嵌套决策粒度约简关系的研究*)

李道国[1,2] 苗夺谦[1] 张年琴[1]

(同济大学计算机科学与技术系 上海 200092)[1] (太原理工大学阳泉学院 山西阳泉 045001)[2]

摘 要 在基于粗糙集理论的知识发现中,知识约简是核心内容。因此,本文从理论上分析了相容决策表中嵌套决策粒度约简的关系,得出并证明了粗决策粒度的相对D核一定是细决策粒度相对D核的子集,粗决策粒度的一个相对D约简在满足相容性的条件下一定可以扩张成细决策粒度的一个约简。反之,细决策粒度的一个约简一定可以缩减为粗决策粒度的一个约简。研究结果对基于决策表的知识约简和知识发现有一定的实际意义。

关键词 相对D核,相对D约简,包容强度,偏序关系,粒划分

Research on Reduction Relationships of Nested Granularities

LI Dao-Guo[1,2] MIAO Duo-Qian[1] ZHANG Nian-Qin[1]

(Department of Computer Science and Technology, Tongji University, Shanghai 200092)[1]

(Yangquan College of Taiyuan University of Technology, Yangquan 045001)[2]

Abstract Knowledge Reduction plays an important role in knowledge discovery based on rough set theory. In this paper, therefore, the reduction relationships of nested granularities are theoretically analyzed in consistent information tables. We educe and prove that relative core of coarser decision granularity must be a subset of relative core of finer decision granularity. At the same time, a relative reduction of coarser decision granularity that satisfies consistency can be extended into a relative reduction of its finer decision granularity. Inversely, a relative reduction of finer decision granularity can also be cut back a relative reduction of its coarser decision granularity. The conclusions are helpful for knowledge reduction and knowledge discovery based on rough sets.

Keywords Relative core, Relative reduction, Inclusion degree, Information granular partition

1 引言

基于决策表的知识表达系统一般包含了某一领域大量的数据和信息记录,这些记录可构成领域的实例关系数据库,且在一定程度上反映条件属性与决策的关系,它们是领域知识的载体。知识发现(knowledge discovery in database,缩写KDD)的目的是通过分析知识表达系统,从模糊的、不精确的、不完整的和海量的数据和信息中获取潜在的、新颖的、正确的和有利用价值的知识,然后进一步用于求解该领域的相关问题。知识发现的过程本质上是一个动态过程,因此不仅要研究静态决策系统的知识发现的方法,还需要研究决策系统变化情况下的知识发现的修正方法与递推方法,从而不断完善信息处理的理论、方法与技术。由于知识约简是知识发现的核心内容和算法设计的关键技术,因此本文着重分析研究了基于相容决策表的嵌套决策粒度约简的关系,得出并证明了粗决策粒度的相对D核一定包含在细决策粒度的相对D核之中,粗决策粒度的一个相对D约简在满足相容性的条件下一定可以扩张成细决策粒度的一个约简。反之,细决策粒度的一个约简一定可以缩减为粗决策粒度的一个约简。研究结果对基于决策表的知识约简、决策分析和动态求解问题有一定的实际意义。

2 基本的概念

定义1 知识表达系统KRS 形式化地称一个七元组(U,At,V,f,L,N,Γ)是一个广义的知识表达系统,其中,

U:对象的非空有限集合,称为论域。

At:属性的集合,一般情形下可表示为$At=C\cup D$(C为条件属性集,D为决策属性集)。(i)若$D=\phi$,则称知识表达系统是一个信息系统;(ii)若满足条件$C\cap D=\phi, C\neq\phi\wedge D\neq\phi$,则称知识表达系统是一个决策系统(决策表或信息表)。

V:属性集的值域,$V=\bigcup_{\forall a_i\in A_t}V_{a_i}$,$V_{a_i}$表示属性$a_i$的值域,属性的值域可以是定量的值或定性值。由于在知识发现的问题中主要是分类问题,我们用不同的数值替代不同的定性值,并不会影响知识发现的过程与结果,因此属性值域可以完全数量化。

f:信息函数(关系集),对$\forall a_i\in At$,$f_{a_i}:U\to V_{a_i}$,即

$$f_{a_i}(x_l)=v,\ \forall x_l\in U, v\in V_{a_i} \tag{1}$$

在知识表达系统中,信息函数非常重要。如果它不存在,对象集与属性集之间就是孤立的,因此信息函数表达了二者之间的关联,这正是知识发现所需要的信息基础。现将信息函数的定义拓广到属性的子集上,即对$\forall B\subseteq At$,$f_B:U\to V_{a'_1}\times V_{a'_2}\cdots\times V_{a'_k}$,其中属性子集$B=\{a'_1,a'_2,\cdots,a'_k\}$,如此$f_B(x)$就表示对象$x\in U$在属性子集$B$上的值,实际上对应于一个$K$维向量。

L:语言集,它可以用DL-语言来描述,语言的语意可以通过$Tarki$[2]的方法来定义。

N:信息粒的数量特征集,例如概率测度、信息熵、包含强度、隶属度、覆盖度、支持度等等,它们可以使我们从定量化的角度出发分析和研究知识表达系统。

*)国家自然科学基金项目(60175016,60475019)。李道国 副教授,博士生,主要研究方向:模式识别与智能系统、粗糙集理论、粒度计算。

Γ:论域的结构,例如拓扑结构、半序或拟半序、代数结构和商结构[3]等,这些结构对于问题求解、粒逻辑与推理、时空规划、智能搜索技术和信息压缩等非常重要。

总之,基于知识表达系统[1]的知识发现(概念及其关系的发现)实质上是按照属性特征将对象进行分类,对每一个类(信息粒)冠以不同的名称,同时依据上、下近似算子便可以得到不同的确定性概念和近似概念。而这些概念表达的知识利用了满足某些条件的属性来表达,因而是易于理解的。

定义 2 信息粒划分(划分[2])

设U是给定的一个论域,若通过论域上的一个等价关系R获得一簇信息粒$G_i\subseteq U, i=1,2,\cdots,n$。满足:(1)$G_i\neq\phi$;(2)$G_i\cap G_j=\phi, i=j$,即两两互不相交;(3)$\bigcup_{i=1}^{n}G_i=U$,则称$\pi(U/R)=\{[x]_R:\forall x\in U\}=\{G_1,G_2,\cdots,G_n\}=\{G_i\}_{i=1}^{n}$为论域$U$的一个粒划分(划分),也可标记为$\pi(R)$,$\pi(ind(R))$,$U/R$,$U/ind(R)$。当论域与等价关系明确时,可简记为$\pi$,其中$[x]_R=\{y|xRy,\forall x,y\in U\}$。由于有限个等价关系的交仍然是等价关系,因此可以将信息粒划分的定义扩展如下:

设给定一个论域U和U上的一族等价关系R,若$B\subseteq R$,且$B\neq\phi$,则B中所有等价关系的交集:

$$ind(B)=R_B=\bigcap_{\forall\alpha\in B}\alpha=\cap B \tag{2}$$

仍然是U上的一个等价关系[3],称之为B上的不可区分(indiscernibility)关系,简记为$ind(B)$或B或R_B。因此,称$\pi(B)=\pi(ind(B))=\pi(U/B)=\pi(U/ind(B))$为不可区分关系$B$导出的粒划分,其中

$$[x]_B=\{y|\bigwedge_{\forall R\in B}xRy,\forall x,y\in U\} \tag{3}$$

$\zeta(\pi)=\{\pi|\pi=U/B,\forall B\subseteq\mathscr{R}\vee B=\phi\vee B=\text{“=”}\}$表示论域$U$上的粒划分全体,其中$\mathscr{R}$是论域$U$上的全体等价关系;$B=\phi$表示空关系$R_\phi$,则$U/\phi=U/R_\phi=[U]$,即论域$U$的最粗的粒划分;“=”表示对象的恒等关系$R_=$,则$U/R_==U$,即$\forall x,y\in U, xR_=y\Leftrightarrow x=y$表明根据现有的信息可以区分论域中的任何两个元素,它被认为是最细的粒划分。将这两种极端情形考虑进来,有益于对粒划分完备性的理解。

定义 3 粒划分的粗细

设$\pi_1(R_1)=\{G_{11},G_{12},\cdots,G_{1m}\}$,$\pi_2(R_2)=\{G_{21},G_{22},\cdots,G_{2n}\}$是论域$U$上的两个粒划分。如果对$\forall G_{1i}\in\pi_1$,总$\exists G_{2j}\in\pi_2$,使得$G_{1i}\subseteq G_{2j}$,则称$\pi_1$是比$\pi_2$更细的粒划分,简记为$\pi_1\leqslant\pi_2$。注意,粒划分之间存在不能比较粗细的情形。

定义 4 集合 U 上的等价关系的粗细

设给定集合 U 及 U 上的一切等价关系$\mathscr{R}$,对$\forall R_1,R_2\in\mathscr{R},\forall x,y\in U$,若满足$xR_1y\Rightarrow xR_2y$,则称$R_1$比$R_2$细,简记为$R_1\leqslant R_2$;若满足$xR_1y\Leftrightarrow xR_2y$,则称$R_1$与$R_2$等价(同样粗细);若上述两式均不成立,则称$R_1$与$R_2$不能比较粗细。

定义 5[3] 包含强度 设$(U,\leqslant)$是一个偏序集,如果对$\forall x,y\in U$,都有一个实数$\partial(y/x)$与之对应,且满足:(1)$0\leqslant\partial(y/x)\leqslant1$,(2)$x\leqslant y\Rightarrow\partial(y/x)=1$,(3)$x\leqslant y\leqslant z\Rightarrow\partial(x/z)\leqslant\partial(x/y)$,则称$\partial$为$U$上的包含强度。包含强度融合了许多的粒计算模型,为研究各种知识表达系统提供了更通用的度量。

例 1 设U是一个有限论域,$P(U)$是U的幂集,表示U的全体子集,$(P(U),\subseteq)$构成一个偏序集[2],式(2)中的P表示U上的概率测度,则对$\forall X,Y\in P(U)$,如果定义:

$$\partial(Y/X)=|X\cap Y|/|X| \tag{4}$$

$$\partial(Y/X)=P(X\cap Y)/P(X) \tag{5}$$

求证:(4)和(5)均为$P(U)$上的包含强度。

证明:(1)对$\forall X,Y\subseteq P(U)$,

$\because 0\leqslant|X\cap Y|\leqslant|X|$,

$\therefore 0\leqslant\partial(Y/X)=|X\cap Y|/|X|\leqslant1$,

(2)$X\subseteq Y,\partial(Y/X)=|X\cap Y|/|X|=|X|/|X|=1$,

(3)$X\subseteq Y\subseteq Z$,则$|X|\leqslant|Y|\leqslant|Z|$,$\partial(X/Z)=|X\cap Z|/|Z|=|X|/|Z|\leqslant|X|/|Y|=|X\cap Y|/|Y|=\partial(X/Y)$。

根据定义 5 可知(4)是$P(U)$上的包含强度。该包含强度可用来表示规则强度或支持度。

同理可证(5)也是$P(U)$上的包含强度,它可应用于变精度的粗糙集模型。

定义 6[3] 条件属性 $\alpha(\forall\alpha\in C)$相对于决策 D 的重要度

设(U,A_t,V,f,L,N,Γ)是一个广义的决策表,$\pi(ind(D))=\pi(R_D)=\pi(U/D)$表示决策属性$D$形成的论域$U$的粒划分。对$\forall\alpha\in C$,定义属性$\alpha$相对于决策$D$的重要度为:

$$sig(\alpha)=1-\partial(\pi(D)/\pi(C-\{\alpha\})) \tag{6}$$

其中,∂是$\zeta_U(\pi)$上的一个包含强度。重要度在分析属性相依性和构造决策表的知识约简的启发式算法中起着重要的作用。

定义 7[2] Q 的 P 正域 设给定论域U和论域上的等价关系P和Q,称$pos_P(Q)=\bigcup_{X\in U/Q}\underline{P}(X)$为$Q$的$P$正域,它表明了论域$U$中所有根据分类$U/P$的信息可以准确地划分到关系$Q$的等价类中去的对象集合,其中

$$\underline{P}(X)=\{Y_i|(Y_i\in U/P)\wedge(Y_i\subseteq X)\} \tag{7}$$

定义 8 决策表的核属性和核[5] 设给定一个广义的决策表(U,A_t,V,f,L,N,Γ),对$\forall\alpha\in C$,若单属性α满足以下条件之一:

(1)$pos_C(D)\neq pos_{C-\{\alpha\}}(D)$,

(2)在基于对象的差别矩阵$M=(r_{ij})_{n\times n}$中,$\exists r_{ij}=\{\alpha\}$单属性集,其中$n=|U|$,

(3)在基于粒划分$\pi(U/C)=\{G_i\}_{i=1}^{s}$的差别矩阵$M_G=(r_{ij}^G)_{s\times s}$中,$\exists r_{ij}^G=\{\alpha\}$。其中,$s=|U/C|$,

$$r_{ij}^G=\alpha(G_i,G_j)=\{\alpha\in C|(f_D(G_i)\neq f_D(G_i))\wedge(f_\alpha(G_i)\neq f_\alpha(G_j))\}\ \text{or}\ \{\phi|(f_\alpha(G_i)=f_\alpha(G_j),\forall\alpha\in C)\wedge(f_D(G_i)\neq f_D(G_j))\}\ \text{or}\ \{0|(f_D(G_i)=f_D(G_j),G_i\neq G_j)\vee(i=j)\} \tag{8}$$

$\forall G_i,G_j\in U/C,i,j=1,2,\cdots,s$

(4)在相容决策表中,$ind(C-\{\alpha\})\leqslant ind(D)$不成立,即$\pi(C-\{\alpha\})\leqslant\pi(D)$不成立,或$R_{C-\{\alpha\}}\leqslant R_D$不成立,则称$\alpha$为决策表的一个核属性。决策表的全体核属性称为它的相对核,简记为$core_C(D)$。核属性是绝对必要的属性。去掉任何一个核属性,都会改变决策表的分类能力。

定义 9[3] 相对约简 设给定一个决策表$(U,C\cup D,V,f,L,N,\Gamma)$,若$\exists P\subseteq C$,满足:(1)$P$是独立的,即对$\forall\alpha\in P$,$pos_{P-\{\alpha\}}(D)\neq pos_P(D)$,

(2)$pos_P(D)=pos_C(D)$,则称P是一个相对D约简,简记为$P\in red_C(D)$。

定义 10 协调的决策表 设给定一个决策表$(U,C\cup D,V,f,L,N,\Gamma)$,若满足$\pi(C)\leqslant\pi(D)$,即$pos_C(D)=U$,则称该决策表是协调的(相容的或一致的)决策表,否则称为不协调的决策表。基于协调的决策表获取的知识都是确定的知识,表明不含有不一致的(冲突的)信息(实例或样本),而不协调的决策表却相反。

3 基本定理

定理 1[3] 设$\zeta_U(\pi)$表示论域U上的全体粒划分,$(\zeta(\pi),$

$\leqslant$)为一个偏序集，给定$(P(U),\subseteq)$上的一个包含强度∂，对$\forall \pi_1,\pi_2\in\zeta_U(\pi)$如果定义：

$$\partial(\pi_2/\pi_1)=\bigwedge_{i=1}^{k}\bigvee_{j=1}^{l}\partial(Y_j/X_i) \tag{9}$$

则∂为偏序集$(\zeta(\pi),\leqslant)$上的包含强度，其中$\pi_1=\{X_i\}_{i=1}^{K}$，$\pi_2=\{Y_j\}_{j=1}^{L}$均为论域U的一个粒划分。

证明：(1)$\because \forall i,j,i=1,2,\cdots,k,j=1,2,\cdots,l,0\leqslant D(Y_j/X_i)\leqslant 1,\therefore 0\leqslant\partial(\pi_2/\pi_1)\leqslant 1$。

(2)$\pi_1\leqslant\pi_2$，对$\forall X_i\in\pi_i,\exists Y_j\in\pi_2,X_i\subseteq Y_j,\bigvee_{j=1}^{1}\partial(Y_j/X_i)=1\Leftrightarrow\partial(\pi_2/\pi_1)=1$。

(3)对$\forall\pi_1\leqslant\pi_2\leqslant\pi_3,\pi_3=\{Z_t\}_{t=1}^{r}$，也是论域的一个粒划分。由包含强度的定义可知，对$\forall X_i\in\pi_1,\exists Y'_j\in\pi_2$和$Z_p\in\pi_3$，使得$X_i\subseteq Y_j\subseteq Z_p,\therefore\partial(X_i/Z_p)\leqslant\partial(X_i/Y_j)\Rightarrow\bigwedge_{t=1}^{r}\bigvee_{i=1}^{k}\partial(X_i/Z_t)\leqslant\bigwedge_{j=1}^{l}\bigvee_{i=1}^{k}\partial(X_i/Y_j)\Rightarrow\partial(\pi_1/\pi_3)\leqslant\partial(\pi_1/\pi_2)$。由定义5可知$\partial$是一个包含强度。

定理2 设$(U,C\cup D,V,f,L,N,\Gamma)$是给定的一个协调的广义决策表，则$\alpha\in C$为核属性的充分必要条件是：

$$sig(\alpha)=1-\partial(\pi(ind(D))/\pi(ind(C-\{\alpha\})))>0 \tag{10}$$

证明：

$sig(\alpha)=1-\partial(\pi(ind(D))/\pi(ind)(C-\{\alpha\}))>0$

$\partial(\pi(ind(D))/\pi(ind(C-\{\alpha\})))<1$

$\pi(ind(C-\{\alpha\}))\leqslant\pi(ind(D))$不成立

$pos_{C-\{\alpha\}}(D)\neq pos_C(D)$，所以$\alpha$为决策表的一个核属性。

定理3 设给定一个协调的广义决策表$(U,C\cup D,V,f,L,N,\Gamma)$，则下面关于决策表的$C$的相对$D$约简的定义是等价的：

(1)若$\exists B\subseteq C$，满足：

(i)$pos_B(D)=pos_C(D)$，

(ii)$\forall\alpha\in B,pos_{B-\{\alpha\}}(D)\neq pos_B(D)$，则$B\in red_C(D)$。

(2)若$\exists B\subseteq C$，满足：

(ⅰ)$\pi(B)\leqslant\pi(D)$，即$R_B\leqslant R_D$，

(ⅱ)$\forall\alpha\in B,\pi(B-\{\alpha\})\leqslant\pi(D)$不成立，即$R_{B-\{\alpha\}}\leqslant R_D$不成立，则$B\in red_C(D)$。

(3)若$\exists B\subseteq C$满足：

(ⅰ)$\partial^+(\pi(D)/\pi(B))=1$，

(ⅱ)$\forall\alpha\in B,\partial^+(\pi(D)/\pi(B-\{\alpha\}))<1$，则$B\in red_C(D)$，其中，$\partial^+(\pi(D)/\pi(B))=\min_{x\in U}\partial([x]_D/[x]_B)$，或$\partial^+(\pi(D)/\pi(B))=\frac{1}{|U|}\sum_{X\in U/D}|\underline{B}(X)|$，$\partial$是幂集$P(U)$上的一个包含度，$\partial^+$是由$\partial$诱导的论域$U$的粒划分$\zeta_U(\pi)$上的包含度。

(4)若$\exists B\subseteq C$，满足：

(ⅰ)$B\cap r_{ij}^G=\phi,\forall r_{ij}^G(\neq 0)\in M_G,(i>j)or\ (i<j)$，即$\pi(B)\leqslant\pi(D)$，

(ⅱ)B是相对D独立，则$B\in red_C(D)$。

讨论知识约简的多种形式，有利于灵活处理知识约简问题和知识约简算法的设计。

4 嵌套决策粒度约简的关系分析

在面对众多实际问题时，常常需要将决策粒度进行细化或粗化，以便能够动态求解问题。因此，下面研究嵌套决策粒度的相对D核的关系和相对D约简的关系。

定义11 设$(U,C\cup D_i,V,f,L,N,\Gamma)$是给定的一广义的决策表，若满足：(1)$\pi(C)\leqslant\pi(D_i)$，(2)$\pi(D_1)\geqslant\pi(D_2)\geqslant\cdots\geqslant\pi(D_N)$，$i=1,2,\cdots,N$，则称这一族决策表为协调的嵌套决策粒度信息表。

定理4 设给定一族协调的嵌套决策粒度信息表$(U,C\cup D_i,V,f,L,N,\Gamma)$，$i=1,2,\cdots,N$，而$\partial$是$\zeta_U(\pi)$上的包含度，则有：

(1)$core_C(D_1)\subseteq core_C(D_2)\subseteq\cdots\subseteq core_C(D_N)$

(2)对$\forall B'\in red_C(D_{i+1})\Rightarrow\exists B\subseteq B'\wedge B\in red_C(D_i)$

即细粒度的约简可以缩减为粗粒度的约简。反之，对$\forall B\in red_C(D_i)\wedge\pi(B)\leqslant\pi(D_{i+1})\Rightarrow\exists B'\supseteq B\wedge B'\in red_C(D_{i+1})$，即粗决策粒度的约简在满足细粒度的相容性的条件下可以扩展为细粒度的一个约简。

证明：(1)设$\pi(D_i)\geqslant\pi(D_{i+1})$，下证，

$\forall\alpha\in core_C(D_i)\Rightarrow\alpha\in core_C(D_{i+1})$。

$\because\pi(C)\leqslant\pi(D_{i+1})\leqslant\pi(D_i)$，

对$\forall\alpha\in core_C(D_i),i=1,2,\cdots,N\Leftrightarrow sig(\alpha)=1-\partial(\pi(D_i)/\pi(C-\{\alpha\}))>0$

$\Leftrightarrow\partial(\pi(D_i)/\pi(C-\{\alpha\}))<1$

$\Leftrightarrow\pi(C-\{\alpha\})\leqslant\pi(D_i)$不成立，

$\Rightarrow\pi(C-\{\alpha\})\leqslant\pi(D_{i+1})$不成立，

$\Rightarrow\alpha\in core_C(D_{i+1})$，否则$\pi(C-\{\alpha\})\leqslant\pi(D_{i+1})$且又$\because\pi(D_{i+1})\leqslant\pi(D_i)\Rightarrow\pi(C-\{\alpha\})\leqslant\pi(D_{i+1})\leqslant\pi(D_i)\Rightarrow\alpha\notin core_C(D_1)$产生矛盾。所以$core_C(D_i)\subseteq core_C(D_{i+1})$。以此类推，可得命题(1)成立。

(2)设$\pi(D_i)\geqslant\pi(D_{i+1})$，下证

(i)对$\forall B'\in red_C(D_{i+1})$，总存在$B\subseteq B'$，使得$B\in red_C(D_i)$。

(ii)对$\forall B\in red_C(D_i)\wedge\pi(B)\leqslant\pi(D_{i+1})\Rightarrow(\exists B'\supseteq B)\wedge(B'\in red_C(D_{i+1}))$。

令$B'\in red_C(D_{i+1})\Leftrightarrow(\pi(B')\leqslant\pi(D_{i+1}))$且$(\forall\alpha\in B',\pi(B'-\{\alpha\})\leqslant\pi(D_{i+1})$不成立)，$\because\pi(D_i)\geqslant\pi(D_{i+1})\Rightarrow\pi(B')\leqslant\pi(D_{i+1})\leqslant\pi(D_i)$，若对$\forall\alpha\in B',\pi(B'-\{\alpha\})\leqslant\pi(D_i)$不成立，则$B'\in red_C(D_i)$。我们取$B=B'$，即$B$满足$B\subseteq B'$，且$B\in red_C(D_i)$；若$\exists\alpha\in B'$，使$\pi(B'-\{\alpha\}\leqslant\pi(D_i))$成立，则令$B''=B'-\{\alpha\}$，再判断对$\forall\beta\in B''$是否有$\pi(B''-\{\beta\}\leqslant\pi(D_i))$成立。若不成立，则$(B''=B'-\{\alpha\})\in red_C(D_i)$，取$B=B''$，即有$B=B''\subset B'$且$B\in red_C(D_i)$；若成立，则重复上述的过程。由于属性集合$C$是有限集，这样总能够找到$B\subseteq B'$，满足$(\pi(B)\leqslant\pi(D_i))$且$(\pi(B-\{b\})\leqslant\pi(D_i)$不成立，$\forall b\in B)$，即$B\in red_C(D_i)$，可类推到嵌套的决策粒度上，命题(i)得证。同理可证(ii)。

上述定理证明了基于相容决策表的嵌套决策粒度的相对D核的关系和D约简的关系，这为我们动态分析决策表和求解实际问题提供了理论上的支持。

5 试验结果

对给定的一个相容决策表(表1)的决策属性进行连续的细化，得到嵌套决策粒度。然后，利用一个自主开发的知识约简的试验平台 Mixed Rough Set Knowledge Reduction Systems (MRSKRS 系统)求出嵌套的决策粒度的核和约简，以检验结果是否符合定理4。其中，条件属性$a\sim i,D_i,i=1,2,3,4,5$为嵌套的决策粒度。表2显示了嵌套决策粒度的核及其关系，表3显示了嵌套决策粒度的约简及其关系。

表 1 相容决策表

U	a	b	c	d	e	f	g	h	i	D1	D2	D3	D4	D5
1	A	T	T	T	T	C	G	T	T	1	1	1	1	1
2	A	T	T	T	T	G	A	T	T	1	1	1	1	10
3	A	T	T	T	T	G	T	A	G	1	1	1	1	10
4	A	T	T	T	T	T	T	T	C	1	1	1	5	5
5	A	T	T	T	T	T	T	T	G	1	1	1	5	11
6	C	T	T	G	G	A	G	A	G	1	1	1	5	11
7	C	T	T	G	G	C	G	C	T	1	1	1	5	11
8	C	T	T	G	G	C	T	C	T	1	1	3	3	3
9	C	T	T	G	T	G	C	T	A	1	1	3	3	12
10	C	T	T	T	A	A	G	A	A	1	1	3	3	12
11	C	T	T	T	A	T	T	C	T	1	1	3	6	6
12	C	T	T	T	T	T	A	T	A	1	1	3	6	13
13	C	T	T	T	T	T	C	T	T	1	1	3	6	13
14	C	T	T	T	T	T	T	T	T	1	1	3	6	13
15	G	C	A	G	G	A	A	T	G	1	2	2	2	2
16	G	C	A	G	G	A	G	A	C	1	2	2	2	14
17	G	C	A	G	G	G	A	G	A	1	2	2	2	14
18	G	C	A	G	G	G	C	A	C	1	2	2	2	14
19	G	C	A	G	G	G	G	G	A	1	2	2	7	7
20	G	C	A	G	G	G	T	C	C	1	2	2	7	15
21	G	C	A	G	T	T	G	G	T	1	2	2	7	15
22	G	C	A	T	A	G	A	A	A	1	2	2	7	15
23	G	C	A	T	C	A	A	G	C	1	2	4	4	4
24	T	T	T	T	G	G	G	C	C	1	2	4	4	4
25	T	T	T	T	T	A	T	T	C	1	2	4	4	16
26	T	T	T	T	T	C	C	C	T	1	2	4	4	16
27	T	T	T	T	T	T	C	T	T	1	2	4	9	9
28	T	T	T	T	T	T	G	A	G	1	2	4	9	9
29	T	T	T	T	T	T	T	C	A	1	2	4	9	17
30	T	T	T	T	T	T	T	T	T	1	2	4	9	17

表 2 相对 D 核

ϕ	$\{a\}$	$\{a,g\}$	$\{a,g\}$	$\{a,g,i\}$
$core_C(D_1)$	$core_C(D_2)$	$core_C(D_3)$	$core_C(D_4)$	$core_C(D_5)$

从表 2 和表 3 不难看出，完全符合定理 4 的结论。决策粒度越细，它的近似分类精度、分类质量和信息熵[6]就越大。但计算复杂度也相应变大，有时造成计算代价太大。所以，决策粒度不宜太细，决策粒度的细化程度应适度和符合实际需要。

表 3 相对 D 约简

$red_C(D_1)$	$red_C(D_2)$	$red_C(D_3)$	$red_C(D_4)$	$red_C(D_5)$
ϕ	$\{a\}$	$\{a,g,e\}$ $\{a,g,f,d\}$ $\{a,g,h,d\}$ $\{a,g,i\}$	$\{a,g,f,d\}$ $\{a,g,i,h,d\}$ $\{a,g,f,e\}$ $\{a,g,i,h,e\}$ $\{a,g,i,h,f\}$	$\{a,g,i,f,e\}$ $\{a,g,i,h,d\}$ $\{a,g,i,h,e\}$ $\{a,g,i,h,f\}$

结束语 本文分析研究了相容决策表的嵌套决策粒度的相对 D 核的关系和相对 D 约简的关系，得出并证明了粗决策粒度的核一定是细决策粒度核的子集，粗决策粒度的一个相对 D 约简在满足相容性的条件下一定可以扩张为细决策粒度的一个约简，反过来，细决策粒度的一个约简一定可以缩减为粗决策粒度的一个约简。研究结果对知识约简和动态求解问题有一定的实际意义。

5 参 考 文 献

1 王国胤. Rough 集理论与知识获取[M]. 西安：西安交通大学出版社，2001
2 Zhang Wenxiu, Wu Weizhi, Liang Jiye, et al. The theory and methodology of Rough Set [M]. Lin Peng, et al, eds. Beijing: Science Publishing Company, 2000
3 Zhang Wenxiu, Liang Yi, Wu Weizhi. Information System and KDD [M]. Yang Bo Eds. Beijing: Science Publishing Company, 2000
4 张铃，张钹. 模糊商空间理论（模糊粒度计算方法）[J]. 软件学报，2003，14(4)：770～776
5 Skowron A. The rough sets theory and evidence theory [J]. Fundamental Information XIII Intelligence, 1995, 11: 371～388
6 苗夺谦，王珏. 粗糙集理论中概念与运算的信息表示[J]. 软件学报，1999，10(2)：113～116

（上接第 165 页）

根据以上拟合训练和外推预测的结果分析，DCSVM 及传统 SVM 方法均具有较好的推广能力，误差符合预测精度的要求，用于航空航线旅客运输量的预测是有效的。而无论从 MAPE 值还是从 EC、RMSE、MXE 进行比较，均表明 DCSVM 的结果比传统 SVM 要好。

结束语 理论上 SVM 能以任意精度逼近函数。本文建立了一种分而治之的 SVM 学习方法（称为 DCSVM 网络模型），并将该 DCSVM 方法应用于一般的函数逼近与实际中的航空旅客量预测问题，采用 ε 一次及二次不敏感损失函数，分别选用 Linear、RBF、ANOVA 和 Poly 等 4 种核函数建立了 DCSVM 网络模型，其实际结果具有相当理想的精度，比传统 SVM 方法预测的精度更高，可以满足预测要求。如果采用并行计算方法，计算时间将远远小于传统 SVM 模型。

到目前为止，对 SVM 模型的核函数及其参数以及损失函数的选择尚没有确定的方法和结论。对取各种核函数预测后的结果进行二次 SVM 预测，是一个可取的办法。在本文的变量可分离支持向量机 DCSVM 网络模型的基础上，值得考虑的方面包括：一般性的任意变量分离支持向量机、直接影响预测结果的 DCSVM 模型中加权平均系数 a_i 的选择；DCSVM 模型的逼近能力与复杂性分析的理论证明等。限于篇幅，作者将另文论述。

参 考 文 献

1 Adams W, Michael V. Short Term Forecasting of Passenger Demend and Some Application in Quantas . AGIFORS Symposium Proc, 27, Sydney, Australia, 1987
2 Vapnik V N. The Nature of Statistical Learning Theory [M]. NY: Springer2Verlag, 1995
3 边肇祺，张学工. 模式识别[M]. 第 2 版. 北京：清华大学出版社，2000
4 邓乃扬，田英杰. 数据挖掘中的新方法——支持向量机. 北京：科学出版社，2004
5 Platt J C . Fast Training of Support Vector Machines using Sequential Minimal Optimization [R]. Microsoft Research , 2000
6 Cortes C, Vapnik V. Support Vector Networks. Machine Learning, 1995, 20: 273～297
7 Huang Rongbo. A Divide-and-Conquer Hybrid System Based Radial Basis Function Networks: [Doctoral Thesis]. Zhongshan University, 2004
8 Gunn S R. Support Vector Machines for Classification and Regression [R]. University of outhampton , 1998
9 PROS5. 2 Training Course PROS Company 2001
10 中国南方航空股份有限公司. 中国南方航空股份有限公司 2003 年统计年鉴. 2004