

# 动态更新属性值变化时的最优粒度

刘凤玲 林国平

( 闽南师范大学 数学与统计学院 福建 漳州 363000)

E-mail: guoplin@163.com

**摘要:** 针对在实际应用中可能由于人为或其他因素造成某个粒度下属性值变化的问题,研究了某个粒度下属性值变化时最优粒度的选择策略,并根据理论研究内容设计了相应的静态和动态最优粒度选择算法,最后在 UCI 数据库中选取 6 组数据进行了实验.实验结果表明,新提出的方法是可行且有效的,且相对比静态算法动态最优粒度选择算法在一定程度上减少了计算时间,提高了时间效率.

**关键词:** 动态;属性值变化;多粒度标记;最优粒度

中图分类号: TP18

文献标识码: A

文章编号: 1000-4220(2020)10-2063-05

## Updating Optimal Scale in Multi-scale Decision Systems Under Environment of Attribute Value

LIU Feng-ling, LIN Guo-ping

( Department of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, China)

**Abstract:** In practical application, attribute value is changed under a certain scale due to man-made or other factors. Aiming at the problem, the optimal scale selection strategy under environment of attribute value updating was studied, and the corresponding dynamic optimal scale selection algorithm was designed according to the theory proposed in this paper. Finally, comparative experiments by using six data sets from UCI were carried out. The experimental results illustrate the proposed method is feasible and effective, and the dynamic optimal scale selection algorithm reduces computation time and improves time efficiency to some extent.

**Key words:** dynamic; attribute value updating; multi-scale; optimal scale

### 1 引言

数据规模随着信息技术的迅速发展而不断膨胀,粒计算常用于处理复杂的数据系统.粒计算是将所需探究的论域划分成若干相对简单的粒,簇、块或集合构成了这些粒.

Zadeh 于 1979 年首次提出模糊信息粒这一概念<sup>[1]</sup>.随后,相关应用背景的粒计算模型与方法相继被提出<sup>[2-5]</sup>.信息粒在解决复杂问题中起重要作用,其可将较为复杂的信息系统抽象地转化为若干相对简单的信息系统,这样既可以降低处理难度,又可以提高预测信息的准确性.

在大部分信息系统中,同一对象在同一个属性下只有一个属性值,这使得人们只能从固定的粒度或视角分析数据信息.然而,在实际应用中,根据实际问题的不同需要,同一个对象在同一个属性下可以取不同粒度层次标记的值.例如,需要评定奖学金时学生的考试成绩一般记录为 0~100 之间的数,需要知晓成绩优秀、较好等级别的人数时可以将成绩记为“优秀”、“较好”、“中等”、“差”.进一步的,如果需要,还可以将其分为两个值,比如给予学位认定时记为“合格”和“不合格”.基于此, Wu 等人根据对象在决策系统中拥有不同的粒度标记,于 2011 年首次给出了多粒度标记决策系统的概念<sup>[6]</sup>.近年来,基于多粒度标记划分的粗糙集数据分析方法

得到不断完善和发展,多粒度标记决策系统是人们研究的热点.目前,查阅相关文献发现学者们主要从最优粒度选择<sup>[7-17]</sup>、规则约简或提取<sup>[18,19]</sup>及知识表示<sup>[20]</sup>等方面对多粒度标记决策系统进行研究.从系统的协调性与完备性的角度来看,学者们主要研究了协调且完备的、协调不完备的以及不协调但完备的这三种多粒度标记决策系统.在实际应用中,多粒度标记决策系统是常被用于各种信息分析,如被广泛应用于数据挖掘、图像处理、人工智能、地理信息甄别和军事技术等领域.

如今是大数据时代,在许多实际情况下,由于各种需求或人为失误,数据可能会发生变化.因此,需要对动态数据集中隐藏的知识进行相应的更新.增量更新机制是利用前面获得的计算结果来获取动态数据集中的知识而不是重新从头开始计算.在动态环境中,目前基于粗糙集理论的增量更新方法主要涉及对象的变化<sup>[21-23]</sup>、属性的变化<sup>[24,25]</sup>以及属性值的变化<sup>[26,27]</sup>.而有关多粒度标记决策系统方面的动态更新,本人只查阅到学者们讨论了在多粒度决策系统中,对象更新环境下选择最优粒度的策略并设计了相应的算法.

事实上,在实际应用中人为或其他因素难免不出现收集到的多粒度标记决策系统在某个粒度标记下的属性值是错误的,而此时则需要将正确的属性值更新进去,从而得正确的数据,即此时新的系统属性值发生了更新.若当前粒度尺度下属

收稿日期: 2019-12-24 收修改稿日期: 2020-01-19 基金项目: 国家青年科学基金项目(61603173,11871289,61379021,11701258) 资助; 浙江省海洋大数据挖掘与应用重点实验室开放课题项目(OBDMA201603) 资助; 福建省自然科学基金项目(2019J01748) 资助. 作者简介: 刘凤玲,女,1994年生,硕士研究生,研究方向为粗糙集理论、粒计算; 林国平(通讯作者),女,1978年生,博士,教授,研究方向为粗糙集理论、粒计算、人工智能、机器学习.

性值发生变化,则粗于该粒度尺度的粒度尺度可能也随之发生变化.一般来说,在多粒度标记系统中等级越小的粒度越细,然而粒越细获取知识的成本越高,因此选择合适的粒度级别来获取规则及求目标近似集等较为重要,即最优粒度选择是多粒度标记系统中的知识获取较为关键的一步.基于此,本文针对属性值变化环境下的不协调多粒度决策系统最优粒度的选择进行研究.

由于 WU 在文献[9]中分析了不协调多粒度决策系统中 8 种不协调性选择最优粒度的对比,指出实际上只有 4 种:分布协调;最大分布协调;而分配协调、似然协调、上近似协调及广义协调互相等价;下近似协调与信任协调等价.基于此,本文将只讨论 4 种协调性下属性值变化环境下的不协调多粒度决策系统选择最优粒度的策略:分布协调、最大分布协调、上近似协调及下近似协调.

## 2 基本知识

### 2.1 多粒度标记决策系统的概念

定义 1<sup>[6]</sup>. 一个多粒度标记信息系统记为  $S = (U, AT, V_j^k, f)$ . 其中  $U = \{x_1, x_2, \dots, x_n\}$  为一个非空有限对象集,被称为论域;  $AT = \{AT^k \mid AT^k = \{a_1^k, a_2^k, \dots, a_m^k\}, k = 1, 2, \dots, l\}$  为一个非空有限属性集;  $V_j^k = \{f_{a_j^k}(x) \mid x \in U\}$  为属性  $a_j$  在第  $k$  个粒度标记下的值域;  $f_{a_j^k}(x)$  表示对象  $x$  在属性  $a_j^k$  下的值.  $(U, AT, V_j^k, f)$  亦可简记为  $(U, AT)$ .

在多粒度标记系统中,每一个属性  $a_j \in AT$  均为多粒度标记属性,也即对于同一个对象  $x \in U$  在不同粒度层面下有不同的取值.为方便讨论,本文约定没有特别说明情况下在同一系统中,所有条件属性的粒度层次均为  $l$ .

对于  $j \in \{1, 2, \dots, m\}$  与  $1 \leq k \leq l-1$ ,总存在一个满射  $g_j^{k+1}: V_j^k \rightarrow V_j^{k+1}$ ,使得  $f_{a_j^{k+1}} = g_j^{k+1} \circ f_{a_j^k}$ .也即每个对象  $x \in U$  在第  $k+1$  粒度标记层面的值为  $f_{a_j^{k+1}}(x) = g_j^{k+1}(f_{a_j^k}(x))$ ,称  $g_j^{k+1}$  为信息粒度变换.

定义 2<sup>[6]</sup>. 一个多粒度标记决策信息系统记为  $dIS = (U, C \cup \{d\})$ .  $d$  为决策属性;  $C$  为条件属性集,其中  $C = \{C^k \mid C^k = \{a_1^k, a_2^k, \dots, a_m^k\}, k \in \{1, 2, \dots, l\}\}$ .  $C^k$  表示第  $k$  粒度层次下的条件属性集.

对任意的  $C^k \subseteq C$  将论域  $U$  在  $C^k$  下的等价关系、等价类及等价划分分别记为:

$$R_{C^k} = \{(x, y) \in U \times U \mid \forall a_j^k \in C^k, f_{a_j^k}(x) = f_{a_j^k}(y)\}; [x]_{C^k} = \{y \in U \mid (x, y) \in R_{C^k}\}; U/R_{C^k} = \{[x]_{C^k} \mid x \in U\}.$$

类似记:  $R_{\{d\}} = \{(x, y) \in U \times U \mid f_d(x) = f_d(y)\}$ ,  $[x]_{\{d\}} = \{y \in U \mid (x, y) \in R_{\{d\}}\}$ ,  $U/R_{\{d\}} = \{[x]_{\{d\}} \mid x \in U\} = \{D_1, D_2, \dots, D_r\}$ ,  $r$  为决策类划分的个数,即  $r = |U/R_{\{d\}}|$ .

对于论域  $U$  中的一个子集  $X$ ,其关于  $R_{C^k}$  的下近似和上近似分别为:

$$\underline{R}_{C^k}(X) = \{x \in U \mid [x]_{C^k} \subseteq X\}, \\ \overline{R}_{C^k}(X) = \{x \in U \mid [x]_{C^k} \cap X \neq \emptyset\}.$$

若对任意的  $[x]_{R_{C^k}} \in U/R_{C^k}$ ,总存在  $[x]_{R_d} \in U/R_{\{d\}}$  使得  $[x]_{R_{C^k}} \subseteq [x]_{R_d}$  成立,则记  $R_{C^k} \subseteq R_d$ .

假设  $dIS = (U, C \cup \{d\})$  为一个多决策的多粒度标记决策系统,该系统中所有条件属性均有  $l$  个等级粒度,则  $dIS = (U, C \cup \{d\})$  可分解为  $l$  个信息系统,也即  $dIS = dIS^1 \cup dIS^2 \cup \dots \cup dIS^l$ .

### 2.2 不协调多粒度标记决策系统的协调性与最优粒度选择

定义 3<sup>[6]</sup>. 设  $dIS = (U, C \cup \{d\})$  为多粒度决策系统,若  $R_{C^1} \subseteq R_d$ ,则系统  $dIS$  是协调的.若进一步满足  $R_{C^k} \subseteq R_d$ ,则称  $dIS^k = (U, C^k \cup \{d\})$  协调.若  $dIS^k = (U, C^k \cup \{d\})$  协调,则对任意的  $k' < k$  粒度层次,  $dIS^{k'} = (U, C^{k'} \cup \{d\})$  也必协调.若  $R_{C^1} \not\subseteq R_d$ ,则系统  $dIS$  不协调.对所有的粒度层次  $k \in \{1, 2, \dots, l\}$ ,记

$$L_{C^k}(d) = (\underline{R}_{C^k}(D_1), \underline{R}_{C^k}(D_2), \dots, \underline{R}_{C^k}(D_r)), H_{C^k}(d) = (\overline{R}_{C^k}(D_1), \overline{R}_{C^k}(D_2), \dots, \overline{R}_{C^k}(D_r)), \\ \mu_{C^k}(x) = (D([x]_{C^k}, D_1), D([x]_{C^k}, D_2), \dots, D([x]_{C^k}, D_r)), \gamma_{C^k}(x) = \{D \in U/R_{\{d\}} \mid D([x]_{C^k}, D) = \max_{j=1, 2, \dots, r} D([x]_{C^k}, D_j)\}, \text{其中 } D([x]_{C^k}, D_1) = \frac{|[x]_{C^k} \cap D_1|}{|[x]_{C^k}|}.$$

定义 4<sup>[9]</sup>. 设  $dIS = (U, C \cup \{d\})$  为不协调的多粒度决策系统,则

1) 如果  $L_{C^k}(d) = L_{C^1}(d)$ ,就称  $dIS^k$  关于  $dIS$  下近似协调.若  $dIS^k$  关于  $dIS$  下近似协调,但若  $k+1 \leq l$ ,  $dIS^{k+1}$  关于  $dIS$  不是下近似协调,那么  $S$  的下近似最优粒度为第  $k$  粒度层次.

2) 如果  $H_{C^k}(d) = H_{C^1}(d)$ ,就称  $dIS^k$  关于  $dIS$  上近似协调.若  $dIS^k$  关于  $dIS$  上近似协调,但如果  $k+1 \leq l$ ,  $dIS^{k+1}$  关于  $dIS$  不是上近似协调,那么  $S$  的下近似最优粒度为第  $k$  粒度层次.

3) 如果对任意的  $x \in U$ ,有  $\mu_{C^k}(d) = \mu_{C^1}(d)$ ,就称  $dIS^k$  关于  $dIS$  分布协调.若  $dIS^k$  关于  $dIS$  分布协调,但若  $k+1 \leq l$ ,  $dIS^{k+1}$  关于  $dIS$  不是分布协调,那么  $S$  的分布最优粒度为第  $k$  粒度层次.

4) 如果对任意的  $x \in U$ ,有  $\gamma_{C^k}(d) = \gamma_{C^1}(d)$ ,就称  $dIS^k$  关于  $dIS$  最大分布协调.若  $dIS^k$  关于  $dIS$  最大分布协调,但若  $k+1 \leq l$ ,  $dIS^{k+1}$  关于  $dIS$  不是最大分布协调,那么  $S$  的最大分布最优粒度为第  $k$  粒度层次.

## 3 属性值变化环境下的不协调多粒度决策系统的协调性与最优粒度选择

定理 1. 设在  $t$  时刻,不协调多粒度决策系统为  $dIS^{(t)} = (U, C \cup \{d\})$ ,且系统  $dIS^{(t)}$  的下近似最优粒度层次为第  $k'$  粒度层次;在  $t+1$  时刻,  $k' < k$  粒度层次下的属性值保持不变,第  $k$  粒度层次下的属性值发生了变化,  $k'' > k$  粒度层次下的属性值也随之发生变化,记此时的多粒度决策系统为  $dIS^{(t+1)} = (U, C \cup \{d\})$ ,并记此时该系统  $dIS^{(t+1)}$  的下近似最优粒度层次为第  $k'^{t+1}$  粒度层次,则以下两个命题成立:

- 1) 当  $k > k' + 1$  时,则  $k'^{t+1} = k'$ ;
- 2) 当  $k \leq k' + 1$  时,若  $dIS^k$  关于  $dIS^{t+1}$  不是下近似协调的,则  $k'^{t+1} = k-1$ ;若  $dIS^k$  关于  $dIS^{t+1}$  是下近似协调的,则判断  $k+1$  的下近似协调性,若  $dIS^{k+1}$  关于  $dIS^{t+1}$  是下近似协调的,则继续判断  $k+2$  的下近似协调性,以此循环,直至  $k+m$  关于  $dIS^{t+1}$  不是下近似协调的,则  $k'^{t+1} = k+m-1$ .

证明: 在  $t$  时刻, 系统  $dIS^{(t)}$  的下近似最优粒度层次为第  $k^t$  粒度层次, 则由定义 4 易知, 对任意的  $k' \leq k^t$ ,  $dIS^{k'}$  均关于  $dIS^{k^t}$  是下近似协调的;  $dIS^{k^t+1}$  关于  $dIS^{k^t}$  不下近似协调。

1) 若  $k > k^t + 1$ , 则说明在  $t+1$  时刻, 任意的  $k' \leq k^t + 1$  所对应的属性值均未发生变化, 此时对任意的  $k' \leq k^t + 1$ ,  $dIS^{k'}$  均关于  $dIS^{k^t+1}$  是下近似协调的, 又  $dIS^{k^t+1}$  关于  $dIS^{k^t+1}$  不是下近似协调的, 从而由定义 4 可知  $t+1$  时刻系统的下近似最优粒度为  $k^t$ , 即可得  $k^{t+1} = k^t$ 。

2) 若  $k \leq k^t + 1$ , 则说明任意的  $k' \leq k-1$  所对应的属性值均未发生变化, 即此时对任意的  $k' \leq k-1$ ,  $dIS^{k'}$  均关于  $dIS^{k^t+1}$  是下近似协调的, 若  $dIS^k$  关于  $dIS^{k^t+1}$  不是下近似协调的, 又  $dIS^{k-1}$  关于  $dIS^{k^t+1}$  是下近似协调的, 则由定义 4 可知  $t+1$  时刻系统的下近似最优粒度层次为  $k-1$ ; 若  $dIS^k$  关于  $dIS^{k^t+1}$  下近似协调则由定义 4 中的 (1) 可知成立。

下面给出例子来分析及理解所提出的定理。

例 1. 如表 1 为一个多粒度标记决策系统, 记为  $dIS = (U, C \cup \{d\})$ , 该表的条件属性有三个粒度层次, 其中“G, F, B, S, M, L, Y, N”分别表示“好, 中等, 差, 小, 中, 大, 是, 否”。按粒度层次该信息表可分解为三个信息表, 即  $dIS = dIS^1 \cup dIS^2 \cup dIS^3$ , 其中,

$$dIS^1 = \{U \setminus \{a_1^1, a_2^1, a_3^1, a_4^1\} \cup \{d\}\},$$

$$dIS^2 = \{U \setminus \{a_1^2, a_2^2, a_3^2, a_4^2\} \cup \{d\}\},$$

$$dIS^3 = \{U \setminus \{a_1^3, a_2^3, a_3^3, a_4^3\} \cup \{d\}\}.$$

表 1 多决策的多粒度标记信息表

Table 1 Multi-granular labeled table with multi-decision

$U$	$a_1^1$	$a_2^1$	$a_3^1$	$a_1^2$	$a_2^2$	$a_3^2$	$a_1^3$	$a_2^3$	$a_3^3$	$a_4^1$	$a_4^2$	$a_4^3$	$d$
$x_1$	1	G	Y	1	G	Y	1	L	Y	1	L	Y	1
$x_2$	2	G	Y	2	G	Y	1	L	Y	1	L	Y	1
$x_3$	3	G	Y	3	G	Y	2	L	Y	2	L	Y	1
$x_4$	4	F	N	4	F	N	3	M	N	3	M	N	2
$x_5$	5	F	N	5	F	N	4	M	N	4	M	N	2
$x_6$	6	F	N	6	F	N	5	M	N	4	M	N	2
$x_7$	4	F	N	1	G	Y	3	M	N	1	L	Y	2
$x_8$	5	F	N	2	G	Y	4	M	N	1	L	Y	2
$x_9$	6	F	N	3	G	Y	5	M	N	2	L	Y	2
$x_{10}$	4	F	N	4	F	N	3	M	N	1	L	Y	2
$x_{11}$	5	F	N	5	F	N	4	M	N	1	L	Y	2
$x_{12}$	6	F	N	6	F	N	5	M	N	2	L	Y	2
$x_{13}$	4	F	N	1	G	Y	3	M	N	1	L	Y	1
$x_{14}$	5	F	N	2	G	Y	4	M	N	1	L	Y	1
$x_{15}$	6	F	N	3	G	Y	5	M	N	2	L	Y	1
$x_{16}$	1	G	Y	1	G	Y	1	L	Y	1	L	Y	3
$x_{17}$	2	G	Y	2	G	Y	1	L	Y	1	L	Y	3
$x_{18}$	3	G	Y	3	G	Y	2	L	Y	2	L	Y	3
$x_{19}$	7	B	N	7	B	N	6	H	N	5	H	N	3
$x_{20}$	8	B	N	7	B	N	6	H	N	6	H	N	3

定理 2. 设在  $t$  时刻, 不协调多粒度决策系统为  $dIS^{(t)} = (U, C \cup \{d\})$ , 且系统  $dIS^{(t)}$  的上近似最优粒度层次为第  $k^t$  粒度层次; 在  $t+1$  时刻,  $k' < k$  粒度层次下的属性值保持不变, 第  $k$  粒度层次下的属性值发生了变化,  $k'' > k$  粒度层次下的属性值也随之发生变化, 记此时的多粒度决策系统为  $dIS^{t+1} = (U, C \cup \{d\})$ , 并记此时该系统  $dIS^{t+1}$  的上近似最优粒度层次为第  $k^{t+1}$  粒度层次, 则

1) 当  $k > k^t + 1$  时, 则  $k^{t+1} = k^t$ ;

2) 当  $k \leq k^t + 1$  时,

若  $dIS^k$  关于  $dIS^{k^t+1}$  不是上近似协调的, 则  $k^{t+1} = k-1$ ;

若  $dIS^k$  关于  $dIS^{k^t+1}$  是上近似协调的, 则判断  $k+1$  的上近似协调性, 若  $dIS^{k+1}$  关于  $dIS^{k^t+1}$  是上近似协调的, 则继续判断  $k+2$  的上近似协调性, 以此循环, 直至  $k+m$  关于  $dIS^{k^t+1}$  不是上近似协调的, 则  $k^{t+1} = k+m-1$ 。

定理 3. 设在  $t$  时刻, 不协调多粒度决策系统为  $dIS^{(t)} = (U, C \cup \{d\})$ , 且系统  $dIS^{(t)}$  的分布最优粒度层次为第  $k^t$  粒度层次; 在  $t+1$  时刻,  $k' < k$  粒度层次下的属性值保持不变, 第  $k$  粒度层次下的属性值发生了变化,  $k'' > k$  粒度层次下的属性值也随之发生变化, 记此时的多粒度决策系统为  $dIS^{t+1} = (U, C \cup \{d\})$ , 并记此时该系统  $dIS^{t+1}$  的分布最优粒度层次为第  $k^{t+1}$  粒度层次, 则以下两个命题成立:

1) 当  $k > k^t + 1$  时, 则  $k^{t+1} = k^t$ ;

2) 当  $k \leq k^t + 1$  时,

若  $dIS^k$  关于  $dIS^{k^t+1}$  不是分布协调的, 则  $k^{t+1} = k-1$ ;

若  $dIS^k$  关于  $dIS^{k^t+1}$  是分布协调的, 则判断  $k+1$  的分布协调性, 若  $dIS^{k+1}$  关于  $dIS^{k^t+1}$  是分布协调的, 则继续判断  $k+2$  的分布协调性, 以此循环, 直至  $k+m$  关于  $dIS^{k^t+1}$  不是分布协调的, 则  $k^{t+1} = k+m-1$ 。

定理 4. 设在  $t$  时刻, 不协调多粒度决策系统为  $dIS^{(t)} = (U, C \cup \{d\})$ , 且系统  $dIS^{(t)}$  的最大分布最优粒度层次为第  $k^t$  粒度层次; 在  $t+1$  时刻,  $k' < k$  粒度层次下的属性值保持不变, 第  $k$  粒度层次下的属性值发生了变化,  $k'' > k$  粒度层次下的属性值也随之发生变化, 记此时的多粒度决策系统为  $dIS^{t+1} = (U, C \cup \{d\})$ , 并记此时该系统  $dIS^{t+1}$  的最大分布最优粒度层次为第  $k^{t+1}$  粒度层次, 则以下两个命题成立:

1) 当  $k > k^t + 1$  时, 则  $k^{t+1} = k^t$ ;

2) 当  $k \leq k^t + 1$  时,

若  $dIS^k$  关于  $dIS^{k^t+1}$  不是最大分布协调的, 则  $k^{t+1} = k-1$ ;

若  $dIS^k$  关于  $dIS^{k^t+1}$  是最大分布协调的, 则判断  $k+1$  的最大分布协调性, 若  $dIS^{k+1}$  关于  $dIS^{k^t+1}$  是最大分布协调的, 则继续判断  $k+2$  的最大分布似协调性, 以此循环, 直至  $k+m$  关于  $dIS^{k^t+1}$  不是最大分布协调的, 则  $k^{t+1} = k+m-1$ 。

#### 4 属性值变化环境下的不协调多粒度决策系统的 最优粒度选择算法

下面以近似协调为例给出相应的静态算法和动态算法。

算法 1. 静态: 多粒度决策系统的最优粒度选择

输入:  $t+1$  时刻求得的多粒度决策系统  $dIS^{(t+1)} = (U, C \cup \{d\})$ 。

输出: 多粒度决策表的最优粒度  $k^{t+1}$ 。

Step 1. 计算决策划分  $U/R_d = \{D_1, D_2, \dots, D_r\}$ ,  $k'=1$ ;

Step 2. 当  $k'=1: |I|$  时, 则计算  $[x]_{Ck'}$  ( $x \in U$ );

Step 3. 当  $k'=1: |I|$  时, 求出  $L_{Ck'}(d)$ ;

Step 4. 判断  $L_{Ck'}(d)$  与  $L_{C1}(d)$  的关系, 若  $L_{Ck'}(d) \neq L_{C1}(d)$ , 则  $k^{t+1} = k' - 1$ , 停止运算; 若  $L_{Ck'}(d) = L_{C1}(d)$ , 则转 Step 5。

Step 5. 若  $k'+1 > I$ , 则  $k^{t+1} = k'$ , 停止运算; 若  $k'+1 \leq I$ , 则  $k' = k'+1$ , 并转到 Step 2。

算法 2. 动态: 多粒度决策系统在属性值更新环境下的最优粒度选择

输入:  $t$  时刻求得的多粒度决策系统  $dIS^{(t)} = (U, C \cup \{d\})$  及其最优粒度层次  $k^t$  及其决策划分  $\{D_1, D_2, \dots, D_r\}$ ,  $t+1$  时刻的多粒度决策系统  $dIS^{(t+1)}$ , 属性值变化的最小粒度层次  $k$ .

输出: 多粒度决策表在属性值更新之后的最优粒度  $k^{t+1}$ .

Step 1. 判断  $k$  与  $k^t + 1$  的大小关系, 若  $k > k^t + 1$ , 则转 Step2; 若  $k \leq k^t + 1$ , 则转 Step3;

Step 2.  $k^{t+1} = k^t$ , 停止运算;

Step 3. 当  $k' = k: |I|$  时, 求出  $L_{Ck'}(d)$ ; 判断  $L_{Ck'}(d)$  与  $L_{C1}(d)$  的关系, 若  $L_{Ck'}(d) \neq L_{C1}(d)$ , 则  $k^{t+1} = k' - 1$ , 停止运算; 若  $L_{Ck'}(d) = L_{C1}(d)$ , 则转 Step 4.

Step 4. 若  $k' + 1 > I$ , 则  $k^{t+1} = k'$ , 停止运算; 若  $k' + 1 \leq I$ , 则  $k' = k' + 1$ , 并转回 Step 3.

## 5 实验分析

这部分通过在 6 个 UCI 数据集来比较静态和动态算法的计算性能. 6 个数据集的相关信息见表 2.

表 2 数据集描述

Table 2 Data set description

数据集	实例数	属性数
seeds	210	7 + 1
abalone	4177	7 + 1
winequalityred	1599	11 + 1
SGG	4070	36 + 1
shuttle	145000	9 + 1
letterrecognition	20000	16 + 1

在这个实验中, 比较了随着数据规模增大, 静态和动态算法的执行时间. 所有的算法运行在个人电脑, 配置 Windows 7, Intel(R) Core(TM) i7-3632QM CPU 2.20GHz, 4GB 内存. 编程语言 matlab.

数据设计: UCI 数据库的数据集基本上都是单粒度标记的, 为将数据处理为多粒度标记, 先对数据集进行预处理: 对于每个条件属性下原始值作为第一粒度标记属性下的值, 记为  $a_i^1$ ; 再根据第一粒度标记属性下的值进行确定第二粒度下的值, 记  $\delta_{a_i^1} = (\max_{a_i^1} - \min_{a_i^1}) / 12$ , 对象在第一粒度标记属性值  $[\min_{a_i^1}, \min_{a_i^1} + \delta_{a_i^1})$ 、 $[\min_{a_i^1} + \delta_{a_i^1}, \min_{a_i^1} + 2\delta_{a_i^1})$ 、 $[\min_{a_i^1} + 2\delta_{a_i^1}, \min_{a_i^1} + 3\delta_{a_i^1})$ 、 $\dots$ 、 $[\min_{a_i^1} + 11\delta_{a_i^1}, \min_{a_i^1} + 12\delta_{a_i^1})$  分别记为 1、2、3、 $\dots$ 、12, 作为第二粒度属性下的值; 再对第二粒度属性下的值做同样的处理, 记为第三粒度标记属性下的值; 以此类推, 依次将属性值合并, 处理成 5 个粒度标记层次.

将预处理好的每个数据集分割为等大小的 10 份. 第 1 份作为第 1 个数据集, 第 1 和第 2 份的组合作为第 2 个数据集, 第 2 个数据集和第 3 份数据的组合作为第 3 个数据集, 以此类推, 全部 10 份的数据组合作为第 10 个数据集. 10 个数据集的目标概念都使用所对应的数据集的决策类. 这 10 个数据集可以用来观察静态和动态算法执行时间随着数据集规模增大的变化情况.

在本小节中, 不失一般性, 对于所有粒度标记下属性值变

化的可能都进行了实验, 即对  $k = 1, 2, 3, 4, 5$  这五种情况都进行了静态和动态的实验. 由于  $k = 1, 2, 3, 4, 5$  时静态算法的计算时间一致, 故本文只罗列一次. 为了公平, 两个算法没有进行其它的优化.

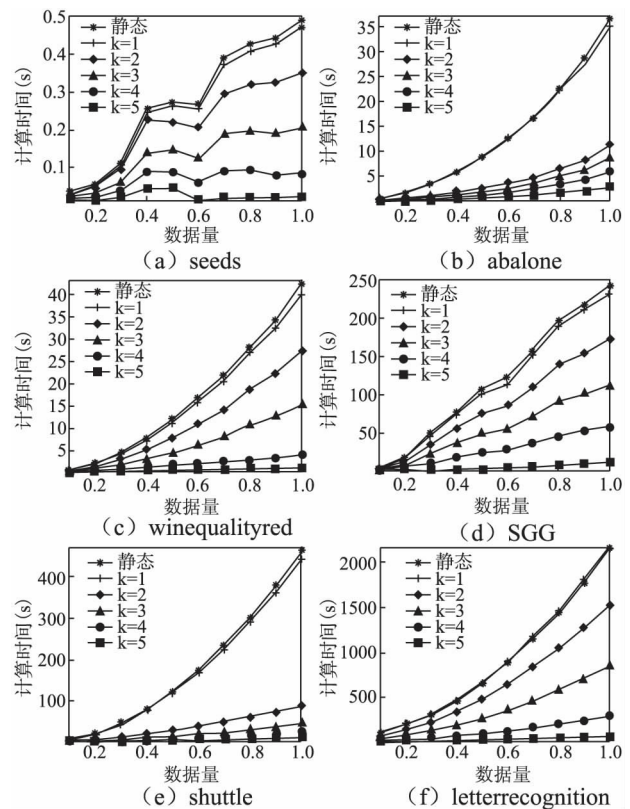


图 1 静态和动态算法最优粒度选择的计算时间比较

Fig. 1 Comparison of computation time of optimal scale selection between static and dynamic algorithms

实验结果见图 1 和表 3. 图 1 展示了两个算法计算时间随数据集规模的变化情况. 从图 1 ( $x$  轴表示数据集占总数据的比例,  $y$  轴表示计算时间) 不难发现, 对于相同大小的数据集, 当  $k \geq 2$  时静态最优粒度选择算法时间消耗一致地低于静态最优粒度选择算法. 另外, 在大多数数据中, 计算时间的差值随着数据的增加而增大.  $k = 1$  时计算时间基本持平. 表 3

表 3 静态和动态算法计算时间比较(时间: 秒)

Table 3 Comparison of computation time between static and dynamic algorithms( time: s)

数据集	seeds	abalone	winequalityred	SGG	shuttle	letterrecognition
静态 ( $k = 1, 2, 3, 4, 5$ )	0.4908	36.5331	42.1728	242.1620	460.3384	2152.3
动态 k=1	0.4714	35.1546	40.0332	232.1183	444.5350	2170.9
k=2	0.3514	11.3875	27.2826	172.5067	87.1637	1519.4
k=3	0.2043	8.7322	15.3246	112.8675	42.9623	838.0964
k=4	0.0842	5.9104	4.1220	59.0640	23.4637	294.5273
k=5	0.0212	2.9949	1.1787	12.6941	11.6658	64.4093

展示了两个算法在 6 个数据集各自产生的第 10 个数据集上的运行时间. 结果显示局部多粒度下近似算法仅仅花费了相应全局算法的十分之一执行时间.

## 6 结 语

研究了某个粒度下属性值变化时最优粒度的选择策略, 并设计了相应的静态和动态最优粒度选择算法并进行了相关实验. 实验结果表明, 新提出的方法可以对于属性值变化的系统可以有效进行选择最优粒度, 且设计的动态最优粒度选择算法在一定程度上提高了计算效率.

### References:

- [1] Zadeh L A. Fuzzy sets and information granularity [C]//Advances in Fuzzy Set Theory and Applications, Amsterdam: North-Holland Press, 1996.
- [2] Yao Yi-yu. Granular computing: basic issues and possible solutions [C]//Proceeding of the 5th Joint Conference on Information Science 2000, 1: 186-189.
- [3] Liang Ji-ye, Qian Yu-hua, Li De-yu, et al. Theory and method of granular computing for big data mining [J]. Science China: Information Science 2015, 45(11): 1355-1369.
- [4] Yao J T, Vasilakos A V, Pedrycz W. Granular computing: perspectives and challenges [J]. IEEE Trans on Cybernetics 2013, 43(6): 1977-1989.
- [5] Xu Ji, Wang Guo-yin, Yu Hong. Big data processing based on particle computing [J]. Journal of Computer Science, 2015, 38(8): 1497-1517.
- [6] Wu Wei-zhi, Leung Y. Theory and applications of granular labeled partitions in multi-scale decision tables [J]. Information Sciences, 2011, 181(18): 3878-3897.
- [7] Wu Wei-zhi, Leung Y. Optimal scale selection for multi-scale decision tables [J]. International Journal of Approximate Reasoning, 2013, 54(8): 1107-1129.
- [8] Wu Wei-zhi, Chen Ying, Xu You-hong, et al. Optimal granularity selection for consistent incomplete multi-scale decision system [J]. Pattern Recognition and Artificial Intelligence 2016, 29(2): 108-115.
- [9] Wu Wei-zhi, Chen Chao-jun, Li Tong-jun, et al. Comparison of optimal granularity for inconsistent multi-scale decision system [J]. Pattern Recognition and Artificial Intelligence 2016, 29(12): 1095-1103.
- [10] Gu Shen-ming, Gu Jin-yan, Wu Wei-zhi, et al. Optimal granularity selection for incomplete multi-scale decision systems [J]. Computer Research and Development 2017, 54(7): 99-108.
- [11] Xu You-hong, Wu Wei-zhi, Tan An-hui. Optimal scale selections in consistent generalized multi-scale decision tables [C]//Proceeding of the International Joint Conference on Rough Sets, Springer, Cham 2017: 185-198.
- [12] Wu Wei-zhi, Yang Li, Tan An-hui, et al. Granularity selection of generalized incomplete multi-granular labeled decision system [J]. Computer Research and Development 2018, 55(6): 1263-1272.
- [13] Li Feng, Hu Bao-qing. A new approach of optimal scale selection to multi-scale decision tables [J]. Information Sciences 2017, 381: 193-208.
- [14] Shi Jin-ling, Zhang Qian-qian, Xu Jiu-cheng. Optimal granularity selection for attribute reduction of multi-scale decision system [J]. Computer Science 2018, 45(2): 152-156.
- [15] Gu Shen-ming, Wan Ya-hong, Wu Wei-zhi, et al. Local optimal granularity selection for multi-scale decision system [J]. Journal of Nanjing University( Natural Science) 2016, 52(2): 280-288.
- [16] Tie Wen-yan, Fan Min, Li Jin-hai. Optimal granularity selection for multi-granular labeled decision system under object updating environment [J]. Computer Science 2018, 45(1): 113-121.
- [17] Xie Jun-peng, Yang Min-hua, Li Jin-hai, et al. Rule acquisition and optimal scale selection in multi-scale formal decision contexts and their applications to smart city [J]. Future Generation Computer Systems 2018, 83: 564-581. <https://doi.org/10.1016/j.future.2017.03.011>.
- [18] She Yan-hong, Li Jin-hai, Yang Hai-long. A local approach to rule induction in multi-scale decision tables [J]. Knowledge-Based Systems 2015, 89: 398-410. doi: 10.1016/j.knsys.2015.07.020.
- [19] Wu Wei-zhi, Qian Yu-hua, Li Tong-jun, et al. On rule acquisition in incomplete multi-scale decision tables [J]. Information Sciences, 2017, 378: 282-302.
- [20] Gu Shen-ming, Wu Wei-zhi. On knowledge acquisition in multi-scale decision systems [J]. International Journal of Machine Learning and Cybernetics 2013, 4(5): 477-486.
- [21] Fan Y N, Tseng T L, Huang C C, et al. Rule induction based on an incremental rough set [J]. Expert Systems with Applications 2009, 36(9): 11439-11450.
- [22] Luo Chuan, Li Tian-rui, Zhang Yi, et al. Matrix approach to decision-theoretic rough sets for evolving data [J]. Knowledge-Based Systems 2016, 99: 123-134. doi: 10.1016/j.knsys.2016.01.042.
- [23] Zhang Jun-bo, Li Tian-rui, Chen Hong-mei. Composite rough sets for dynamic data mining [J]. Information Sciences 2014, 257: 81-100.
- [24] Liu Dun, Li Tian-rui, Zhang Jun-bo. Incremental updating approximations in probabilistic rough sets under the variation of attributes [J]. Knowledge-Based Systems 2015, 73: 81-96.
- [25] Shu Wen-hao, Shen Hong. Updating attribute reduction in incomplete decision systems with the variation of attribute set [J]. International Journal of Approximate Reasoning 2014, 55(3): 867-884.
- [26] Chen Hong-mei, Li Tian-rui, Ruan D. Maintenance of approximations in incomplete ordered decision systems while attribute values coarsening or refining [J]. Knowledge-Based Systems 2012, 31: 140-161.
- [27] Luo Chuan, Li Tian-rui, Chen Hong-mei, et al. Fast algorithms for computing rough approximations in set-valued decision systems while updating criteria values [J]. Information Science 2015, 299: 221-242.

### 附中文参考文献:

- [3] 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.
- [5] 徐 计, 王国胤, 于 洪. 基于粒计算的大数据处理 [J]. 计算机学报 2015, 38(8): 1497-1517.
- [8] 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.
- [9] 吴伟志, 陈超君, 李同军, 等. 不协调多粒度标记决策系统最优粒度的对比 [J]. 模式识别与人工智能 2016, 29(12): 1095-1103.
- [10] 顾沈明, 顾金燕, 吴伟志, 等. 不完备多粒度决策系统的局部最优粒度选择 [J]. 计算机研究与发展 2017, 54(7): 99-108.
- [12] 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.
- [14] 史进玲, 张倩倩, 徐久成. 多粒度决策系统属性约简的最优粒度选择 [J]. 计算机科学 2018, 45(2): 152-156.
- [15] 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.
- [16] 铁文彦, 范 敏, 李金海. 对象更新环境下的多粒度决策系统的最优粒度选择 [J]. 计算机科学 2018, 45(1): 113-121.