第45卷 第1期
2018年1月
计 算 机 科 学
COMPUTER SCIENCE
Vol. 45 No. 1
Jan. 2018

对象更新环境下的多粒度决策系统的最优粒度选择

铁文彦　范　敏　李金海

（昆明理工大学理学院　昆明 650500）（昆明理工大学数据科学研究中心　昆明 650500）

摘　要　多粒度决策系统是一类重要的关系数据库，最优粒度选择是研究多粒度决策系统的主要目的之一。讨论了对象更新环境下的多粒度决策系统的最优粒度选择。首先，介绍了多粒度信息系统和多粒度决策系统；然后，引入了广义决策函数，并利用此函数定义多粒度决策系统的协调性和最优粒度；最后，讨论了对象更新环境下不同协调性的多粒度决策系统的最优粒度的变化规律。

关键词　多粒度决策系统，广义决策函数，协调性，最优粒度

中图法分类号　TP18　　**文献标识码**　A　　**DOI**　10.11896/j.issn.1002-137X.2018.01.018

Optimal Scale Selection in Multi-scale Decision Systems under Environment of Objects Updating

TIE Wen-yan　FAN Min　LI Jin-hai

(Faculty of Science, Kunming University of Science and Technology, Kunming 650500, China)

(Data Science Research Center, Kunming University of Science and Technology, Kunming 650500, China)

Abstract　Multi-scale decision system is an important kind of relational databases, and the optimal scale selection is one of the main research targets in the study of multi-scale decision system. This paper discussed the optimal scale selection of multi-scale decision systems under the environment of objects updating. First, the notions of the multi-scale information system and multi-scale decision system were presented. Then, a generalized decision function was defined and used to introduce the consistency and the optimal scale in the multi-scale decision systems. At last, the optimal scale transformation mechanism of different consistent multi-scale decision systems were investigated under the environment of objects updating.

Keywords　Multi-scale decision system, Generalized decision function, Consistency, Optimal scale

1　引言

随着信息技术的不断发展，数据规模逐渐膨胀，在面对复杂系统的问题时，经常使用粒计算这一数学方法。粒计算将整个论域划分为若干相对简单的粒，而这些粒由簇、块或集合组成。

粒计算由 Zadeh L A 首次提出，他于1979年在模糊集的基础上提出了信息粒的概念[1]，并于1997年强调了信息粒的重要性[2]。近年来，不少学者对粒计算进行了深入研究[3-8]。信息粒较好地解决了复杂问题，它可以将复杂系统抽象地转化为若干个相对简单的系统，既降低了处理难度，又提高了预测信息的准确性。

在进行数据处理时，很多信息系统中的每个对象在每个属性下只有一个属性值，研究人员只能从固定的视角或粒度来了解数据信息。目前社会已经进入信息爆炸时代，人们每天接触成千上万的信息，采集到的数据多样、多源且复杂性高[9]，为了方便管理和分析数据，需要进行数据集成。在实际应用中，多粒度决策系统[10]是一类特殊的、重要的关系数据库，它常被用于各种信息分析，如在图像处理、地理信息甄别、数据挖掘、人工智能和军事技术等领域中都有广泛的应用。

目前，已有不少学者对多粒度决策系统进行了大量研究。吴伟志等[10-11]根据对象在决策系统中拥有的不同知识粒度，提出了多粒度信息系统和多粒度决策系统，并探讨了最优粒度选择问题。顾沈明等[12]以粗糙集为基础，解释了在不同粒度下知识近似的变化规律。折延宏等[13]提出了对强协调、一般协调和完全不协调3种多粒度决策系统的规则约简。郝晨等[14]在形式背景中引入了多粒度，讨论了多标记单属性决策系统的最优粒度的选择问题。李峰等[15]将格模型和补充模型引入多粒度决策系统中，讨论了在两种模型下的最优粒度选择问题，并给出了各自的算法。

现有多粒度数据分析主要是基于静态的数据集，但是数据是可变的，很多信息系统和决策系统都是动态更新的。

到稿日期：2017-03-03　返修日期：2017-06-11　本文受国家自然科学基金资助项目(61562050,61573173)资助。

铁文彦(1991－)，男，硕士生，主要研究方向为粗糙集与粒计算，E-mail：wenyantie@163.com；**范　敏**(1975－)，女，博士，副教授，硕士生导师，主要研究方向为概念格、数据分析，E-mail：fmkmust@163.com(通信作者)；**李金海**(1984－)，男，博士，教授，主要研究方向为粗糙集、概念格与粒计算，E-mail：jhlixjtu@163.com。

因此，研究动态更新环境下的多粒度决策系统是必要的，但目前未见报道。

本文面向对象更新环境下的多粒度决策系统，通过广义决策函数讨论不同协调性的多粒度决策系统，进而选择最优粒度。

2 基础知识

定义 1 一个信息系统可以表示为二元组(U,AT)，其中$U=\{x_1,x_2,\cdots,x_n\}$为一个非空有限的对象集，称为论域；$AT=\{a_1,a_2,\cdots,a_m\}$为一个非空有限的属性集。任意$a\in AT$，满足$a:U\rightarrow V_a$，$V_a=\{a(x)|x\in U\}$，$a(x)$为对象$x$在属性$a$下的取值，$V_a$为$a$的论域。那么可以在$U$上定义二元关系$R$：

$$R=\{(x,y)\in U\times U|a(x)=a(y)\}$$

显然，R是一个等价关系，由它能够产生一个划分，记为：

$$U/R=\{[x]_R|x\in U\}$$

定义 2 一个单粒度信息系统可以表示为$S=(U,AT)$，它是一个特殊的信息系统。$AT=\{a_1,a_2,\cdots,a_m\}$为非空有限的单粒度多属性集。

定义 1 中的信息系统也是一个单粒度信息系统。

定义 3 一个多粒度信息系统可以表示为$S=(U,AT)$，其中$AT=\{a_{1i},a_{2i},\cdots,a_{mi}|i=1,2,\cdots,p\}$，$p$为$S$系统的粒度个数。

此外，一个多粒度信息系统可以被分解为m个多粒度单属性信息系统，也可以被分解为p个单粒度信息系统。当$i=1,2,\cdots,p$时，$S^i=(U,AT^i)$为多粒度信息系统S在第i粒度的单粒度信息系统。在第i个粒度下的单粒度信息系统(U,AT^i)均能产生一个划分U/AT^i；换言之，U被p次粒化。

一个多粒度信息系统应满足充分必要条件：所有的粒度可以按粗细顺序依次排列。

设$u,v\in\{1,2,\cdots,p\}$，第u粒度和第v粒度可以产生划分U/AT^u和U/AT^v，若$U/AT^u\sqsubseteq U/AT^v$，则第u粒度比第v粒度细。

为了方便叙述，本文假设每个多粒度信息系统的粒度都是从细到粗，且每个属性下的粒度都是恒定的。

一个粒度从细到粗排列的多粒度信息系统有如下性质：

$$U/AT^1\sqsubseteq U/AT^2\sqsubseteq\cdots\sqsubseteq U/AT^p$$

定义 4 一个多粒度决策系统是由一个多粒度信息系统与一个决策属性d相结合得到的，它可以表示为$S=(U,AT\cup\{d\})=(U,\{a_{1i},a_{2i},\cdots,a_{mi}|i=1,2,\cdots,p\}\cup(d))$，其中$d\notin AT$。$AT$是$S$的条件属性集，$d$是$S$的决策属性。此外，$S^i=(U,AT^i\cup\{d\})=(U,\{a_{1i},a_{2i},\cdots,a_{mi}\}\cup\{d\})$是第$i$粒度下的单粒度决策系统。

一个多粒度决策系统可以被分解为p个单粒度决策系统$S^i=(U,AT\cup\{d\})=(U,\{a_{1i},a_{2i},\cdots,a_{mi}\}\cup\{d\})$，$i=1,2,\cdots,p$。

利用$d(x)$表示对象x在决策属性d下的取值，可以定义一个等价关系：

$$R_d=\{(x,y)\in U\times U|d(x)=d(y)\}$$

根据上式，对象x在R_d的等价类可以表示为：

$$[x]_{R^d}=\{y\in U|(x,y)\in R_d\}$$

例 1 表 1 是一个多粒度决策系统$S=(U,AT\cup\{d\})$。对象集$U=\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9\}$表示 9 个人；条件属性集$AT=\{a_{1i},a_{2i},a_{3i}|i=1,2,\cdots,4\}$表示在不同粒度下 3 类信用程度的评分，决策属性表示信用是否合格。此多粒度决策系统有 4 个粒度，第 1 粒度的单粒度决策系统$S^1=(U,AT^1\cup\{d\})$以百分制记录信用评分，第 2 粒度下的单粒度决策系统$S^2=(U,AT^2\cup\{d\})$以十级制记录信用评分，第 3 粒度下的单粒度决策系统$S^3=(U,AT^3\cup\{d\})$以三级制记录信用评分，第 4 粒度下的单粒度决策系统$S^4=(U,AT^4\cup\{d\})$以二级制记录信用评分。表 2－表 5 是多粒度决策系统S各个粒度下的单粒度决策系统。

如果将决策属性d去掉，那么每个粒度下的决策系统也是一个信息系统。

表 1 例 1 中的多粒度决策表$S=(U,AT\cup\{d\})$

Table 1 Multi-granularity decision table $S=(U,AT\cup\{d\})$ in example 1

U	a_{11}	a_{12}	a_{13}	a_{14}	a_{21}	a_{22}	a_{23}	a_{24}	a_{31}	a_{32}	a_{33}	a_{34}	d
x_1	11	1	L	N	21	2	L	N	48	4	L	N	－
x_2	32	3	L	N	49	4	L	N	55	5	L	N	－
x_3	78	7	M	Y	91	9	H	Y	84	8	H	Y	＋
x_4	24	2	L	N	35	3	L	N	17	1	L	N	－
x_5	59	5	L	N	13	1	L	N	27	2	L	N	－
x_6	84	8	H	Y	75	7	M	Y	91	9	H	Y	＋
x_7	63	6	M	Y	81	8	H	Y	62	6	M	Y	＋
x_8	48	4	L	N	53	5	L	N	39	3	L	N	－
x_9	95	9	H	Y	62	6	M	Y	73	7	M	Y	＋

表 2 表 1 的第 1 粒度下的决策系统

Table 2 Decision system under the first level of granularity of Table 1

U	a_{11}	a_{21}	a_{31}	d
x_1	11	21	48	－
x_2	32	49	55	－
x_3	78	91	84	＋
x_4	24	35	17	－
x_5	59	13	27	－
x_6	84	75	91	＋
x_7	63	81	62	＋
x_8	48	53	39	－
x_9	95	62	73	＋

表 3 表 1 的第 2 粒度下的决策系统

Table 3 Decision system under the second level of granularity of Table 1

U	a_{12}	a_{22}	a_{32}	d
x_1	1	2	4	－
x_2	3	4	5	－
x_3	7	9	8	＋
x_4	2	3	1	－
x_5	5	1	2	－
x_6	8	7	9	＋
x_7	6	8	6	＋
x_8	4	5	3	－
x_9	9	6	7	＋

表 4　表 1 的第 3 粒度下的决策系统

Table 4　Decision system under the third level granularity of Table 1

U	a_{13}	a_{23}	a_{33}	d
x_1	L	L	L	−
x_2	L	L	L	−
x_3	M	H	H	+
x_4	L	L	L	−
x_5	L	L	L	−
x_6	H	M	H	+
x_7	M	H	M	+
x_8	L	L	L	−
x_9	H	M	M	+

表 5　表 1 的第 4 粒度下的决策系统

Table 5　Decision system under the fourth level granularity of Table 1

U	a_{14}	a_{24}	a_{34}	d
x_1	N	N	N	−
x_2	N	N	N	−
x_3	Y	Y	Y	+
x_4	N	N	N	−
x_5	N	N	N	−
x_6	Y	Y	Y	+
x_7	Y	Y	Y	+
x_8	N	N	N	−
x_9	Y	Y	Y	+

3　多粒度决策系统的协调性与最优粒度

为了更好地研究和说明多粒度决策系统的协调性和最优粒度，引入一个广义决策函数：

$$\partial(x)=\{d(y)\mid y\in[x]_{AT}\}$$

其中，$\partial(x)$ 是指对象 x 所在类在决策属性 d 下的取值的集合。

在多粒度决策系统 $S=(U,AT\cup\{d\})=(U,\{a_{1i},a_{2i},\cdots,a_{mi}\mid i=1,2,\cdots,p\}\cup\{d\})$ 中，此函数变更为 $\partial_{AT^i}(x)=\{d(y)\mid y\in[x]_{AT^i}\}$，指在第 i 粒度下的单粒度决策系统中对象 x 所在类在决策属性 d 下的取值的集合。

根据多粒度决策系统的定义，可得：

$$\partial_{AT^1}(x)\subseteq\partial_{AT^2}(x)\subseteq\cdots\subseteq\partial_{AT^p}(x),x\in U$$

定义 5　设 $S=(U,AT\cup\{d\})=(U,\{a_{1i},a_{2i},\cdots,a_{mi}\mid i=1,2,\cdots,p\}\cup\{d\})$ 是一个完全协调的多粒度决策系统，那么 $R_{AT^p}\subseteq R_d$。当 $1\leqslant k<i\leqslant p$ 时，若 $S^i=(U,AT^i\cup\{d\})$ 为一个协调的单粒度决策系统，即 $R_{AT^i}\subseteq R_d$，那么 $R_{AT^k}\subseteq R_{AT^i}\subseteq R_d$。因此，$(U,AT^k\cup\{d\})$ 也是一个协调的单粒度决策系统。根据广义决策函数的定义，在一个完全协调的多粒度决策系统中，当 $i\in\{1,2,\cdots,p\}$ 时，在每个粒度下的广义决策函数值的数量恒为 1，即：

$$card(\partial_{AT^i}(x))\equiv 1$$

定义 6　设 $S=(U,AT\cup\{d\})=(U,\{a_{1i},a_{2i},\cdots,a_{mi}\mid i=1,2,\cdots,p\}\cup\{d\})$ 为一个部分协调的多粒度决策系统，则必有一个粒度 k 使 S^k 是协调的且 S^{k+1} 是不协调的 $(k+1\leqslant p)$。

根据广义决策函数的定义，在一个部分协调的多粒度决策系统中，必有：

$$card(\partial_{AT^k}(x))\equiv 1，且\ card(\partial_{AT^{k+1}}(x))\not\equiv 1$$

定义 7　设 $S=(U,AT\cup\{d\})=(U,\{a_{1i},a_{2i},\cdots,a_{mi}\mid i=1,2,\cdots,p\}\cup\{d\})$ 为一个完全不协调的多粒度决策系统，那么 S^p 是完全不协调的，当然也没有任何一个粒度下的单粒度决策系统 S^i 是协调的。

根据广义决策函数的定义，在一个完全不协调的多粒度决策系统中，均有 $card(\partial_{AT^i}(x))\not\equiv 1$。

综上所述，判断一个多粒度决策系统的协调性时，只需要知道首尾粒度下的单粒度决策系统的协调性，即最细的粒度和最粗的粒度。

多粒度决策系统的知识获取是相对困难的。在此系统中，每个对象的特征表达方式复杂且多样。本文主要讨论的是广义决策函数上的多粒度决策系统的协调性和最优粒度。为了方便叙述，将其简称为多粒度决策系统的协调性和最优粒度。

根据以上对 3 种协调性的多粒度决策系统的定义，容易得到 3 种协调性的多粒度决策系统的最优粒度定义。

定义 8　设 $S=(U,AT\cup\{d\})=(U,\{a_{1i},a_{2i},\cdots,a_{mi}\mid i=1,2,\cdots,p\}\cup\{d\})$ 为一个多粒度决策系统。若 S 是完全协调的，则它的最优粒度为第 p 粒度；若 S 是部分协调的，则它的最优粒度为第 k 粒度，必有 S^k 是协调的，S^{k+1} 是不协调的；若 S 是完全不协调的，则没有粒度是协调的，不存在最优粒度，但为了讨论方便，最优粒度取为第 1 粒度。

虽然多粒度决策系统在现实中有广泛应用，但是研究人员面对庞大的数据时更倾向于选择最好的单粒度的决策系统来解决实际问题。因此，最优粒度选择问题是至关重要的。

4　对象更新环境下多粒度决策系统的最优粒度选择

第 3 节介绍了多粒度决策系统的协调性和最优粒度的概念。在现实世界中，多粒度决策系统会随着时间的变化而进行数据的更新和优化。因此，本节讨论对象更新环境下的多粒度决策系统的最优粒度选择问题。其中，对象更新指更新单个对象，且不增加新的条件属性值和决策属性值。

令初始的多粒度决策系统为 $S^{(t)}$，即在时间 t 时的多粒度决策系统。更新之后的多粒度决策系统为 $S^{(t+1)}$，即在时间 $t+1$ 时的多粒度决策系统。$S^{(t+1)}=(U^{(t+1)},AT\cup\{d\})$，$U^{(t+1)}=U^{(t)}\cup\{x'\}$，$\{x'\}$ 为更新过程中新增加的对象。

令初始的多粒度决策系统的决策划分为 $U^{(t)}/R_d=\{D_1^{(t)},D_2^{(t)},\cdots,D_s^{(t)}\}$，对象更新后的多粒度决策系统的决策划分为 $U^{(t+1)}/R_d=\{D_1^{(t+1)},D_2^{(t+1)},\cdots,D_s^{(t+1)}\}$。当初始多粒度决策系统更新后，决策等价类的数量是不会改变的。$S^{(t+1)}$ 由 $S^{(t)}$ 和更新对象 $\{x'\}$ 构成，决策划分 $U^{(t)}/R_d$ 也更新为 $U^{(t+1)}/R_d$，如下所示：

$$D_j^{(t+1)}=\begin{cases}D_j^{(t)}\cup\{x'\},d(x)=d(x'), & x\in D_j^{(t)}\\ D_j^{(t)}, & 其他\end{cases}$$

命题 1　设 $S=(U,AT\cup\{d\})=(U,\{a_{1i},a_{2i},\cdots,a_{mi}\mid i=1,2,\cdots,p\}\cup\{d\})$ 是一个完全协调的多粒度决策系统，那么

$S^{(t+1)}=(U^{(t+1)},AT\cup\{d\})$为一个完全协调或部分协调或完全不协调的多粒度决策系统。

上述命题显然成立,证略。

命题2 设$S^{(t)}=(U^{(t)},AT\cup\{d\})=(U,\{a_{1i},a_{2i},\cdots,a_{mi}\mid i=1,2,\cdots,p\}\cup\{d\})$是一个部分协调的多粒度决策系统,经过对象更新,更新对象为x',那么$S^{(t+1)}=(U^{(t+1)},AT\cup\{d\})$为一个部分协调或完全不协调的多粒度决策系统。

证明:(反证法)假设$S^{(t)}=(U^{(t)},AT\cup\{d\})=(U^{(t)},\{a_{1i},a_{2i},\cdots,a_{mi}\mid i=1,2,\cdots,p\}\cup\{d\})$为一个部分协调的多粒度决策系统,经过对象更新,$S^{(t+1)}=(U^{(t+1)},AT\cup\{d\})$为一个完全协调的多粒度决策系统。

根据定义6,在$S^{(t)}$中必有一个粒度k令S_k协调,而且S_{k+1}是不协调的。在对象更新之前,从第$k+1$粒度到第p粒度的单粒度决策系统是不协调的,即$k+1\leqslant q\leqslant p$时,存在$x\in U$使得广义决策函数值$card(\partial^{(t)}_{AT^q}(x))\neq1$。在对象更新之后,新增对象$x'$加入等价类中,依然不能令广义决策函数值的数量恒等于1。因此,假设不成立,即原命题成立。

命题3 设$S^{(t)}=(U^{(t)},AT\cup\{d\})$为一个完全不协调的多粒度决策系统,经过对象更新,更新对象为x',$S^{(t+1)}=(U^{(t+1)},AT\cup\{d\})$只能是一个完全不协调的多粒度决策系统。

证明过程类似于命题2的证明。

命题4 设$S^{(t)}=(U^{(t)},AT\cup\{d\})$为一个完全协调的多粒度决策系统,$S^{(t+1)}=(U^{(t+1)},AT\cup\{d\})$为一个完全协调的多粒度决策系统。二者的最优粒度分别为r和r',那么$r=r'$。

证明:根据定义5,在完全协调的多粒度决策系统中,$card(\partial_{AT^i}(x))=1$,最优粒度是粒度最大值。因为$S^{(t)}$和$S^{(t+1)}$都是完全协调的,所以$r=r'$。

命题5 设$S^{(t)}=(U^{(t)},AT\cup\{d\})$为一个完全协调的多粒度决策系统,$S^{(t+1)}=(U^{(t+1)},AT\cup\{d\})$为一个部分协调的多粒度决策系统。二者的最优粒度分别为r和r',那么$r>r'$。

证明:根据定义8,一个完全协调的多粒度决策系统的最优粒度是粒度最大值,一个部分协调的多粒度决策系统的最优粒度为k。因为$r=p,r'=k$,但$1<k<p$,所以$r>r'$。

命题6 设$S^{(t)}=(U^{(t)},AT\cup\{d\})$为一个部分协调的多粒度决策系统,经过对象更新,更新对象为x',更新后的多粒度决策系统$S^{(t+1)}=(U^{(t+1)},AT\cup\{d\})$也为一个部分协调的多粒度决策系统。二者的最优粒度分别为r和r',那么$r\geqslant r'$。

证明:(反证法)假设$S^{(t)}=(U^{(t)},AT\cup\{d\})$为一个部分协调的多粒度决策系统,经过对象更新,更新对象为x',更新后的多粒度决策系统$S^{(t+1)}=(U^{(t+1)},AT\cup\{d\})$也为一个部分协调的多粒度决策系统。二者的最优粒度分别为r和r',那么$r<r'$。

在$S^{(t)}$中,$card(\partial_{AT^k}(x))\equiv1$,且$card(\partial_{AT^{k+1}}(x))\neq1$。增加对象$x'$后,$x'$的最优粒度为第$p$粒度[13],$r'=r$,与假设矛盾。因此,假设不成立,命题6成立。

命题7 设$S^{(t)}=(U^{(t)},AT\cup\{d\})$为一个部分协调的多粒度决策系统,经过对象更新,更新对象为x',更新后的多粒度决策系统$S^{(t+1)}=(U^{(t+1)},AT\cup\{d\})$为一个完全不协调的多粒度决策系统,二者的最优粒度分别为r和r',那么$r\geqslant r'$。

本文主要研究完全协调的多粒度决策系统和部分协调的多粒度决策系统在对象更新之后的最优粒度变化。对于完全不协调的多粒度决策系统在对象更新环境下的最优粒度变化,仅在后续算法中体现。

当$S^{(t)}$为完全协调的多粒度决策系统时,经过对象更新后,增加对象为x'。

(1)若$card(\partial_{AT^i}(x'))\equiv1$,那么$S^{(t+1)}$是完全协调的。

(2)若$card(\partial_{AT^k}(x'))\equiv1$,$card(\partial_{AT^{k+1}}(x'))\not\equiv1$,当且仅当$card(\partial_{AT^1}(x'))\equiv1$,那么$S^{(t+1)}$是部分协调的$(1\leqslant k<p)$。

(3)若$card(\partial_{AT^1}(x'))\not\equiv1$,那么$S^{(t+1)}$是完全不协调的。

当$S^{(t)}$为部分协调的多粒度决策系统时,经过对象更新后,增加对象为x'。

(4)若$card(\partial_{AT^i}(x'))\equiv1$,那么$S^{(t+1)}$是部分协调的。

(5)若$card(\partial_{AT^k}(x'))\equiv1$,$card(\partial_{AT^{k+1}}(x'))\not\equiv1$,当且仅当$card(\partial_{AT^1}(x'))\equiv1$,那么$S^{(t+1)}$是部分协调的$(1\leqslant k<p)$。

(6)若$card(\partial_{AT^1}(x'))\not\equiv1$,那么$S^{(t+1)}$是完全不协调的。

下面给出一个寻找最优粒度的算法,此算法以广义决策函数为依托,以广义决策函数值的数目来判断协调性,并利用协调性的定义求出最优粒度。

算法1 对象更新环境下的多粒度决策系统的最优粒度选择

输入:一个多粒度决策系统$S^{(t)}$及它的粒度数p,新增对象x′及新增对象在条件属性和决策属性下的取值

输出:对象更新之后的多粒度决策表的最优粒度r′

Step1 S^1和$S^p\leftarrow S^{(t)}$,$r\leftarrow S^{(t)}$。

Step2 若S^1和S^p都是协调的,则转Step3;若S^1是协调的,但S^p是不协调的,则转Step4;若S^1和S^p都是不协调的,则$r=r'=1$。

Step3 若x′满足(1),r′=r=p;若x′满足(2),找到满足情况(2)的最小k值,r′=k;若x′满足(3),r′=1。

Step4 若x′满足(4),r′=r;若x′满足(5),则转Step5;若x′满足(6),r′=1。

Step5 找到满足情况(5)的最小k值,若r>k,则r′=k;若r≤k,则r′=r。

下面给出例2来进一步说明此算法。

例2 表1给出了一个初始多粒度决策系统。现对此决策系统进行对象更新,第一次的更新对象为x_{10},第二次的更新对象为x_{11},第三次的更新对象为x_{12}。表6是经过3次对象更新后的多粒度决策系统。第一次的更新对象为x_{10},$S^{(t)}$的粒度数为4,每个粒度下的单粒度决策系统分别如表2-表5所列。将x_{10}加入$S^{(t)}$,x_{10}皆属于第(2)种情况,那么$r'=2$。第二次的更新对象为x_{11},$S^{(t)}$是经过第一次更新对象之后的多粒度决策系统。将x_{11}加入$S^{(t)}$,$S^{(t+1)}$是更新之后的多粒度决策表。x_{11}属于第(5)种情况,那么$r'=2$。第三次的更新对象为x_{12},$S^{(t)}$是经过第二次更新对象之后的多粒度决策系统。将x_{12}加入$S^{(t)}$,$S^{(t+1)}$是更新之后的多粒度决策表。x_{12}属于第(6)种情况,$S^{(t+1)}$是完全不协调的,那么$r'=1$。

表 6　更新对象后的多粒度决策系统

Table 6　Multi-scale decision system after objects updating

U	a_{11}	a_{12}	a_{13}	a_{14}	a_{21}	a_{22}	a_{23}	a_{24}	a_{31}	a_{32}	a_{33}	a_{34}	d
x_1	11	1	L	N	21	2	L	N	48	4	L	N	−
x_2	32	3	L	N	49	4	L	N	55	5	L	N	−
x_3	78	7	M	Y	91	9	H	Y	84	8	H	Y	+
x_4	24	2	L	N	35	3	L	N	17	1	L	N	−
x_5	59	5	L	N	13	1	L	N	27	2	L	N	−
x_6	84	8	H	Y	75	7	M	Y	91	9	H	Y	+
x_7	63	6	M	Y	81	8	H	Y	62	6	M	Y	+
x_8	48	4	L	N	53	5	L	N	39	3	L	N	−
x_9	95	9	H	Y	62	6	M	Y	73	7	M	Y	+
x_{10}	58	5	L	N	45	4	L	N	59	5	L	N	+
x_{11}	61	6	M	Y	83	8	H	Y	75	7	M	Y	−
x_{12}	24	2	L	N	35	3	L	N	17	1	L	N	+

5　数值实验

为了验证上述算法的可行性与有效性,从 UCI 数据库中选出一些数据集进行数值模拟。这些数据集分别为 Lenses,HayesRoth_1,HayesRoth_2,Haberman's Survival,Airfoil Self Noise,Combined Cycle Power Plant,CarEvaluation,Seeds 和 BalanceScale。

在每个多粒度决策系统中,每个对象根据不同的粒度在每个条件属性下都有不同的取值。但这些数据集都是以单粒度的形式表达对象在各个属性下的取值,因此对这些数据集进行属性值的扩展是非常有必要的。

在进行数据预处理时,有一些数据集没有决策属性,可以选择一条属性进行多值化处理,并且能使其他分类属性形成完全协调或部分协调或完全不协调的多粒度决策系统。在对条件属性进行扩展时,若条件属性值为离散型,则尽量对属性值进行合并,保证其不重复;若条件属性值为连续型,则将属性值划分为多个区间,并将每个区间看作一个属性值。

本文考虑增加对象的个数为 1,但在实际应用中,一个多粒度决策系统在对象更新环境下往往有多个增加对象。如果增加对象较多,可将新增对象分为几批,对上述算法进行多次迭代,即可得到相应的结果。

每一个多粒度决策系统的规模较大,每次更新的对象相对较少。下面是对上述算法进行数值模拟的实验结果,其中对每个多粒度决策系统增加 10 个对象。表 7 列出了数据集经过预处理之后的特征,表 8 列出了表 7 中数据集增加对象后运行算法的实验结果。

表 7　原数据集的特征

Table 7　Characteristics of original datasets

数据集	对象数	属性规模	协调性及最优粒度
Lenses	24	3×2+1	完全协调,2
HayesRoth_1	132	3×5+1	完全协调,5
HayesRoth_2	132	3×5+1	部分协调,2
Haberman's Survival	306	3×3+1	部分协调,1
Airfoil Self Noise	1503	6×3+1	部分协调,1
Combined Cycle Power Plant	9568	5×3+1	完全不协调,1
CarEvaluation	1728	6×3+1	部分协调,2
Seeds	210	7×3+1	部分协调,1
BalanceScale	625	4×3+1	部分协调,2

表 8　更新之后的数据集的最优粒度选择的实验结果

Table 8　Experimental results of optimal scale selection of datasets after objects updating

数据集	协调性	最优粒度	运行时间/s
Lenses	部分协调	1	0.0016
HayesRoth_1	完全协调	5	0.0082
HayesRoth_2	部分协调	2	0.0091
Haberman's Survival	部分协调	1	0.0212
Airfoil Self Noise	完全不协调	1	1.6472
Combined Cycle Power Plant	完全不协调	1	83.5628
CarEvaluation	部分协调	2	2.1454
Seeds	完全不协调	1	0.0225
BalanceScale	部分协调	1	0.2124

结束语　本文主要介绍以广义决策函数描述多粒度决策系统的 3 种协调性以及对象更新环境下的多粒度决策系统的最优粒度选择。在对象更新过程中,只增加对象,不增加条件属性值和决策属性值。增加新的条件属性值和决策属性值后的相关问题,还有待后续进一步研究。

参 考 文 献

[1] ZADEH L A. Fuzzy sets and information granularity[M]// Fuzzy Sets Fuzzy Logic & Fuzzy Systems. 1979:433-448.

[2] ZADEH L A. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic[J]. Fuzzy Sets and Systems,1997,90(2):111-127.

[3] LIN T Y. Granular computing:structures,representations,and application[C]// RSFDGrC'03. 2003:16-24.

[4] YAO Y Y. Granular Computing:Basic Issues and Possible Solutions[C]// Proceedings of the 5 Joint Conference on Information Sciences. 2000:186-189.

[5] 张钹,张玲. 问题求解理论及应用[M]. 北京:清华大学出版社,1990.

[6] BARGIELA A,PEDRYCZ W. Toward a theory of granular computing for human-centered information processing[J]. IEEE Transaction on Fuzzy Systems,2008,16(2):320-330.

[7] HU Q H,LIU J F,YU D R. Mixed feature selection based on granulation and approximation[J]. Knowledge-Based Systems,2008,21(4):294-304.

[8] QIAN Y H,LIANG J Y,DANG C Y. Knowledge structure,knowledge granulation and knowledge distance in a knowledge base[J]. International Journal of Approximate Reasoning,2009,50(1):174-188.

(下转第 121 页)

算法复杂度分析：文献[8]提出的并行化简算法在不同输入下利用矩阵相乘进行迭代运算，当迭代次数为ω时，时间复杂度为$O(|X||\omega|^2)$；文献[6]提出了基于粒计算理论的状态化简算法，该算法的主要时间开销为等价类的求交过程，时间复杂度为$O(|\omega|^2)$；本文提出基于等价关系的状态化简算法，只需比较矩阵中的相同列即可，时间复杂度为$O(|\omega|)$，得到大幅降低。

结束语　本文提出了一种基于等价关系的状态化简算法，与传统的隐含表法或根据输出矩阵和状态转移矩阵相乘进行迭代计算的方法不同，首先定义了输出矩阵，根据输出矩阵得到初级分类并标记，然后又定义了初态标记矩阵和次态标记矩阵并构建状态转移系统矩阵，通过寻找相同的列即可得到等价类的划分。本文的创新之处在于：1)将粒计算理论中的分层粒化思想在时序逻辑电路状态化简问题中进行了新的定义和应用；2)将对应多个输入的相同行输出矩阵定义为只有一行的初态标记矩阵，将状态转移矩阵由大规模稀疏矩阵转化为规模很小的次态标记矩阵，大大减少了矩阵的存储空间；3)每次根据状态转移系统矩阵得到的等价类划分即为该次迭代的最终划分结果，避免了其他算法中等价类的求交运算；4)算法基于矩阵求解，但避免了矩阵相乘，只需判断状态转移系统矩阵中相同的列即可，降低了算法复杂度。下一步可以尝试将该算法应用于求取不完全确定时序逻辑电路[10]中的最大相容类[11]和最小闭覆盖[12]，并进行状态化简。

参考文献

[1] 刘宝琴，罗嵘，王德生．数字电路与系统(第2版)[M]．北京：清华大学出版社，2007．

[2] 刘培植．数字电路与逻辑设计[M]．北京：北京邮电大学出版社，2009．

[3] ZHAO H，SUN F R．Discussion on State Reduction Method in Synchronous Sequential Circuit Design [J]．Automation and Instrumentation，2010(5)：120-122．(in Chinese)
赵贺，孙凤茹．同步时序电路设计中状态化简方法探讨[J]．自动化与仪器仪表，2010(5)：120-122．

[4] 边计年，薛宏熙，苏明，等．数字系统设计自动化(第2版)[M]．北京：清华大学出版社，2005．

[5] CHANG Q M，YUE C Q．A Graph Theory Method for State Reduction of Digital Systems [C]// Academic Annual Conference on Electrician Theory．2006．(in Chinese)
常青美，岳彩青．一种数字系统状态化简的图论方法[C]//电工理论学术年会．2006．

[6] ZHANG K Y，ZHANG Y，CHEN Z H．GrC-Based State Reduction Algorithm for Completely Specified Sequential Logic Circuit [J]．Journal of Chinese Computer Systems，2016，37(8)：1786-1789．(in Chinese)
张凯英，张裕，陈泽华．基于粒计算的完全确定时序逻辑电路状态化简算法[J]．小型微型计算机系统，2016，37(8)：1786-1789．

[7] 苗夺谦，王国胤，刘清，等．粒计算：过去，现在与展望[M]．北京：科学出版社，2007．

[8] WANG W Z．A Parallel Algorithm for State Simplification[C]// CAD/CG 1988．1988：831-838．(in Chinese)
王文章．状态化简的一个并行算法[C]//全国计算机辅助设计与图形学学术会议．1988：831-838．

[9] LU Y P，LIN Y P，YANG G Z，et al．A State Minimization Algorithm for Completely Specified Sequential Machine[J]．Journal of Hunan University，1997，24(1)：97-102．(in Chinese)
陆应平，林亚平，杨贯中，等．完全定义时序机的状态化简算法[J]．湖南大学学报 1997，24(1)：97-102．

[10] PAULL M C，UNGER S H．Minimizing the Number of States in Incompletely Specified Sequential Switching Functions[J]．Ire Transactions on Electronic Computers，1959，EC-8(3)：356-367．

[11] DAMIANI M．The state reduction of nondeterministic finite-state machines[J]．IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems，1997，16(11)：1278-1291．

[12] GRASSELLI A，LUCCIO F．A method for minimizing the number of internal states in incompletely specified machines[OL]．http://www.researchgate.net/publication/238705600_A_method_for_minimizing_the_number_of_internal_states_in_incompletely_specified_machines．

(上接第117页)

[9] MENG X F，CI X．Big data management：concepts，techniques and challenges[J]．Journal of Computer Research and Development，2013，50(1)：146-169．(in Chinese)
孟小峰，慈祥．大数据管理：概念、技术与挑战[J]．计算机研究与发展，2013，50(1)：146-169．

[10] 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．

[11] WU W Z，LEUNG Y．Optimal scale selection for multi-scale decision tables[J]．International Journal of Approximate Reasoning，2013，54(8)：1107-1129．

[12] GU S M，WU W Z，XU Y H．On granulation in incomplete multi-labeled information systems[J]．Journal of Nanjing University(Natural Sciences)，2013，49(5)：567-573．(in Chinese)
顾沈明，吴伟志，徐优红．不完备多标记信息系统中粒度研究[J]．南京大学学报(自然科学)，2013，49(5)：567-573．

[13] SHE Y H，LI J H，YANG H L．A local approach to rule induction in multi-scale decision tables [J]．Knowledge-Based Systems，2015，89(C)：398-410．

[14] HAO C，FAN M，LI J H，et al．Optimal scale selection in multi-scale selection in multi-scale contexts based on granular scale rules[J]．Pattern Recognition and Artificial Intelligence，2016，29(3)：272-280．(in Chinese)
郝晨，范敏，李金海，等．多标记背景下基于粒规则的最优标记选择[J]．模式识别与人工智能，2016，29(3)：272-380．

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