ_i) do
6. If S^K 不协调 then
7. IMAX(φ_i) $=$ IMAX(φ_i) $\cup$ {K}
8. Else
9. 将 K 作为初始粒度组合，由算法 2 得到 K_i^* 和 ISC(φ_i)
10. Break
11. End if
12. End for
13. OSC $=$ OSC $\cup$ {K_i^*}，ISC(φ_i) $=$ MAX(ISC(φ_i) $\cup$ IMAX(φ_i))，ISC $=$ MAX(ISC $\cup$ ISC(φ_i))
14. 由命题 2 计算 φ_{i+1}
15. End while
16. Return OSC

算法 3 从介粒度空间 φ 开始(算法 3 中的第 1 行)，因为 $\varphi_1 \neq \emptyset$(算法 3 中的第 2 行)，所以将 $(n,n,\cdots,n)$ 作为第 1 轮初始粒度组合，调用算法 2 得到一个最优粒度约简和最大不协调元素集合(算法 3 中的第 9 行)，然后计算得到需要保存的信息 ISC 和 OSC(算法 3 中的第 13 行)。接着由命题 2 计算剩余空间 φ_{i+1}(算法 3 中的第 14 行)。若 φ_{i+1} 不为空集(Step 2)，则计算满足命题 4 的剩余空间的最大元素集合 $MAX(\varphi_{i+1})$。通过判断这些元素的协调性，将不协调的元素并入 $IMAX(\varphi_{i+1})$，将协调的元素作为初始粒度组合进行迭代(算法 3 中的第 5—12 行)。若 φ_{i+1} 为空集(算法 3 中的第 2 行)，则表示找到了所有最优粒度约简，终止搜索。依次迭代，直到找到所有最优粒度约简。

算法 3 中，设 c 为多粒度决策形式背景 $S^e = (G, A^e, I^e, D, J)(e = 1,2,\cdots,n)$ 中最优粒度约简的个数。由命题 3 可知，算法 3 中的第 2－15 行迭代搜索次数为 c。由命题 4 可知，每轮搜索结束后，最多有 $(n+1)s$ 个最大元素，因此算法 3 中的第 5－12 行判断其协调性最多循环 $(n+1)s$ 次，其中第 9 行调用 1 次算法 2 判断协调性的时间复杂度为 $O(ns)$。因此，算法 3 的协调性检查的时间复杂度为 $O(c \cdot ((n+1)s + ns)) = O(cns)$，远低于依次判断所有粒度组合协调性的时间复杂度 $O((n+1)^s)$。

定理 1 证明了逐步找到的最优粒度约简都是 φ 中的最优粒度约简。

定理 1 设 $S^e = (G, A^e, I^e, D, J)(e = 1,2,\cdots,n)$ 为多粒度决策形式背景，φ 为 S^e 的介粒度空间。对于 $\varphi_1 = \varphi$，$\varphi_{i+1} = BND(\varphi_i)$。若 K_{i+1}^* 是 φ_{i+1} 的最优粒度约简，则 K_{i+1}^* 是 φ 的最优粒度约简。

证明：只需要证明 K_{i+1}^* 是 φ_i 的最优粒度约简，以此类推就可以得到 K_{i+1}^* 是 φ 的最优粒度约简。假设 K_{i+1}^* 不是 φ_i 的最优粒度约简，当 $CMAX(\varphi_i) = \emptyset$ 时，由命题 3 可知，φ_i 中没有最优粒度约简，因此 $\varphi_{i+1} = \emptyset$，故不存在 K_{i+1}^*，与假设矛盾。当 $CMAX(\varphi_i) \neq \emptyset$ 时，由于 K_{i+1}^* 是 φ_{i+1} 的最优粒度约简，因此 $S^{K_{i+1}^*}$ 是协调的，但由于 K_{i+1}^* 不是 φ_i 的最优粒度约简，因此 φ_i 中存在最优粒度约简 K 满足 $K \prec K_{i+1}^*$，即 $K_{i+1}^* \in UB_\varphi(K)$。因此，$K_{i+1}^* \in NEG(\varphi_i)$，与 $K_{i+1}^* \in \varphi_{i+1} = BND(\varphi_i)$ 矛盾。故定理 1 成立。

定理 2 证明了最优粒度约简之间没有偏序关系，特别地 $K_i^* \neq K_j^*$，其中 $i \neq j$。

定理 2　设 $S^e=(G,A^e,I^e,D,J)(e=1,2,\cdots,n)$ 为多粒度决策形式背景，φ 为 S^e 的介粒度空间。对于 $\varphi_1=\varphi$，$\varphi_{i+1}=BND(\varphi_i)$。若 K_{i+1}^* 是 φ_{i+1} 中的最优粒度约简，则对于任意的 $j\in\{1,2,\cdots,i\}$，K_{i+1}^* 和 K_j^* 都没有偏序关系。

证明：由 K_{i+1}^* 是 φ_{i+1} 中的最优粒度约简可知，$K_{i+1}^*\in\varphi_{i+1}$。由命题 2 可知，$K\in\varphi_{i+1}$ 当且仅当 $K\notin\{OSC_i \cup \bigcup_{K_1\in OSC_i} UB_\varphi(K_1) \bigcup_{K_1\in OSC_i} LB_\varphi(K_1) \cup ISC_i \cup \bigcup_{K_2\in ISC_i} LB_\varphi(K_2)\}$，其中 $OSC_i=\{K_1^*,\cdots,K_j^*,\cdots,K_i^*\}$，即任意的 K_j^* 和 K_{i+1}^* 都没有偏序关系。因此，定理 2 成立。

例 7　例 2 中的多粒度决策形式背景 $S^e=(G,A^e,I^e,D,J)(e=1,2)$ 寻找所有最优粒度约简的完整过程如下。

从介粒度空间 φ 开始，第 1 轮搜索过程如例 5 所示，得到 $K_1^*=(0,2,0)$，$ISC(\varphi_1)=\{(0,1,2)\}$，故第 1 轮搜索后有 $ISC=\{(0,1,2)\}$，$OSC=\{(0,2,0)\}$。如例 6 所示，由命题 2 计算 $\varphi_2\neq\varnothing$，由命题 4 计算满足条件 $MAX(\varphi_2)=\{(2,1,2)\}$。由于 $K=(2,1,2)$ 对应的 S^K 是协调的，因此将 $(2,1,2)$ 作为初始粒度组合，调用算法 2 开始第 2 轮迭代。初始组合为 $K_2^*=(2,1,2)$。开始搜索第一个分量，令 $K_2^*[1]=0$。由于 $K_2^*=(0,1,2)$ 不在 φ_2 内，因此需要验证 $K_2^*=(1,1,2)$。由于 $(1,1,2)$ 对应的 $S^{K_2^*}$ 是不协调的，因此 $K_2^*=(2,1,2)$。由于第一个分量无法再增加，记录不协调的粒度组合 $ISC(\varphi_2)=\{(1,1,2)\}$。开始搜索第二个分量，令 $K_2^*[2]=0$，$K_2^*=(2,0,2)$ 对应的 $S^{K_2^*}$ 是协调的。类似地，$K_2^*=(2,0,0)$ 对应的 $S^{K_2^*}$ 也是协调的，因此得到 $K_2^*=(2,0,0)$，$ISC(\varphi_2)=\{(1,1,2)\}$。第 2 轮搜索后有 $ISC=MAX(\{(0,1,2)\}\cup\{(1,1,2)\})=\{(1,1,2)\}$，$OSC=\{(0,2,0),(2,0,0)\}$。由命题 2 计算 $\varphi_3=\varnothing$，因此结束搜索。输出所有最优粒度约简 $OSC=\{(0,2,0),(2,0,0)\}$。

5　实验结果及分析

本文验证了所提出的逐步搜索所有最优粒度约简方法的性能，结果如表 5 所列，从 UCI 数据集[1] 中选取 5 个公开数据集 Iris，Cloud，WineQuality-red，WineQuality-white，Abalone。

表 5　实验数据集

Table 5　Datasets

数据集	对象个数	条件变量	决策变量
Iris	150	4	1
Cloud	1024	10	1
WineQuality-red	1599	11	1
WineQuality-white	4898	11	1
Abalone	4177	7	1

为了获得进行实验所需的多粒度决策形式背景，对 5 个数据集进行相应的处理，将数据集的最后一列标签类别属性作为决策属性，其他列作条件属性。此时，以条件属性为类属性块，将所有属性值都转化为 0-1 布尔属性，对于决策属性，只将其转化为 0-1 布尔属性。具体来说，对于原始数据集中的连续值属性，根据数据的分布特点进行分段处理，然后使用新的布尔属性来代替每个分段区间；对于原始数据集中的多值属性，将每个属性取值都视为一个新的布尔属性进行处理。需要指出的是，将预处理后得到的数据作为最细粒度下的决策形式背景。若最细粒度下的决策形式背景不协调，则后续判断协调性转化为与最细粒度下的决策形式背景的决策保持一致。

在此基础上，将每个类属性块内相邻的属性进行适当的合并，从而得到更粗粒度的决策形式背景。最终得到了从细粒度到粗粒度的多粒度决策形式背景。将原始数据集经过上述预处理后记为数据集 1—数据集 5，最优粒度约简结果如表 6 所列，协调性检查次数、运行时间结果如表 7 所列，其中时间数据是实验重复 100 次取平均值得到。

表 6　最优粒度约简结果

Table 6　Results of optimal franularity reduction

数据集	最细粒度组合	最优粒度约简
数据集 1	(2,3,2,1)	(1,1,1,0)
数据集 2	(7,1,2,1,3,2,1,4,7,4)	(1,1,1,1,1,0,1,2,7,4)；(1,1,1,1,1,1,0,2,7,4)
数据集 3	(3,4,2,7,7,3,2,4,5,5,3)	(1,1,1,1,1,1,1,1,2,2,1,1)
数据集 4	(7,7,7,4,11,6,6,4,4,4,2)	(3,2,3,1,1,2,2,1,3,2,1)
数据集 5	(3,2,4,2,3,3,4)	(2,2,3,1,3,2,2)

表 7　搜索协调性检查次数和时间结果

Table 7　Results of coordination check times and running time for search

数据集	粒度组合数量	检查次数	时间/s
数据集 1	72	11	0.0007
数据集 2	460800	65	0.0802
数据集 3	33177600	36	0.0068
数据集 4	564480000	44	0.0218
数据集 5	14400	32	0.0059

表 6 表明，大部分多粒度决策形式背景只有一个最优粒度约简，对象个数较少时容易存在可约简的类属性块。表 7 的结果表明，本文方法基本有效，通过减少协调性检查的次数，可以在较短的时间内找到所有最优粒度约简。

为了进一步评估方法在对象变化时的灵敏性，分别计算了数据集 1/4，1/2 和全部对象时的最优粒度约简，结果如表 8 所列。

表 8　最优粒度约简变化结果

Table 8　Changing results of optimal granularity reduction

数据集	前 1/4 对象	前 1/2 对象	全部对象
数据集 1	(0,0,1,0)	(0,0,1,1)	(1,1,1,0)
数据集 2	(0,1,0,1,1,0,3,0,4,3)	(1,1,1,1,1,1,2,4,5,4)	(1,1,1,1,1,0,1,2,7,4)；(1,1,1,1,1,1,0,2,7,4)
数据集 3	(1,1,1,1,1,1,1,1,1,1,1)	(1,1,1,1,1,1,1,1,2,1,1)	(1,1,1,1,1,1,1,2,2,1,1)
数据集 4	(1,3,1,1,1,1,1,1,1,1,1)	(1,2,2,1,1,1,2,1,2,2,1)	(3,2,3,1,1,2,2,1,3,2,1)
数据集 5	(2,2,2,1,1,2,2)	(2,2,2,2,2,2,2)	(2,2,3,1,3,2,2)

表 8 的结果表明，随着对象的增加，最优粒度约简的粒度整体上会变得越来越细。

[1] https://archive.ics.uci.edu/ml/index.php

为了了解对象变化后时间开销的变化，进一步评估方法对对象数量的敏感性，计算对象变化时搜索时间、协调性检查次数的结果，如表9和表10所列。

表9 搜索时间变化结果

Table 9 Changing results of time for search

(单位：s)

数据集	前1/4对象	前1/2对象	全部对象
数据集1	0.0005	0.0006	0.0007
数据集2	0.0015	0.0023	0.0802
数据集3	0.0032	0.0042	0.0068
数据集4	0.0061	0.0095	0.0218
数据集5	0.0024	0.0035	0.0059

表10 搜索协调性检查次数变化结果

Table 10 Changing results of coordination check times for search

数据集	前1/4对象	前1/2对象	全部对象
数据集1	6	7	11
数据集2	30	41	65
数据集3	32	33	36
数据集4	34	38	44
数据集5	27	31	32

表9和表10的结果表明，随着数据表中对象数量的增加，搜索的时间和协调性检查的次数增加较少。

根据实验结果，可以得出以下结论：

1)多粒度决策形式背景中的对象个数越多，选出的最优粒度约简粒度相对越细，确定性决策需要越强的约束。

2)多粒度决策形式背景中最优粒度约简个数通常很少。当最优粒度约简有多个时，确定剩余的粒度空间需要较多的时间开销。

3)该方法对数据的对象数量不敏感，处理一些复杂数据时，可以在较短的时间内找到所有最优粒度约简。

结束语 为了研究如何在多粒度数据的粒度组合中降低决策代价，本文通过从属性融合的方式中分析最优粒度选择和属性约简的联系，创新之处是定义了一种可以同时进行最优粒度选择和属性约简的介粒度空间，扩大了粒度选择的范围。在此基础上，本文提出了逐步搜索所有最优粒度约简的方法，在介粒度空间上搜索最优粒度约简并通过已搜索的信息更新剩余的粒度空间，去除大量非最优粒度约简，显著减少了单粒度决策形式背景所需的协调性检查次数。此外，本文的结果有助于今后对多粒度决策形式背景的最优粒度约简进行进一步研究。

然而，基于介粒度空间的最优粒度约简方法仍存在一些不足和需要进一步探讨的问题。

1)本文不考虑各个多粒度类属性块的权重因素。但是，在实际情况中，人们通常会对某些重要属性进行特殊对待，因此当用户提出具体的类属性块重要性时，判断最优粒度约简也值得进一步探讨。

2)现实数据中存在数据不完备或者数据模糊的情况，因此如何在更加复杂的现实情况下判断最优粒度约简也是一个有意义的研究课题。

3)在进行形式背景的属性融合时，会导致信息的损失，该过程是不可逆的。因此，如何在精度、效率和计算代价之间进行综合权衡，以平衡这些因素，是一个需要进一步探讨的问题。

参 考 文 献

[1] WILLE R. Restructuring lattice theory: An approach based on hierarchies of concepts[C]// IVAN R. Ordered Sets. Berlin, Germany: Springer, 1982: 445-470.

[2] ZHANG L, ZHANG B. Quotient space based problem solving: A theoretical foundation of granular computing[M]. Beijing, China: TsinghuaUniversity Press, 2014.

[3] MULKAR-MEHTA R, HOBBS J, HOVY E. Granularity in natural language discourse[C]// Proceedings of the 9th International Conference on Computational Semantics. Stroudsburg, USA: ACL, 2011: 360-364.

[4] WEI L, WAN Q. Granular Transformation and irreducible element judgment theory based on pictorial diagrams[J]. IEEE Transactions on Cybernetics, 2016, 46(2): 380-387.

[5] LI F, HU B Q. A new approach of optimal scale selection to multi-scale decision tables[J]. Information Sciences, 2017, 381, 193-208.

[6] YAO Y Y. Three-way decisions with probabilistic rough sets [J]. Information Sciences, 2010, 180(3): 341-353.

[7] CHENG Y L, ZHANG Q H, WANG G Y, et al. Optimal scale selection and attribute reduction in multi-scale decision tables based on three-way decision[J]. Information Sciences, 2020, 541(1): 36-59.

[8] HAO C, LI J H, FAN M, et al. Optimal scale selection in dynamic multi-scale decision tables based on sequential three-way decisions[J]. Information Sciences, 2017, 415/416: 213-232.

[9] MIAO D Q, ZHANG, Q H, QIAN Y H, et al. From human intelligence to machine implementation model: Theories and applications based ongranular computing[J]. CAAI Transactions on Intelligent Systems, 2016, 11(6): 743-757.

[10] ZENG W L, ZHE Y H. Object-oriented multigranulation formal concept analysis[J]. Computer Science, 2018, 45(10): 51-53, 63.

[11] ZHANG C, BAI W H, LI D Y, et al. Multiple attribute group decision making based on multigranulation probabilistic models, MULTIMOORA and TPOP in incomplete q-rung orthopair fuzzy information systems[J]. International Journal of Approximate Reasoning, 2022, 143(4): 102-120.

[12] PANG J F, SONG P, LIANG J Y. Review on multi-granulation computing models and methods for decision analysis[J]. Pattern Recognition and Artificial Intelligence, 2021, 34(12): 1120-1130.

[13] WANG Q, LI D Y, ZHAI Y H, et al. Parameterized fuzzy decision implication[J]. Journal of Computer Research and Development, 2022, 59(9): 2066-2074.

[14] ZHANG S X, LI D Y, ZHAI Y H. Incremental method of generating decision implication canonical basis[J]. Soft Computing: A Fusion of Foundations, Methodologies and Applications, 2022, 26(3): 1067-1083.

[15] ZHAI Y H, JIA N, ZHANG S X, et al. Study on deduction process and inference methods of decision implications[J]. Inter-

national Journal of Machine Learning and Cybernetics，2022，13(7)：1959-1979.

[16] ZHANG C，LI D Y. Interval-valued hesitant fuzzy graphs decision making with correlations and prioritization relationships [J]. Journal of Computer Research and Development，2019，56(11)：2438-2447.

[17] LI J H，WU W Z，DENG S. Multi-scale theory in formal concept analysis[J]. Journal of Shandong University(Natural Science)，2019，54(2)：30-40.

[18] LI J H，LI Y F，MI Y L，et al. Meso-granularity labeled method for multi-granularity formalconcept analysis[J]. Journal of Computer Research and Development，2020，57(2)：447-458.

[19] WU W Z，LEUNG Y，MI J S. Granular computing and knowledge reduction in formal contexts[J]. IEEE Transaction on Knowledge and Data Engineering，2009，10(21)：1461-1474.

[20] HAO C，FAN M，LI J H，et al. Optimal scale selection in multiscale contexts based on granular scale rules[J]. Pattern Recognition and Artificial Intelligence，2016，29(3)：272-280.

[21] LI J H，HE J J. Uncertainty measurement and optimal granularity selection for multi-granularity formal context[J]. Control and Decision，2022，37(5)：1299-1308.

[22] LI J H，ZHOU X R. Attribute reduction in multi-granularity formal decision contexts[J]. Pattern Recognition and Artificial Intelligence，2022，35(5)：387-400.

[23] ZHANG W X，QIU G F. Uncertain decision making based on rough set [M]. Beijing，China：Tshinghua University Press，2005.

LI Teng，born in 1999，master. His main research interests include data mining and formal concept analysis.

LI Deyu，born in 1965，Ph.D，professor，Ph.D supervisor，is a senior member of China Computer Federation. His main research interests include data mining and multi-label learning.

（责任编辑：喻藜）