粗糙集的划分贴近度及基于划分贴近度的属性约简算法*)

徐久成　孟慧丽　郭林鹏　史进玲

(河南师范大学计算机与信息技术学院　河南新乡 453007)

摘　要　Rough集理论是近年来发展起来的一种处理不确定、不精确、不完整数据的数学工具。属性约简是粗糙集的核心内容之一。本文提出了一个新的不确定性度量—划分贴近度，并基于划分贴近度分别提出了对一般信息系统和决策信息系统进行属性约简的算法，对决策信息系统进行约简的算法不仅可以对一致决策表进行约简，还可以对不一致决策表进行有效的约简。

关键词　Rough集，划分贴近度，属性约简

The Partition Close-degree of Rough Sets with Attribute Reduction Algorithm

XU Jiu-Cheng　MENG Hui-Li　GUO Lin-Peng　SHI Jin-Ling

(School of Computer and Information Technology, Henan Normal University, Xinxiang 453007)

Abstract　Rough set theory is a new developed mathematic tool which can deal with uncertain, imprecise and incomplete data. Attribute reduction is one of the most important aspect in this theory. In this paper, we define a new measure of uncertainty—partition close-degree, and give two attribute reduction algorithms based on it. The first one can deal with information system. The second one can deal with not only a consistent decision table but also a inconsistent decision table with good performance.

Keywords　Rough sets, Partition close-degree, Attribute reduction

1　引言

粗糙集理论是近年来发展起来的一种处理不确定、不精确、不完整数据的数学工具[1]。粗糙集理论自波兰科学家Pawlak于1982年提出以来，已经被成功地应用于机器学习、数据挖掘、软计算等领域[2,3]。属性约简是粗糙集的核心内容之一，目前很多研究者已提出了许多属性约简算法，主要有基于区分矩阵[4,5]和信息熵理论[6~8]的约简算法。

文[7]中提出的CEBARKNC算法是一种比较典型的基于信息熵的属性约简算法，文[8]通过例子指出CEBARKNC算法对于某些不一致决策表约简，会存在冗余属性。本文结合粗糙集的集合运算思想，提出了一个新的不确定性度量—划分贴近度，并基于划分贴近度分别提出了对一般信息系统和决策信息系统进行属性约简的算法，对决策表进行约简的算法不仅可以对一致决策表进行约简，还可以对不一致决策表进行有效的约简。

本文在以下的讨论中采用的均为标准的粗糙集理论术语和记号，有些基本概念就不再重述。

2　粗糙集的划分贴近度

一个信息系统可以定义为四元组 $T=\langle U,A,V,f\rangle$，U 为论域，A 为属性集，当信息系统为决策系统时，$A=C\cup D$，C、D 分别为条件属性集和决策属性集，且 $C\cap D=\phi$。

定义1　若 $T=\langle U,A,V,f\rangle$ 为一信息系统，属性集 P、$Q\subseteq A$，且 P、Q 在 U 上导出的划分分别为 $X=\{X_1,X_2,\cdots,X_m\}$、$Y=\{Y_1,Y_2,\cdots,Y_n\}$，且任意的 $X_i\cap X_j=\phi$，$Y_i\cap Y_j=\phi(i\neq j)$，对任一 X_i、Y_j，定义：

$$p(X_i|Y_j)=\begin{cases}|X_i|/|Y_j|, & \text{其中 } X_i\subseteq Y_j\\ 0, & \text{否则}\end{cases}$$

称 $p(X_i|Y_j)$ 为集合 X_i 对 Y_j 的贴近度。

这里 $|\cdot|$ 表示集合的基数，$0\leqslant p(X_i|Y_j)\leqslant 1$，$p(X_i|Y_j)$ 越大，则集合 X_i 和 Y_j 越接近。

定义2　若 $T=\langle U,A,V,f\rangle$ 为一信息系统，属性集 P、$Q\subseteq A$，P、Q 在 U 上导出的划分分别为 $X=\{X_1,X_2,\cdots,X_m\}$、$Y=\{Y_1,Y_2,\cdots,Y_n\}$，属性集 P 对 Q 的划分贴近度定义为：

$$H(P|Q)=H(X|Y)=\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n}p(X_i|Y_j)$$

这里 $0\leqslant H(P|Q)\leqslant 1$。

定理1　若 $T=\langle U,A,V,f\rangle$ 为一信息系统，某个等价关系 $P1$ 在 U 上形成的划分为 $B1=\{X_1,X_2,\cdots,X_m\}$。将 $P1$ 中某个属性 r 去掉，得到新的属性集合 $P2$。Q 也为 U 上一等价关系，将 U 划分为 $Y=\{Y_1,Y_2,\cdots,Y_n\}$，则 $H(P1|Q)\geqslant H(P2|Q)$。

证明：设属性集 $P2$ 在 U 上形成的划分为 $B2$，因为从 $P1$ 中去掉一个属性 r，会导致对论域分块的合并或不变。

①若 $B1=B2$，即对论域的划分不变，根据定义有 $H(P1|Q)=H(B1|Y)$，$H(P2|Q)=H(B2|Y)$，则 $H(P1|Q)=H(P2|Q)$。

②若 $B1\neq B2$，即在发生分块合并的情况下，设 $P2$ 在 U 上形成的划分 $B2=\{X_1,X_2,\cdots,X_{i-1},X_{i+1},\cdots,X_{k-1},X_{k+1},\cdots,X_m,X_i\cup X_k\}$，即 $B2$ 是将划分 $B1$ 中的某两个等价块 X_i 和 X_k 合并后得到的新划分，记

*)项目基金：河南省自然科学基金项目资助(0511011500)；河南省高校新世纪优秀人才支持计划(2006HANCET-19)资助。**徐久成**　教授，博士，CCF会员，主要研究方向为粗糙集理论、粒计算、数据挖掘等。**孟慧丽**　硕士研究生，主要研究方向为粗糙集理论、数据挖掘。

$$H(P1|Q)=H(B1|Y)=\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n}p(X_i|Y_j)$$

$$H(P2|Q)=H(B2|Y)=H(P1|Q)-\frac{1}{n}[\sum_{j=1}^{n}p(X_i|Y_j)+\sum_{j=1}^{n}p(X_k|Y_j)]+\frac{1}{n}[\sum_{j=1}^{n}p(X_i\cup X_k|Y_j)]$$

$$\therefore H(P1|Q)-H(P2|Q)=\frac{1}{n}[\sum_{j=1}^{n}p(X_i|Y_j)+\sum_{j=1}^{n}p(X_k|Y_j)]-\frac{1}{n}[\sum_{j=1}^{n}p(X_i\cup X_k|Y_j)]=\frac{1}{n}[p(X_i|Y_{j1})+p(X_k|Y_{j2})-p(X_i\cup X_k|Y_{j3})]$$

若 $p(X_i\cup X_k|Y_{j3})>0$ 则根据定义 $X_i\cup X_k\subseteq Y_{j3}$，因此 $X_i\subseteq Y_{j3}$ 且 $X_k\subseteq Y_{j3}$，因为任一 X_i 只能包含于唯一的 Y_j，所以 $Y_{j1}=Y_{j2}=Y_{j3}$，因而 $H(P1|Q)-H(P2|Q)=0$。

若 $p(X_i\cup X_k|Y_{j3})=0$，则根据定义 $X_i\cup X_k$ 不包含于任意的 Y_j，且 $Y_{j1}\neq Y_{j2}$。若 $Y_{j1}=Y_{j2}$，则 $X_i\cup X_k\subseteq Y_{j1}$，与 $X_i\cup X_k$ 不包含于任意的 Y_j 矛盾，所以 $Y_{j1}\neq Y_{j2}$。又因为 $p(X_i|Y_{j1})\geqslant 0$ 且 $p(X_k|Y_{j2})\geqslant 0$，从而有 $H(P1|Q)-H(P2|Q)\geqslant 0$。

综上可得，当 X_i、X_k、$X_i\cup X_k$ 都不包含于任意的 Y_j 时，或同时包含于同一 Y_j 时，$H(P1|Q)-H(P2|Q)=0$；其他情况下 $H(P1|Q)-H(P2|Q)>0$，所以 $H(P1|Q)-H(P2|Q)\geqslant 0$。

由定理 1 可知，对于同一属性集 Q，属性集 P 对 Q 的贴近度随 P 中的属性的增加而呈非递减变化，这说明增加属性会导致对论域 U 的细分，导致划分的贴近度不变或增加。

定义 3 若 $T=\langle U,A,V,f\rangle$ 为一信息系统，属性集 B、$P\subseteq A$，当 $H(B|P)=1$ 时，称 B 完全贴近于 P；当 $H(B|P)=0$ 时，称 B 完全不贴近于 P；当 $0<H(B|P)<1$ 时，称 B 以贴近度 $H(B|P)$ 贴近于 P。

注：当 B 完全贴近于 P 时，这时 B 对 U 的划分等于或细于 P 对 U 的划分。反之不一定成立。

定理 2 若 $T=\langle U,A,V,f\rangle$ 为一信息系统，属性集 B、$P\subseteq A$，若 $H(B|P)=H(P|B)=1$，则 B 对 U 的划分等价于 P 对 U 的划分。

证明：设属性集 B、P 对于论域 U 的划分分别为 $X=\{X_1,X_2,\cdots,X_m\}$、$Y=\{Y_1,Y_2,\cdots,Y_n\}$

则 $$H(B|P)=H(X|Y)=\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n}p(X_i|Y_j)=1$$

且 $$H(P|B)=H(Y|X)=\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{m}p(Y_j|X_i)=1$$

$$\because H(B|P)=\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n}p(X_i|Y_j)=1$$

$$\therefore\sum_{i=1}^{m}\sum_{j=1}^{n}p(X_i|Y_j)=n$$

$$\sum_{j=1}^{n}\sum_{i=1}^{m}p(X_i|Y_j)=n \tag{1}$$

$$\because p(X_i|Y_j)=\begin{cases}|X_i|/|Y_j|, & \text{其中 } X_i\subseteq Y_j\\ 0, & \text{否则}\end{cases}$$

$$\therefore\sum_{i=1}^{m}p(X_i|Y_j)\leqslant 1 \tag{2}$$

由(1)和(2)可知 $\sum_{i=1}^{m}p(X_i|Y_j)=1$，所以对任意的 Y_j，$\exists X_i$ 使得 $X_i\subseteq Y_j$。同理可证 $\sum_{j=1}^{n}p(Y_j|X_i)=1$，对 X_i，$\exists Y_k$ 使得 $Y_k\subseteq X_i$，若 $Y_k\neq Y_j$，则 $Y_k\cap Y_j\neq\phi$，与定义 $Y_k\cap Y_j=\phi$ 矛盾，从而有 $X_i=Y_j=Y_k$。因此对任意的 X_i 只能包含于唯一的 Y_j，同样对任意的 Y_j 只能包含于唯一的 X_i，所以属性集 B 对 U 的划分等价于 P 对 U 的划分。

定义 4 若 $T=\langle U,A,V,f\rangle$ 为一信息系统，属性集 $P\subseteq A$，对属性集 P 中任意属性 $a\in P$，属性 a 在 P 中的重要性为 $SGF(a,P)=1-H(P-\{a\}|P)$。当 $SGF(a,P)>0$，说明 a 是 P 中必要的；当 $SGF(a,P)=0$ 时，说明 a 是 P 中不必要的。若每一个 $a\in P$ 都为 P 中必要的，则称 P 为独立的，否则称 P 为依赖的。

定理 3 若 $T=\langle U,A,V,f\rangle$ 为一信息系统，对属性集 B、P，设 $B\subseteq P$，如果 B 是独立的且 $H(B|P)=1$，则 B 为 P 的一个约简。

证明：由定理 2 的证明可知当 $H(B|P)=1$ 时，$\sum_{i=1}^{m}p(X_i|Y_j)=1$，因此对任意的 Y_j，$\exists X_i$ 使得 $X_i\subseteq Y_j$，又因为 $B\subseteq P$，所以 P 对 U 形成的划分应等于或细于 B 对 U 形成的划分，因此对任意的 Y_j，$\exists X_k$ 使得 $Y_j\subseteq X_k$，从而有 $X_i\subseteq Y_j\subseteq X_k$，若 $X_i\neq X_k$，则 $X_i\cap X_k\neq\phi$，与定义 $X_i\cap X_k=\phi$ 矛盾，从而有 $X_i=Y_j=X_k$，所以 B、P 对 U 形成的划分是相同的，又因为 B 是独立的，B 中每个属性都是必要的，所以 B 是 P 的约简。

定义 5 若 $T=\langle U,A,V,f\rangle$ 为决策信息系统，$A=C\cup D$，C 为条件属性集，D 为决策属性集，属性集 $P\subseteq C$ 中任意属性 $a\in P$，属性 a 在 P 中的重要性为 $SGF(a,P,D)=H(P|D)-H(P-\{a\}|D)$。当 $SGF(a,P,D)>0$，说明 a 是 P 相对于 D 必要的；当 $SGF(a,P,D)=0$ 时，说明 a 是 P 相对于 D 不必要的。如果 P 中每个属性 a 都为 D 必要的，则 P 为 D 独立的。

定理 4 若 $T=\langle U,A,V,f\rangle$ 为决策信息系统，$A=C\cup D$，C 为条件属性集，D 为决策属性集，如果 B 是 D 独立的且 $B\subseteq P\subseteq C$，则 B 为 P 相对于决策属性集 D 的约简的充分必要条件为：$H(B|D)=H(P|D)$。

证明：(1)(充分性) 当 B 为 P 相对于决策属性集的约简时，则有 $Pos_B(D)=Pos_P(D)$。设 $Pos_B(D)=Pos_P(D)=\{x_1,x_2,\cdots,x_k\}$，则有 $H(B|D)=H(P|D)=\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n}p(x_i|Y_j)$，从而 $H(B|D)=H(P|D)$。

(2)(必要性) 因为 $B\subseteq P$，所以 P 对 U 形成的划分等于或细于 B 对 U 形成的划分。当 P 对 U 形成的划分等于 B 对 U 形成的划分时，因为 B 是 D 独立的，则 B 显然为 P 的约简；当 P 对 U 形成的划分细于 B 对 U 形成的划分时，设 P 对 U 形成的划分为 $X=\{X_1,X_2,\cdots,X_m\}$，B 对 U 形成的划分为 $Z=\{X_1,X_2,\cdots,X_{i-1},X_{i+1},\cdots,X_{k-1},X_{k+1},\cdots,X_m,X_i\cup X_k\}$，$Z$ 是将划分 X 中的某两个等价块 X_i、X_k 合并后得到的划分，设决策属性集 D 对论域 U 形成的划分为 $Y=\{Y_1,Y_2,\cdots,Y_n\}$。

$$H(P|D)=\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n}p(X_i|Y_j)$$

$$H(B|D)=H(P|D)-\frac{1}{n}\sum_{j=1}^{n}p(X_i|Y_j)-\frac{1}{n}\sum_{j=1}^{n}p(X_k|Y_j)+\frac{1}{n}[\sum_{j=1}^{n}p(X_i\cup X_k|Y_j)]$$

所以当 $H(B|D)=H(P|D)$ 时，则有

$$\sum_{j=1}^{n}p(X_i|Y_j)+\sum_{j=1}^{n}p(X_k|Y_j)=[\sum_{j=1}^{n}p(X_i\cup X_k|Y_j)]$$

所以当且仅当 X_i，X_k 都是同一个 Y_j 的子集或都不是任意的 Y_j 的子集时，上式左右相等。当 X_i，X_k 都是同一个 Y_j

的子集时有 $X_i \cup X_k \subseteq Y_j$，所以 $Pos_B(D)=Pos_P(D)$；当 X_i，X_k 都不是任意的 Y_j 的子集时，同样有 $Pos_B(D)=Pos_P(D)$，又因为 B 是 D 独立的，所以 B 是 P 相对于 D 的约简。

定理 5 若 $T=\langle U,A,V,f\rangle$ 为决策信息系统，$A=C\cup D$，C 为条件属性集，D 为决策属性集，C、D 对于论域 U 的划分分别为 $X=\{X_1,X_2,\cdots,X_m\}$、$Y=\{Y_1,Y_2,\cdots,Y_n\}$。若 $H(C|D)=1$，则该决策表为一致的；若 $H(C|D)<1$，则该决策表为不一致的。

证明：当 $H(C|D)=1$ 时，即

$$H(C|D)=\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n}p(X_i|Y_j)=1$$

$$\because p(X_i|Y_j)=\begin{cases}|X_i|/|Y_j|, & X_i\subseteq Y_j\\ 0, & \text{否则}\end{cases}$$

$\therefore$ 当 $X_i\subseteq Y_j$ 时，$X_i\subseteq Pos_C(D)$

$Pos_C(D)=\bigcup X_i(X_i\subseteq Y_j)$

设 $Pos_C(D)=\{x_1,x_2,\cdots,x_k\}$，所以

$$H(C|D)=\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n}p(X_i|Y_j)=\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n}p(x_i|Y_j)=1$$

$$\sum_{i=1}^{k}\sum_{j=1}^{n}p(x_i|Y_j)=n$$

$$\sum_{j=1}^{n}\sum_{i=1}^{k}p(x_i|Y_j)=n$$

$$\therefore\sum_{i=1}^{k}p(x_i|Y_j)=1$$

即对任意 Y_j，$Pos_C(D)$ 中包含 Y_j 的所有元素，所以 $Pos_C(D)=\bigcup X_i=U\ (X_i\subseteq Y_j)$，由于根据 C 所得 U 的划分可以完全包含到 D 对 U 的划分中，所以决策表为一致的；当 $H(C|D)<1$ 时，即 $Pos_C(D)=\bigcup X_i\subset U(X_i\subseteq Y_j)$，根据 C 所得 U 的划分部分可以包含到 D 对 U 的划分中，所以决策表为不一致的。

3 基于划分贴近度的属性约简算法

我们根据上面的理论设计了两个对一般信息系统和决策信息系统进行约简的算法，算法 1 是对一般信息系统进行属性约简，算法 2 是对决策信息系统进行约简。两个算法都从属性集中依次去掉对划分没有影响的属性，最后得到属性集的一个约简。算法 2 不仅可以对一致决策表进行约简，同时可以对不一致决策表进行有效的属性约简。

由定理 3 的证明可知，对一个一般信息系统 $T=\langle U,P,V,f\rangle$，当属性集 $B1\subseteq P$，且 $H(B1|P)=1$ 时，$B1$ 与 P 对 U 形成的划分是相同的，所以采用 $1-H(B1-\{a_i\}|P)$ 作为启发信息，当 $1-H(B1-\{a_i\}|P)=0$，说明从 $B1$ 中去掉属性 a_i，$B1-\{a_i\}$ 与 P 对 U 形成的划分仍然是相同的；当 $1-H(B1-\{a_i\}|P)>0$，说明从 $B1$ 中去掉属性 a_i，$B1-\{a_i\}$ 对 U 形成的划分与 P 对 U 形成的划分不同。所以可以依次去掉 $B1$ 中不影响对论域划分的属性，最后得到 P 的一个约简。

算法 1 对一般信息系统进行属性约简的算法

输入：一个一般信息系统 $T=\langle U,P,V,f\rangle$，U 为论域，P 为属性集，$P=\{a_1,a_2,\cdots,a_n\}$。
输出：属性集 P 的一个约简 B。
Step1. 令 $B1=P=\{a_1,a_2,\cdots,a_n\}$，$Att=\phi$。
Step2. 如果 $B1-Att=\phi$，则 $B=B1$，转 Step5；否则任选 $B1-Att$ 中一个属性 a_i，计算 $1-H(B1-\{a_i\}|P)$。
Step3. 若 $1-H(B1-\{a_i\}|P)=0$，则 $B1=B1-\{a_i\}$，转 Step2。
Step4. 若 $1-H(B1-\{a_i\}|P)>0$，则 $Att=Att\cup\{a_i\}$，转 Step2。
Step5. 终止。

由定理 4 的证明可知，对决策系统 $T=\langle U,C\cup D,V,f\rangle$，当属性集 $B1\subseteq C$，且 $H(C|D)=H(B1|D)$ 时，$Pos_{B1}(D)=Pos_C(D)$，此时 $B1$ 与 C 对 U 的划分可能不同，但形成的划分相对于 D 的正域是相同的，所以采用 $H(C|D)-H(B1-\{a_i\}|D)$ 作为启发信息，当 $H(C|D)-H(B1-\{a_i\}|D)=0$ 时，说明从 $B1$ 中去掉 a_i，$Pos_{B1-\{a_i\}}(D)=Pos_C(D)$；当 $H(C|D)-H(B1-\{a_i\}|D)>0$ 时，$Pos_{B1-\{a_i\}}(D)\neq Pos_C(D)$。所以可以依次去掉 $B1$ 中不使 D 正域发生变化的属性，最后得到 C 相对 D 的一个约简。

算法 2 对决策信息系统进行属性约简的算法

输入：一个决策信息系统 $T=\langle U,C\cup D,V,f\rangle$，$U$ 为论域，C 为条件属性集，D 为决策属性集，$C=\{a_1,a_2,\cdots,a_n\}$。
输出：属性集 C 的一个相对约简 B。
Step1. 令 $B1=C$，$Att=\phi$。
Step2. 如果 $B1-Att=\phi$，则 $B=B1$，转 Step5；否则任选 $B1-Att$ 中一个属性 a_i，计算 $H(C|D)-H(B1-\{a_i\}|D)$。
Step3. 若 $H(C|D)-H(B1-\{a_i\}|D)=0$，则 $B1=B1-\{a_i\}$，转 Step2。
Step4. 若 $H(C|D)-H(B1-\{a_i\}|D)>0$，则 $Att=Att\cup\{a_i\}$，转 Step2。
Step5. 终止。

算法复杂性分析：设一般信息系统中，$|U|=n$，$|P|=m$，通过对算法 1 进行分析，可得算法 1 时间复杂度为 $O(m^2n^2)$；设决策表中 $|U|=n$，$|C|=m$，通过对算法 2 进行分析，可得算法 2 时间复杂度也为 $O(m^2n^2)$。两个算法的时间复杂度主要由计算属性对论域的划分所引起。

4 实例分析

为了考察以上算法的有效性，我们采用文[9]中的一个一般信息系统对算法 1 作一个测试，另外采用文[8]中的一个不一致决策表对算法 2 进行测试。

表 1 一般信息系统

U	a	b	c	d
1	0	1	2	0
2	1	2	0	2
3	1	0	1	1
4	2	1	0	1
5	1	1	0	2

表 2 不一致决策表

U	C		D	
	a	b	c	d
U_1	1	0	2	−1
U_2	1	1	0	−1
U_3	1	1	0	0
U_4	1	0	2	−1
U_5	1	1	2	0
U_6	1	1	0	−1
U_7	0	1	1	1
U_8	0	1	0	1
U_9	0	1	0	0
U_{10}	0	1	1	1

表 1 为文[9]中的一个一般信息系统，属性集 $P=\{a,b,c,d\}$，文[9]中采用的是区分矩阵的方法，可以得到两个最小约简为 $\{a,b\}$ 和 $\{b,d\}$，利用我们提出的算法 1，根据选择属性的顺序不同，总可以得到该信息系统属性的一个约简为 $\{a,b\}$ 或 $\{b,d\}$。

表 2 为文[8]中的一个不一致决策表，条件属性集 $C=\{a,b,c\}$，决策属性集 $D=\{d\}$。文[8]指出采用文[7]中的 CEBARKNC 算法，得出任何属性均不可约简，根据文[8]中

（下转第 260 页）

和信息需要间的关系的抽象通过可测量的概念这一元素表达。在上面的例子中，应当是“软件开发的生产率”。作为可测量的属性，将采用开发产品的规模和开发的工作量表示。

所有可测量的属性与度量相关，度量是用于量化的不同类型测量的进一步抽象，并为相关实体进行决策。所有度量均有隶属于某一度量刻度下的测量单元[12]。三种度量类型为：

(1)基测量

基测量的数学描述为：$y=f(x)$，其中 y 为基测量目标值，f 为测量方法，x 为属性。测量方法由一系列的逻辑运算构成，通常用于量化某一刻度下的属性[13]，如统计某时间段内的数量或经历的时间等。

(2)派生测量

软件派生测量的数学模型为 $z=g(y_1,y_2,\cdots,z_1,z_2,\cdots)$，其中 z 为目标派生测量值，g 为测量函数，$y_1,y_2,\cdots$ 为基测量值，$z_1,z_2,\cdots$ 为派生测量值。测量函数是在两个或多个基测量基础上的算法或运算[13]。

(3)指示器

指示器的形式化描述：

$$\begin{cases} W=h(y_1,y_2,\cdots,z_1,z_2,\cdots) \\ w\in g, \text{其中 } g \text{ 为决策准则或决策条件} \end{cases}$$

其中 W 为指示器目标值，h 为度量分析模型，$y_1,y_2,\cdots$ 为基本度量值，$z_1,z_2,\cdots$ 为派生测量值[13]。

分析模型产生与已定义的信息需求相关的估算或评价。它由与确定的相关的决策标准结合了一个或多个基测量与/或派生测量的算法或计算得到。所有的决策标准由一系列的限值、确定研究需要所使用的对象或达到确定结果相关的可信度构成。这些标准用于解释度量结果。

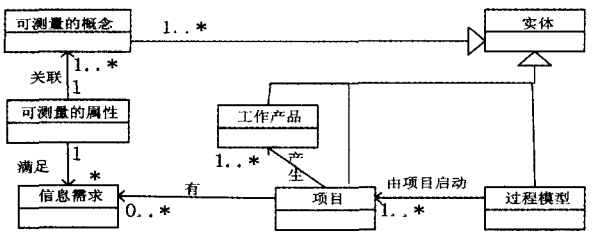

图 3 软件过程元模型和度量元模型间的关系

使用图 2 给出的元模型可度量过程或数据模型的任何元素。图 3 给出软件过程元模型和软件度量元模型的主要关系。从图 3 我们可以看出，任何过程模型都是在具体的项目下执行的。作为软件项目执行的结果，产生工作产品且所有的软件项目需要满足某些信息需求。

小结 创建度量的过程模型是软件度量过程的难点之一，本文对度量的概念模型研究的过程中，为保证术语的一致性和度量的可复用性，采用已有的度量标准 ISO15939 作为基础。通过实践证明度量的过程概念模型的建立有助于减少对度量的误解，提高度量的有效性。

参 考 文 献

1 Koch F J. Metrics and the Immature Software Process. 2000. http://www.qpmg.com/metrics.htm

2 Brockers A, Differding C. The Role of Software Process Modeling in Planning Industrial Measurement Programs. In: Proceedings of the 3rd International Metrics Symposium, Berlin, March 1996

3 Morisio M. A methodology to measure the software process. In: Proceedings of the 7th Annual Oregon Workshop on Software Metrics, 1995

4 Aarsten A, Morisio M. Using object oriented technology to measure a software process. In: Proceedings of Achieving Quality in Software (AQUIS96), Florence, Italy

5 Webby R G, Becker U. Towards a Logical Schema Integrating Software Process Modeling and Software Measurement. In: Proceedings of the Process Modeling and Studies of Software Evolution Workshop, ICSE97, ACM and IEEE, Boston, May 1997

6 李娟，李明树，武占春，等. 基于 SPEM 的 CMM 软件过程元模型. 软件学报，2005，16(8)

7 李娟，袁峰，李明树. 一种基于模型融合的 CMM 实施过程建模方法. 计算机学报，2006(1)

8 García F, Ruiz F, Cruz J A, et al. Integrated Measurement for the Evaluation and Improvement of Software Processes, 2003

9 Object Management Group. Meta object facility (MOF) specification. Version1.4, formal/02-04-03; Needham, Object Management Group, 2002. http://doc.omg.org/formal/02-04-03

10 McGarry J, Card D, Jones C, et al. Practical software Measurement: Objective Information for Decision Makers, Pearson Education, Inc, 2001, 10

11 International Organization for Standard. Software Engineering: Software measurement process, ISO/IEC 15939. 2002(E)

12 侯红，郝克刚. 软件测量刻度及选择方法. 计算机科学，2005，32(5)

13 李兴南. 软件测试度量的研究及其工具 STMT 开发：[硕士论文]. 西北大学，2005

(上接第 215 页)

的改进算法可得约简为{b,c}，根据我们提出的算法 2，也可得到该不一致决策表的约简为{b,c}。

结论 本文所提出的划分贴近度理论是基于粗糙集的集合思想，所提出的属性约简算法 1 能够对一般信息系统进行有效的约简，算法 2 对一致和不一致决策信息系统都能进行有效的约简，并通过实例验证了两个算法的有效性。

参 考 文 献

1 Pawlak Z. Rough Set-Theoretical Aspect of Reasoning About Data [M]. Kluwer Academic pub, 1991

2 Lingras P J, Yao Y Y. Data mining using extensions of the rough set model[J]. Journal of the American Society for Information Science, 1998, 49(5): 415～422

3 Li Y C, Fang T J. Rough Set Methods for Constructing Support Vector Machines [A]. In: 9th International Conference Rough Sets, Fuzzy Sets, Data Mining And Granular Computing[C]. Chongqing China, 2003. 334～338

4 Skowron A, Rauszer C. The discernibility matrices and functions in information[A]. Intelligent Decision Support Handbook of Applications and Advances of the Rough Sets Theory[C]. Kluwer Academic Pub, 1992. 331～338

5 WangJ. Reduction Algorithms Based on Discernibility Matrix The Ordered Attributes Method[J]. Jounary Computer Science Technology, 2001, 16(6): 498～504

6 苗夺谦，胡桂荣. 知识约简的一种启发式算法[J]. 计算机研究与发展，1999，36(6)：681～684

7 王国胤，于洪，杨大春. 基于条件信息熵的决策表约简[J]. 计算机学报，2002，25(7)：759～766

8 李明，黄文涛，刘智云. 关于决策表约简的 CEBARKNC 算法改进[J]. 计算机应用，2006，26(4)

9 张文修，吴伟志，梁吉业，李德玉. 粗糙集理论与方法[M]. 北京：科学出版社，2001