第 23 卷第 6 期
2006 年 12 月

控 制 理 论 与 应 用
Control Theory & Applications

Vol. 23 No. 6
Dec. 2006

文章编号: 1000−8152(2006)06−0867−06

基于粗糙集的案例属性约简技术

常春光[1,2], 汪定伟[2], 胡琨元[2], 陶　志[2]

(1. 沈阳建筑大学 管理学院, 辽宁 沈阳 110168; 2. 东北大学 信息科学与工程学院, 辽宁 沈阳 110004)

摘要: 为了提高案例推理(case-based reasoning, CBR)系统的案例匹配效率, 引入粗糙集理论对案例属性约简技术加以研究. 提出准约简的概念, 依次证明了某一条件属性集成为准约简与准约简成为约简的充要条件. 进而, 以核为出发点, 提出了一种改进的基于分辨矩阵的属性最小约简算法. 为使之仍然适用于连续属性, 提出一种基于逼近精度敏感性的离散化算法. 最后, 将此约简技术应用于某钢铁企业的实际动态调度问题中, 计算试验表明, 该技术消除了冗余信息, 提高了案例匹配的效率.

关键词: 案例推理; 案例匹配; 粗糙集; 属性约简; 分辨矩阵; 动态调度

中图分类号: TP182　　**文献标识码:** A

Rough-set-based reduction technique for case attributes

CHANG Chun-guang[1,2], WANG Ding-wei[2], HU Kun-yuan[2], TAO Zhi[2]

(1. School of Management, Shenyang Jianzhu University, Shenyang Liaoning 110168, China;
2. School of Information Science & Engineering, Northeastern University , Shenyang Liaoning 110004, China)

Abstract: To improve the efficiency of case retrieving in CBR(case-based reasoning), the rough-set theory is introduced in this paper to study the reduction technique for case attributes. Firstly, the concept of quasi-reduction is presented. The necessary and sufficient conditions for some attribute-set to become quasi-reduction, and the quasi-reduction to become reduction are then proved. Secondly, starting from the core, a differentiating matrix-based improved algorithm for minimal attribute reduction is then proposed. To maintain its application to continuous attributes, the dispersing algorithm based on the sensitivity of approximation precision is also proposed. Finally, the technique is applied to a practical dynamic scheduling problem of an iron and steel works. The computation experiment shows that it eliminates redundant information, improving the efficiency of case retrieving.

Key words: case-based reasoning; case retrieving; rough set; attributes reduction; differentiating matrix; dynamic scheduling

1　引言(Introduction)

钢铁企业生产是介于离散与连续流程行业之间的一种混杂生产方式, 其考虑因素众多, 因素间存在复杂的相互关系, 其动态调度问题一般要难于后两者. 同时由于钢水温度的限制, 又具有极强的实时性要求. 为解决这一问题, 笔者已经开发了基于两级案例推理(case-based reasoning)技术的动态调度系统. 在该系统实施中的两个重要的环节即案例整理与匹配都与案例属性的约简有着密切的关系.

粗糙集(rougn set, RS)理论是由波兰学者Pawlak Z在1982 年提出的[1], 作为刻划不完整性和不确定性的数学工具, 能有效地分析和处理不精确、不一致、不完整等各种不完备信息, 并从中发现隐含的知识, 揭示潜在的规律[2]. 与模糊集和概率统计方法相比, RS仅利用数据本身提供的信息, 无须任何先验知识. RS能在保留关键信息的前提下对数据进行化简并求得知识的最小表达[3]. 国内外学者对属性简化展开了相关研究, 文献[4]与文献[5]研究了基于遗传算法的属性约简方法. 文献[6]提出基于动态聚类和减少不相容性的属性量化算法. 文献[7] 提 出 了 4 种基于熵的属性约简方法. 为提高案例匹配的效率, 本文采用了基于粗糙集理论的属性约简技术, 其效

收稿日期: 2004−01−07; 收修改稿日期: 2006−03−07.
基金项目: 国家自然科学基金资助项目(70431003, 70171056); 国家 863/CIMS 项目(2002AA412010, 2002AA414610).

果在实践中得以验证.

2 粗糙集理论的基本概念(Basic concepts of rough set theory)

定义 1[8] 四元组S=(U,A,V,f)是一个知识表达系统. 其中: U为非空有限对象集合, 称为论域; $A = C \cup D$, $C \cap D = \phi$, C称为条件属性集, D为决策属性集; $V = \cup V_a$, $a \in V_a$, V_a是属性a的值域; f是$U \times A \to V$的一个函数.

其中条件属性集C对应于案例的状态属性集, 决策属性集D相当于案例的解的属性集.

定义 2[9] 对$\forall P \subseteq A, \forall x \in U$, 称$[x]_p = \{y \in U | \forall a \in P, f(x,a) = f(y,a)\}$为$x$的$P$等价类.

定义 3[10] 分辨矩阵***DM***, 是一个$w \times w$矩阵, 每一个元素$DM_{ij} \subseteq A$, 其中, $DM_{ij} = \{dm_{ij1}, dm_{ij2}, \cdots, dm_{ijn}\}$, $i,j = 1,2,\cdots,w$, 其中dm_{ijk} 确定如下:

$$dm_{ijk} = \begin{cases} \phi, & a_{ik} = a_{jk}, \\ A_k, & a_{ik} \neq a_{jk}, \end{cases} k = 1,2\cdots,n. \quad (1)$$

定义 4[5] 在$S = (U,A,V,f)$中, 对于每个子集$X \subseteq U$和一个等效关系$R \subseteq A$, 称子集$\underline{R}X = \{x \in U | [x]_R \subseteq X\}$为$X$的$R$下逼近集, 也称作$X$的$R$正域, 记作$\mathrm{Pos}_R(X)$.

$\mathrm{Pos}_R(X)$或$\underline{R}X$是由那些根据知识R判定属于X的U中元素组成的集合. 同理, 上逼近集$\overline{R}X$是那些根据知识R判定可能属于X的R中元素组成的集合.

定义 5[3] 逼近精度, 是在等效关系R下逼近集合X的精度, 定义如下, 其中$|\cdot|$表示集合的基数, 对有限集来说表示集合中包含元素个数. 显然, $0 \leqslant \sigma_R(X) \leqslant 1$,

$$\sigma_R(X) = |\underline{R}X|/|\overline{R}X|. \quad (2)$$

定义 6 约简. 设$S = (U, A, V, f)$中, $P \subseteq C$, $Q \subseteq D, K \subseteq P$, 如果$\mathrm{Pos}_K(Q) = \mathrm{Pos}_P(Q)$, 且对于$\forall R \subseteq K$, 有$\mathrm{Pos}_K(Q) \neq \mathrm{Pos}_{(K-\{R\})}(Q)$, 则称$K$是$P$的一个$Q$约简, 记作$\mathrm{Red}_P(Q)$.

约简实质上为不含多余属性并保证分类正确的最小条件属性集. 一个决策表可能同时存在几个约简.

定义 7 准约简. 设$S = (U,A,V,f)$中, $P \subseteq C$, $Q \subseteq D, K \subseteq P$, 如果$\mathrm{Pos}_K(Q) = \mathrm{Pos}_P(Q)$, 称$K$是$P$的一个$Q$准约简.

准约简与约简相比, 可能是约去部分多余属性并保证分类正确的条件属性集, 也可能就是约简. 约简的交集定义为决策表的核. 核中的属性是影响分类的重要属性,是任何一个约简都必须包含的属性.

3 案例属性约简原理(Principle of case attributes reduction)

3.1 离散案例属性的约简(Reduction of discrete case attributes)

定理 1[10] 任一条件属性$C_k \in C$关于d不可约简, $d \subseteq D$, 充要条件为: 在分辨矩阵DM中, 至少存在一个$DM_{ij}(C,d) \in DM(C,d)$, 满足$DM_{ij}(C,d) = \{C_k, d\}$.

定理1实际上提供了一种求解核的方法, 所有满足定理1的条件属性所组成的集合就是核.

定理 2 设$B_1 \subseteq B = C - \mathrm{Core}_d C$, 则$\mathrm{Core}_d C + B_1$是$C$的一个$d$准约简的充要条件为: 对任何含有$d$的$DM_{ij}(C,d)$, 均满足

$$\{DM_{ij}(C,d)\backslash d\} \underline{\subseteq} (B - B_1).$$

证 充分性. 若对任何含有d的$DM_{ij}(C,d)$, 均满足$\{DM_{ij}(C,d)\backslash d\} \underline{\subseteq} (B–B_1)$. 则说明将$B–B_1$中属性同时约简后, 仍能保证在任何含有$d$的$DM_{ij}(C,d)$中有$DM_{ij}(C,d)\backslash d - (B-B_1) \neq \phi$. 这说明原有的分辨能力不变, 即$\mathrm{Pos}_{\mathrm{Core}_d C + B_1}(d) = \mathrm{Pos}_C(d)$, 同时考虑到$C \subseteq C, d \subseteq D$, $\mathrm{Core}_d C + B_1 \subseteq C$, 根据定义7, 可证明$\mathrm{Core}_d C + B_1$是$C$一个$d$准约简. 充分性得证.

必要性. 反证法. 假设$\mathrm{Core}_d C + B_1$是C的一个d准约简, 且至少存在一个$DM^*_{ij}(C,d)\backslash d \subseteq B - B_1$, 则当$B - B_1$中属性同时约简后, 必有$DM^*_{ij}(C,d)\backslash d - (B - B_1)) = \phi$. 说明$B - B_1$中属性同时约简后分辨能力发生改变, 即$\mathrm{Pos}_{\mathrm{Core}_d C + B_1}(d) \neq \mathrm{Pos}_C(d)$, 与假设矛盾. 必要性得证.

从定理2可以看出, 核$\mathrm{Core}_d C$是保证原有分辨能力不变的必要属性集, 而B_1中属性是得到某一准约简的补充属性集, 因此属性集B_1可以在含有d不含核$\mathrm{Core}_d C$的$DM_{ij}(C,d)$中得到. 定理2提供了求解约简的一个思路, 即从核出发, 逐渐加入非核元素, 检验未被加入的非核元素$B - B_1$是否可以同时约简掉. 直到可以同时约简掉时停止. 但由于定理2每次检验的标准只是未被加入的非核属性是否可以同时约简掉, 因此直接求得的结果并不一定是约简, 很有可能最终结果中包括某些冗余属性, 而得到一个准约简. 为了在准约简的基础上求出约简, 本文证明了定理3.

定理 3 准约简$\mathrm{Core}_d C + B_1$, 是C的d约简充分必要条件为: 对$\forall b \in B_1$都不满足$\{DM_{ij}(C,d)\backslash d\} \underline{\subseteq} (B - B_1 + b)$.

证　必要性. 因准约简Core_dC+B_1是C的d约简, 故$\mathrm{Pos}_{\mathrm{Core}_dC+B_1}(d)=\mathrm{Pos}_P(d)$, 且对$\forall b\subseteq \mathrm{Core}_dC+B_1$, 有$\mathrm{Pos}_{\mathrm{Core}_dC+B_1}(d)\neq \mathrm{Pos}_{(\mathrm{Core}_dC+B_1-b)}(d)$, 说明从$B_1$中去掉任一属性$b$, 都会改变原有的分辨能力. 即在分辨矩阵中, 对$\forall b\in B_1$都不满足$\{DM_{ij}(C,d)\backslash d\}\underline{\subset}(B-B_1+b)$. 必要性得证.

充分性. 设知识表达系统$S=(U,A,V,f)$中,$P\subseteq C, d\subseteq D, \mathrm{Core}_dC+B_1\subseteq P, \mathrm{Core}_dC+B_1$ 是C的d准约简可得Core_dC+B_1不会改变原有的分辨能力, 即$\mathrm{Pos}_{\mathrm{Core}_dC+B_1}(d)=\mathrm{Pos}_P(d)$. 因在分辨矩阵中, 对$\forall b\in B_1$都不满足$\{DM_{ij}(C,d)\backslash d\}\underline{\subset}(B-B_1+b)$, 所以, 对于$\forall b\subseteq \mathrm{Core}_dC+B_1$, 有$\mathrm{Pos}_{\mathrm{Core}_dC+B_1}(d)\neq \mathrm{Pos}_{(\mathrm{Core}_dC+B_1-b)}(d)$, 根据定义6, Core_dC+B_1是P的一个d约简. 充分性得证.

定理3是求得一个约简的重要依据, 但约简不是唯一的, 有时笔者更想求得所有约简中属性最少的一种约简即最小约简. 为此设计最小约简算法如下:

步骤 1　根据定义3求出决策表的分辨矩阵DM.

步骤 2　根据定理1求出所有不可约简的条件属性C_k, k=1, 2, $\cdots$, m. 确定核Core_dC=$\{C_1, C_2, \cdots, C_m\}$. 令$B_1=\phi$.

步骤 3　在含有d而不含Core_dC+B_1的$DM_{ij}(C,d)$中取出一个属性b, $B_1=B_1\cup b$, 检验是否对任何含d的$DM_{ij}(C,d)$均满足$\{DM_{ij}(C,d)\backslash d\}\underline{\subset}(B-B_1)$. 如果是则得到一个准约简$\mathrm{Core}_dC+B_1$. 到步骤5.

步骤 4　取另一个不同的属性b, 重复步骤3, 直到含有d而不含Core_dC+B_1的$DM_{ij}(C,d)$中所有元素被取完.

步骤 5　对此时的B_1进行约简, 设共有z种组合顺序. 设约去属性个数为y, y=0. j=0.

步骤 6　选择其中一个顺序j. 依次去掉B_1中属性, 验证是否仍然满足定理2, 若满足则该属性可约简掉, 直到B_1中的任一元素都验证完为止. 此时得到一个约简Core_dC+B_{1j}和相应的约去属性个数y_1. 如果$y_1>y$. 则$\mathrm{Red}^{\min}=\mathrm{Core}_dC+B_{1j}$.

步骤 7　$j=j+1$, 重复步骤6, 直到$j>z$. 根据推论2, 此时的$\mathrm{Red}^{\min}$就是最小约简.

3.2　连续案例属性的约简(Reduction of continuous case attributes)

粗糙集理论适于处理离散属性问题, 却不能直接处理连续属性, 这限制了其应用范围. 在实际的案例属性中, 既包含离散的又包含连续的, 因此, 用粗糙集方法进行属性约简时, 首先必须将其离散化. 连续属性离散化有多种划分方法: 1) 局部和全局的离散化方法; 2) 静态和动态的离散化方法[11,12]; 3) 有监督和无监督的离散化方法[11]. 其中, 典型的无监督离散化方法包括又可分为等距离算法、等频率算法和基于类信息熵的离散化算法[13]. 典型的有监督算法又可分为自然算法、布尔推理算法[12,13]. 这里, 笔者提出了一种以自然算法为初始状态, 基于逼近精度敏感性的有监督离散化算法.

连续属性离散化的精细度要适中, 如果划分过细虽然可以提高决策表整体分类能力, 但往往会带来很多冗余信息, 降低了约简效率, 更会直接影响案例匹配的效率; 如果划分较粗, 则可能增加决策表中的不相容信息. 因此笔者对连续属性离散化是保持较高逼近精度的前提下(即在逼近精度敏感系数阈值范围内), 寻找使得约简效率最高的划分的过程. 为此, 引入逼近精度敏感系数的定义.

定义 8　逼近精度敏感系数,设连续属性a_i离散化的分割点为$a_{ij}(i=1,2,\cdots,k, j=1,2,\cdots,F_i$, 其中$k$为连续属性的个数, F_i为属性a_i的分割点的个数). 任一属性a_i的分割点a_{ij}与$a_{i,j+1}$的合并后逼近精度的变动量与合并前的逼近精度$\sigma_R(X)$的比值, 如式(3)所示$V_{\sigma j}^{(i)}$称为逼近精度关于属性a_i在分割点a_{ij}的敏感系数.

$$V_{\sigma j}^{(i)}=|\sigma_{R(a_{ij}\cup a_{i,j+1})}(X)-\sigma_R(X)|/\sigma_R(X). \quad (3)$$

有了上述定义, 该算法的具体步骤如下:

步骤 1　将条件属性集合C中所有待离散化的连续属性分别设为a_i, $i=1,2,\cdots,k$, 令$i=1$.

步骤 2　将对象U_p, p=1, 2, $\cdots$, n, 依连续属性a_i的值V_{ip}从小到大排序, 得到新的对象序列$\{U^{(q)}\}$, $q=1,2,\cdots,n$, 令$a_{i1}=\min\{V_{iq}\}$, $j=1$. $F_i=1$.

步骤 3　$q=q+1$, 判断对象$U^{(q)}$的决策值是否满足$d_q\neq d_{q-1}$, 如果是则产生一新分割点, $j=j$+1, $a_{ij}=V_{iq}$. $F_i=F_i+1$.

步骤 4　循环步骤3, 直到$q=n$.

步骤 5　$i=i+1$, 循环步骤1到4, 直到$i=k$. 设得到的各个连续属性a_i离散化的初始分割点为a_{ij}, $i=1,2,\cdots,k, j=1,2,\cdots,F_i$. 重置$i$=1.

步骤 6　计算连续属性a_i在初始离散状态的逼近精度$\sigma_R^{(i)}(X)_0$.

步骤 7　依次去掉分割点a_{ij}, 据式(3)计算$V_{\sigma j}^{(i)}$, $j=1,2,\cdots,F_i$. 如果存在$\exists V_{\sigma j}^{(i)}\leqslant\gamma$($\gamma$为事先规定的逼近精度的敏感系数阈值, 这里取0.25), 则取$\min\{V_{\sigma j}^{(i)}\}$, 对应的分割点$a_{ij}^{\min}=a_{ir}$, 将其删去, $F_i=F_i-1$, 重新修正分割点, 即如果$a_{ij}>a_{ir}$, 则$a_{ij}=a_{ij+1}$.

步骤 8　循环步骤7, 直到不再$\exists V_{\sigma j}^{(i)}\leqslant\gamma$.

步骤 9　$i=i+1$, 循环步骤6到8, 直到$i=k$时结束.

4 案例属性约简应用实例(Application example of case attributes reduction)

限于篇幅, 仅列出部分案例的数据, 见表1所示. 其中一部分属性(炉次数CHN、周期SPAN、精炼1值OB1V、精炼2值OB2V、下炉次值CCPV、IC到时间HIC1V、另1CC值OCC1SV、另2CC值OCC2SV)为连续属性, 其余的属性都是离散属性. 调整方案对照表见表2.

表 1 部分案例原始数据

Table 1 Partial case original data

属性	注释	案例属性值																			
C2NO	案例号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
CHN	炉次数	7	6	7	6	8	6	8	9	7	7	8	7	8	7	8	6	7	9	7	7
SPAN	周期/ min	120	130	110	140	125	135	150	135	142	135	145	150	130	140	145	155	150	135	145	135
CSF	改钢种	Ck1	Ck1	Gr4	Ju5	Ju5	Ap1	Ju5	Gg8	Ck1	Ck1	Ck1	Gg8	Gr4	Ju5	Ck1	Ap1	Ju5	Gg8	Ck1	Ck1
OB	升温否	F	F	F	F	T	T	T	T	T	T	F	F	F	F	T	T	T	T	T	T
OBV	精炼选择	0	0	0	0	rh2	rh2	Cas1	Rh1	Cas1	Lf	0	0	0	0	Cas2	lf	Rh1	Rh1	Cas1	Cas2
OB1	精炼1	Cas1	Cas2	Rh1	lf	Rh1	lf	Cas1	Rh2	Cas2	Lf	Cas2	Cas2	Cas2	Rh2	Rh1	Rh2	Rh1	Rh2	Cas2	lf
OB1S	精1状态	on	on	off	off	on	on	off	off	on	on	on	on	off	off	on	on	off	off	on	on
OB1V	精炼1值	16	8	26	2	11	21	14	19	19	6	16	9	26	12	11	32	26	19	29	16
OB2	精炼2	Cas2	Cas1	kip	Rh1	Rh2	Rh2	lf	Rh1	Cas1	Cas2	Cas1	Cas1	kip	lf	Cas2	lf	lf	Rh1	Cas1	Cas2
OB2S	精2状态	on	off	off	on	off	on	off	off	off	on	on	off	on	off	on	on	on	off	off	on
OB2V	精炼2值	8	12	24	19	20	6	4	28	18	23	8	19	5	21	2	17	14	28	26	8
CCP	本CC值	1	1	1	2	2	2	3	3	2	1	1	1	2	3	2	2	3	3	2	1
CCPS	下炉状态	on	on	on	on	on	on	off	off	on	on	on	on	on	on	on	on	off	off	on	on
CCPV	下炉次值	10	3	19	35	46	5	9	37	5	14	10	3	19	35	46	50	9	27	5	20
LTS	调后连浇	T	T	F	F	F	T	T	T	T	F	T	T	F	F	F	F	T	T	T	F
HIC	有IC	F	F	F	F	T	T	T	T	T	T	F	F	F	F	T	T	T	T	T	T
HIC1V	IC到时	100	100	100	100	5	19	6	20	19	8	100	100	100	100	5	19	6	20	19	9
HIC1V1	满足	F	F	F	F	T	F	T	F	F	T	F	F	F	F	T	F	T	F	F	T
OCC	改CC值	2	3	2	3	1	1	2	1	1	3	3	2	1	2	1	1	2	2	3	3
OCC1	另1CC	2	2	2	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1	1	2
OCC1S	1CC状态	on	on	off	on	on	off	on	off	off	on	on	off	off	on	on	off	on	on	off	on
OCC1SV	1CC值	18	9	30	2	19	40	35	29	29	18	18	8	24	12	14	40	35	19	9	18
OCC1SP	1CC钢种	Ck1	Ap1	Gr4	Ju5	Ju5	Ap1	Gr4	Gg8	Gr4	Ck1	Ck1	Ap1	Gg4	Ju5	Ck1	Ap1	Gr4	Gg8	Gr4	Ck1
OCC1A	1CC需调	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
OCC1L	1CC最后	T	F	F	F	T	F	T	F	F	T	T	F	F	F	T	F	T	F	F	T
OCC2	另2CC	3	3	3	3	3	3	2	2	3	3	3	2	3	2	3	3	2	2	3	3
OCC2S	2CC状态	on	off	on	off	on	off	off	on	off	on	on	off	on	off	on	off	off	off	off	on
OCC2SV	2CC值	34	12	8	30	26	19	25	22	10	6	9	9	8	30	26	19	5	22	16	6
OCC2SP	2CC钢种	Ck1	Ap1	Gg8	Ju5	Ap1	Gr4	Gr4	Gg8	Gr4	Ck1	Ck1	Ap1	Gr4	Ju5	Ap1	Ap1	Gr4	Gg8	Gr4	Ck1
OCC2A	2CC需调	F	F	T	F	F	F	F	T	F	F	F	F	T	F	F	F	F	F	F	F
OCC2L	2CC最后	T	F	T	F	F	F	F	F	F	T	T	F	T	F	F	F	F	F	F	T
JTTZ	调整方案	1	2	7	22	15	19	18	19	5	14	2	2	7	8	15	11	18	20	6	16

将表1中的连续属性, 应用上文讨论的基于逼近精度敏感系数的有监督离散化算法, 进行离散处理, 其结果见表3. 再用上文设计的案例属性约简技术, 实现属性的约简. 最终求得的原有决策表的核是去掉属性包括: 炉次数CHN、周期SPAN、精炼1设备OB1、精炼2设备OB2、调后连浇LTS、本CC值CCP、改CC值OCC、另1CC设备OCC1、另CC1需调整OCC1A、另2CC设备OCC2、另

CC2 需调整OCC2A. 得到两个最小约简,

$$\mathrm{Core}_dC + B_1 = \mathrm{Core}_dC + \{\text{精炼选择OBV}\},$$

$$\mathrm{Core}_dC + B_1 = \mathrm{Core}_dC + \{\text{升温否OB}\}.$$

表 2 调整方案对照表

Table 2 Adaptation scheme contrast table

JTTZ	调整方案
1	加速原CC后续charge; 改向到另CC1;
2	加速原CC后续charge; 改向到另CC2;
3	加速原CC后续charge; 升温到精炼1; 改向到另CC1;
4	加速原CC后续charge; 升温到精炼1; 改向到另CC2;
5	加速原CC后续charge; 升温到精炼2; 改向到另CC1;
6	加速原CC后续charge; 升温到精炼2;改向到另CC2;
7	中断原CC浇铸; 改向到另CC1;
8	中断原CC浇铸; 改向到另CC2;
9	中断原CC浇铸; 升温到精炼1; 改向到另CC1;
10	中断原CC浇铸; 升温到精炼1; 改向到另CC2;
11	中断原CC浇铸; 升温到精炼2; 改向到另CC1;
12	中断原CC浇铸; 升温到精炼2; 改向到另CC2;
13	IC钢水补充原CC; 升温到精炼1; 改向到另CC1;
14	IC钢水补充原CC; 升温到精炼1; 改向到另CC2;
15	IC钢水补充原CC; 升温到精炼2; 改向到另CC1;
16	IC钢水补充原CC; 升温到精炼2; 改向到另CC2;
17	升温到精炼1; 改向到另CC1;
18	升温到精炼1; 改向到另CC2;
19	升温到精炼2; 改向到另CC1;
20	升温到精炼2; 改向到另CC2;
21	改向到另CC1;
22	改向到另CC2;

这样, 在最小约简的基础上, CBR系统的案例匹配所依据的案例属性大大减少, 使得在案例匹配之前能从众多属性中找出关键性决定性属性. 因此案例匹配(无论是采用最邻近法还是神经网络方法)效率将会被提高. 为了便于结果的对比分析, 以神经网路匹配技术为例, 案例匹配的时间复杂度可以简化为$\mathrm{O}(Nn)$[14], 其中N为案例个数, n为属性个数. 可见, 案例属性的多少直接影响到案例匹配的速度.

为了验证本文提出的基于逼近精度敏感系数的离散化算法的优越性, 以同样的数据应用于基于类信息熵的离散化算法. 由于后者没有逼近精度的阈值, 为了具有可比性, 运用后者方法时, 对每个连续属性离散到出现与前者方法得到相同离散区间个数时为止. 离散化后的结果见表4. 与表3的结果相比, 属性OB1V 和OB2V的离散化区间有较大差别, 其他属性的离散化区间基本接近. 同样应用上文设计的案例属性约简技术, 实现属性的约简. 最终求得的两个最小约简去掉了属性OB1V和OB2V. 这与钢铁生产实际是相违背的, 因为二者数值的大小是决定采取什么调整决策的重要依据. 可见, 本文提出的离散化算法对解决复杂的实际问题更具有优势.

表 3 连续属性离散化后的结果(基于逼近精度敏感系数的方法)

Table 3 Discretization result continuous attributes (approximation precision sensitivity based approach)

属性	注释	离散化后的区间
CHN	炉次数	[5,11]
SPAN	周期	[100,160]
OB1V	精炼1值	[2,9] [10,21] [22,33]
OB2V	精炼2值	[3,8] [9,21] [22,29]
CCPV	下炉次值	[3,10] [11,19] [20,35] [36,50]
HIC1V	IC到时	[5,8] [9,20] [21,100]
OCC1SV	1CC值	[2,9] [10,19] [20,35] [36,48]
OCC2SV	2CC值	[4,10] [11,20] [21,34] [35,50]

表 4 连续属性离散化后的结果(基于类信息熵的方法)

Table 4 Discretization result continuous attributes (classification entropy based approach)

属性	注释	离散化后的区间
CHN	炉次数	[5,11]
SPAN	周期	[100,160]
OB1V	精炼1值	[2,5] [6,17] [18,33]
OB2V	精炼2值	[3,4] [5,16] [17,29]
CCPV	下炉次值	[3,11] [12,19] [20,34][35,50]
HIC1V	IC到时	[5,9] [10,20] [21,100]
OCC1SV	1CC值	[2,10] [11,20] [21,37] [38,48]
OCC2SV	2CC值	[4,11] [12,21] [22,33] [34,50]

5 结论(Conclusion)

在分辨矩阵的基础上, 提出了一种以核为出发点的改进的求解案例属性最小约简的方法. 研究了连续属性的离散化问题, 提出了一种基于逼近精度敏感系数的有监督离散化算法. 以某钢铁企业的实际生产为背景, 对其动态调度案例的属性首先经过连续属性离散化处理, 其次应用属性最小约简的求解方法, 得出该调度领域问题中的属性最小约简. 为下一步的案例匹配工作, 消除了冗

余信息, 提高了案例匹配实现的效率. 算法的编写采用VC++语言编写, 在Pentium 750/128M的兼容机上加以实现. 运行结果表明, 计算速度较快, 去掉了冗余的属性, 达到预期目的. 基于粗糙集理论的属性约简技术以及连续属性的离散化技术的不断发展, 将会为各类CBR系统的案例的建立与匹配提供有力的支持. 该理论在属性繁多的大型复杂CBR系统中具有广泛的应用前景.

参考文献(References):

[1] PAWLAK Z. Rough sets[J]. *Int J of Information and Computer Science*, 1982, 11(5): 341 – 356.

[2] PAWLAK Z, GRZYMALA-BUSSE J W, SLOWINSKI R, et al. Rough sets[J]. *Communications of ACM*, 1995, 38 (11): 89 – 95.

[3] 韩祯祥, 张琦, 文福拴. 粗糙集理论及其应用综述[J]. 控制理论与应用, 1999, 16(2): 153 – 157.
(HAN Zhenxiang, ZHANG Qi, WEN Fushuan. A Survey on rough set theory and its application[J]. *Control Theory & Applications*, 1999, 16(2): 153 – 157.)

[4] WROBLEWSKI J. Finding minimal reducts using genetic algorithms[C] *//Proc of the Second Annual Join Conf on Information Sciences*. Wrightsville Beach, NC: Springer, 1995: 186 – 189.

[5] 陶志, 许宝栋, 汪定伟, 等. 基于遗传算法的粗糙集知识约简方法[J]. 系统工程, 2003, 21(4): 116 – 122.
(TAO Zhi, XU Baodong, WANG Dingwei, et al. Rough set knowledge reduction approach based on GA[J]. *Systems Engineering*, 2003, 21(4): 116 – 122.)

[6] 叶东毅, 陈昭炯. 粗糙集属性量化的一个算法[J]. 小型微型计算机系统, 2002, 1123(10): 1239 – 1240.
(YE Dongyi, CHEN Zhaojiong. An algorithm for attributes quantification in rough set[J]. *Mini-Micro System*, 2002, 1123(10): 1239 – 1240.)

[7] 李玉榕, 乔斌, 蒋静坪. 基于熵的粗糙集属性简约算法[J]. 电路与系统学报, 2002, 17(3): 8 – 12.
(LI Yurong, QIAO Bin, JIANG Jingping. Entropy based attribute reduction algorithms for rough sets[J]. *J of Circuits and Systems*, 2002, 17(3): 8 – 12.)

[8] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001.
(ZHANG Wenxiu, WU Weizhi, LIANG Jiye, et al. *Fuzzy Set Theory and Approach*[M]. Beijing: Science Press, 2001.)

[9] DUNTSCH I, GEDIGA G. Statistical evaluation of rough set dependency analysis[J]. *Int J of Human Computer Study*, 1997, 46(5): 589 – 604.

[10] 唐建国, 谭明术. 粗糙集理论中的求核与约简[J]. 控制与决策, 2003, 18(4): 449 – 452.
(TANG Jianguo, TAN Mingshu. On finding core and reduction in rough set theory[J]. *Control and Decision*, 2003, 18(4): 449 – 452.)

[11] DOUGHERTY J, KOHAVI R, SAHAMI M. Supervised and unsupervised discretization of continuous features[C] *//Proc of the Twelfth Int Conf on Machine Learning* . Tahoe City, CA: Morgan Kaufman, 1995: 194 – 202.

[12] NGUYEN H S. *Discretization of real value attributes: a boolean reasoning approach*[D]. Warsaw: Warsaw University, 1997.

[13] KOHAVI R, SAHAMI M. Error-based and entropy-based discretization of continuous features[C] *// Proc of the Second Int Conf on Knowledge Discovery and Data Mining*. Portland, Oregon, USA: AAAI Press, 1996: 114 – 119.

[14] 汪定伟, 常春光, 胡琨元, 等. 基于多个混沌BP网的案例匹配技术研究[J]. 系统工程学报, 2005, 20(4): 410 – 418.
(WANG Dingwei, CHANG Chunguang, HU Kunyuan, et al. Study on case retrieving technique based on multi-chaotic BP neural network[J]. *J of Systems Engineering*, 2005, 20(4): 410 – 418.)

作者简介:

常春光 (1973—), 男, 在东北大学获博士学位，从事生产计划与调度、MES、建模与决策、智能优化方法与应用软件研制开发工作, E-mail: ccg7788@sohu. com;

汪定伟 (1948—), 男, 教授, 博士生导师, 主要研究方向为生产计划与调度理论、建模与决策、智能优化方法;

胡琨元 (1972—), 男, 博士, 从事生产计划、ERP、建模与决策、智能优化方法与应用软件开发工作;

陶 志 (1963—), 男, 博士, 主要研究方向为专家系统与决策支持系统、粗集理论及其应用等.