# 一种基于规则的属性约简算法*)

杨 明[1] 杨 萍[2] 孙志挥[1]

(东南大学计算机科学与工程系 南京210096)[1] (安徽工程科技学院数理系 芜湖241000)[2]

## An Attribute Reduction Algorithm Based on Rules

YANG Ming[1] YANG Ping[2] SUN Zhi-Hui[1]

(Department of Computer Science and Engineering, Southeast University, Nanjing 210096)[1]

(Department of Math and Phys., Anhui University of Technology and Science, Wuhu 241000)[2]

yangm_163@163.com

**Abstract** Reduction of attributes is one of important topics in the research on rough set theory. Wong S K M and Ziarko W have proved that finding the minimal attribute reduction of decision table is a NP-hard problem. Algorithm A (the improved algorithm to Jelonek) choices optimal candidate attribute by using approximation quality of single attribute, it improves efficiency of attribute reduction, but yet exists the main drawback that the single atribute having maximum approxiamtion quality is probably optimal candidate attribute. Therefore, in this paper, we introduce the concept of compatible decision rule, and propose an attribute reduction algorithm based on rules (ARABR). Algorithm ARABR provides a new method that measures the relevance between extending attribute and the set of present attributes, the method assures that the optimal attribute is extended, and obviously reduces the search space. Theory analysis shows that algorithm ARABR is of lower computational complexity than Jelonek's algorithm, and overcomes effectively the main drawback of algorithm A.

**Keywords** Rough set, Decision rule, Confidence measure

## 1 引言

波兰数学家 Pawlak Z 提出的 Rough Set(RS,粗集)是一种新的处理不精确、不完全与不相容知识的数学方法[1,2]。目前,它正在被广泛应用于人工智能、模式识别与智能信息处理等领域,并取得了一定的成果[3]。

属性约简是粗集理论及应用研究的重要内容之一[4],也是知识获取的关键步骤。属性约简作为粗集理论及应用研究的热点,备受研究者的关注。王珏等提出了基于差别矩阵[5]的有效约简策略[6],该策略利用了一些启发性知识来获取属性约简,对于某些问题能够在一定程度上获得较优的属性约简。所谓属性约简,就是在保持信息系统或决策表的分类或决策能力不变的条件下,删除其中不重要的属性。本文主要讨论对决策表的属性约简,又称属性的相对约简。

一个决策表的属性约简一般不是唯一的,即对同一决策表可能存在多个相对约简。因属性约简旨在挖掘决策表的决策规则,约简中属性的多少直接影响着决策规则的繁简。因而,如何快速有效地找到具有最小属性的约简(简称最小属性约简),受到研究者的重视,不幸的是,Wong S K M 和 Ziarko W 已证明寻找决策表的最小属性约简是 NP-hard 问题[7],而属性的组合爆炸是导致 NP-hard 的主要原因。

众所周知,利用人工智能的启发性方法可有效地解决 NP-hard 问题。于是,研究者提出了一些有效的最小属性约简的启发式算法[8~11],其中,Jelonek[10]等人提出的算法是比较典型的一个(它与 HU 的算法[9]本质上是一样的)。针对 Jelonek 算法必须计算很多不同属性集的近似精度才能决定如何扩展候选属性约简这一不足,文[11]提出了算法 A (Jelonek 算法的改进),算法 A 使 Jelonek 算法的计算复杂度降低了一个数量级,但算法 A 存在的主要不足为:它是利用单属性的近似精度决定如何扩展候选属性约简,未考虑扩展属性和已选择属性集之间的相关性,因而,近似精度最大的单属性未必是首选的候选属性。为此,本文在引入相容决策规则概念之后,提出了基于规则的属性约简算法 ARABR,该算法给出了一种通过约束属性集的确定性决策规则的频率之和来度量扩展属性与已选择属性集的相关度,使相关度最大的候选属性得以扩展,并使搜索空间显著缩小。算法 ARABR 提供了一种属性约简的新方法,该方法可使 Jelonek 算法的计算复杂度降低一个数量级,且弥补了算法 A 存在的不足,因而确保了算法的高效率和有效性。

## 2 Rough Set 概念

粗集理论的要点是将分类与知识联系在一起,并用等价类关系形式化表示分类。可理解为:知识是使用等价类 R 对离散空间 U 的划分,记为 $U/R=\{X_1,X_2,\cdots,X_n\}$,称为 $X_i$ 为 U/R 的等价类。为描述方便,用 card(.)表示集合的基数,$\varnothing$ 表示空集。

### 2.1 决策表

决策表 DT 是一个四元组(U,C∪D,V,F),其中,U 是一

*)本课题得到国家自然科学基金(项目编号79970092)及安徽省教育厅自然科学研究基金资助(项目编号2001kj050)。杨 明 副教授,博士生,主要研究方向为知识发现与数据挖掘及粗集理论及应用等。杨 萍 讲师,主要研究方向为管理决策、知识发现及数据挖掘等。孙志挥 教授,博士生导师,主要研究方向为复杂系统信息集成和数据库系统及应用等。

组对象的非空有限集合，称为论域；设有 n 个对象，则 U 可表示为：$U=\{x_1,x_2,\cdots,x_n\}$。C∪D 为属性的有限集，C 表示条件属性集，D 表示决策属性集且 $C\cap D=\varnothing$，D 通常只含有一个属性，V 为属性的值域集；$f:U\times C\cup D\to V$，f 为信息函数，定义对象的属性值。

### 2.2 无差别关系

对于决策表 DT，$B\subseteq C\cup D$，无差别关系 IND(B)定义为 $\{(x,y)\in U\times U\mid \forall a\in B, f(x,a)=f(y,a)\}$，通过 IND(B)将 U 划分为若干个类 $E_i(1\leqslant i\leqslant \mathrm{card}(U/IND(B)))$。

### 2.3 差别矩阵

对于决策表 DT，$B\subseteq C\cup D$，差别矩阵(discernbility matrix)是一个 n×n 的方阵 $M_D(B)=\{M_D(i,j)\}_{n\times n}$，$1\leqslant i\leqslant n=\mathrm{card}(U/IND(B))$。矩阵单元的定义如下：

$M_D(i,j)=\{a\in B: a(E_i)\neq a(E_j)\}$，$i,j=1,2,\cdots,n$。

矩阵单元的内容是属性集，它表明了两个类在该属性集上的值不同，差别矩阵是一个对称矩阵。

## 3 基于规则的属性约简算法—ARABR 算法

### 3.1 ARABR 算法思路

类似于文[11]，不失一般性，假设决策表 DT 仅有一个决策属性 d，其取值范围是 $1,2,\cdots,k$。由 d 决定的等价类构成 U 的一个划分：$\{Y_1,Y_2,\cdots,Y_k\}$，其中，$Y_i=\{x\in U\mid f(x,d)=i\}$，$i=1,2,\cdots,k$。

**定义1** 设 $P\subseteq C$，对划分 $\{Y_1,Y_2,\cdots,Y_k\}$ 的 P-近似精度(approximation quality)为[4]：

$$\gamma_P=\sum_{i=1}^{k}\mathrm{card}(\underline{PY_i})/\mathrm{card}(U) \qquad (1)$$

其中，$\underline{PY_i}$ 为 $Y_i$ 的 P-下近似(lower approximation)。

**定义2** 设 $R\subseteq C$，若 $\gamma_R=\gamma_C$，且不存在 $R'\subset R$，使得 $\gamma_{R'}=\gamma_R$，则称 R 为 C 的一个属性约简。所有 C 的属性约简的交称为 C 的核，记为 CORE(C)。

**性质1** 如果 R 满足 $\gamma_R=\gamma_C$，且 R=CORE(R)，则 R 为 C 的一个属性约简。

表1 决策表

| TID | $A_1$ | $A_2$ | $A_3$ | $A_4$ | d |
|---|---|---|---|---|---|
| 1 | 1 | 2 | 1 | 2 | 1 |
| 2 | 2 | 2 | 3 | 2 | 1 |
| 3 | 1 | 3 | 1 | 4 | 1 |
| 4 | 1 | 3 | 1 | 4 | 1 |
| 5 | 2 | 3 | 3 | 4 | 2 |

**例1** 表1是一决策表，共有5个对象，其中，$C=\{A_1,A_2,A_3,A_4\}$为条件属性集，$D=\{d\}$为决策属性集。$\{A_1,A_2\}$，$\{A_1,A_4\}$，$\{A_3,A_2\}$，$\{A_3,A_4\}$为决策表的属性约简，可见，CORE(C)=∅。由(1)式可得 $\gamma_{A1}=\gamma_{A3}=3/5=0.6$，$\gamma_{A2}=\gamma_{A4}=2/5=0.4$，若通过单属性的近似精度来扩展候选属性约简[11]，则依次选择的候选属性为 $A_1,A_3,A_2,A_4$，这显然是不合理的。在选择了属性 $A_1$ 之后，下一个候选属性应是 $A_2$ 或 $A_4$，而不是 $A_3$，这是因为 $A_3$ 与已选择的属性集 $\{A_1\}$ 的相关度为0，可见，单属性的近似精度不能反应候选属性与已选择属性集的相关度。为此，下文引入了决策规则的置信度及相容决策规则概念，并提出了一种度量候选属性与已选择属性集的相关度的新方法。

**定义3** 设 $E_i\in U/IND(C)(1\leqslant i\leqslant \mathrm{card}(U/IND(C)))$，$Y_j\in U/IDN(D)(1\leqslant j\leqslant k)$，若 $E_i\cap Y_j\neq\varnothing$，则称 $Des(E_i,C)\to Des(Y_j,D)$ 为决策表 DT 的决策规则(也可表示成表2的形式)。其中，$Des(E_i,C)$ 表示属性集 C 上等价类 $E_i$ 的描述，即等价类 $E_i$ 对应于属性集 C 中各属性值的特定取值。

**定义4** 设 $E_i\in U/IND(C)$，$Y_j\in U/IND(D)$，其中，$1\leqslant i\leqslant \mathrm{card}(U/IND(C))$，$1\leqslant j\leqslant k$，称

$$\mu_c(E_i,Y_j)=\mathrm{card}(E_i\cap Y_j)/\mathrm{card}(E_i)$$

为决策规则 $R:Des(E_i,C)\to Des(Y_j,D)$ 的置信度，其中，$E_i\cap X_j\neq\varnothing$，简记 $R.conf=\mu_c(E_i,Y_j)$，$R.count=\mathrm{card}(E_i)$，称 R.count 为 R 的频度。若 R.conf=1，则称 R 为确定性决策规则。

**定义5** $R_1:Des(E_1,S)\to Des(Y_1,D)$ 和 $R_2:Des(E_2,T)\to Des(Y_1,D)$ 为决策表 DT 的两条决策规则；若 $S\subseteq T$，$Y_1=Y_1$，且 $E_2\subseteq E_1$，则称 $R_1$ 和 $R_2$ 是相容决策规则(也称 $R_2$ 为由 $R_1$ 诱导出的决策规则)。记为 $R_1<R_2$。

**定理1** 若 $R:Des(E,S)\to Des(Y_1,D)$ 是决策表 DT 的一决策规则($S\subseteq C$)，且 $R.conf\geqslant\alpha$，则对任意满足 $S\subset T\subseteq C$ 的属性集 T，一定存在由 R 诱导出的决策规则 R′使得 $R'.conf\geqslant\alpha$。

证明：反证法。假设对任意满足 $S\subset T\subseteq C$ 的属性集 T，T 中的任一满足 $R<R'$ 的决策规则 R′均有 $R'.conf<\alpha$。由 $E\in U/IND(S)$，则必存在 $E_1,\cdots,E_p\in U/IND(T)$ 使得 $E=E_1\cup\cdots\cup E_p$，由假使可知，任意 $R_i:Des(E_i,T)\to Des(Y_1,D)$ 均有 $R_i.conf<\alpha$，即 $\mathrm{card}(E_i\cap Y_1)<\alpha\times\mathrm{card}(E_i)$，其中，$i=1,\cdots,p$；则

$$\sum_{i=1}^{p}card(E_i\cap Y_j)<\alpha\times\sum_{i=1}^{p}card(E_i) \qquad (2)$$

即 $\mathrm{card}(E\cap Y_1)<\alpha\times\mathrm{card}(E)$，矛盾。证毕。

**定理2** R 是决策表 DT 的一决策规则，R′是由 R 诱导出的任意决策规则，若 R.conf=1，则使得 R′.conf=1。

证明：类似定理1可证。

由定理1可知，通过逐步增加候选属性可以得到置信度单调递增的相容决策规则序列，因而，可尽早发现置信度为1的决策规则。由定理2可知，若对条件属性 C 的某属性子集 B，存在决策规则 $R:Des(E,B)\to Des(Y_1,D)$ 的置信度 R.conf=1，则由 R 诱导出任意决策规则的置信度均为1(等价于 $E\subset Y_1$)，即有 $E_1,\cdots,E_q\in U/IND(C)$ 使得 $E=E_1\cup\cdots\cup E_q$，且 $R_i:Des(E,C)\to Des(Y_1,D)$ 的置信度 $R_i.conf=1$(等价于 $E_i\subset Y_1$，即 $E\subset\underline{CY_1}$)，其中，$i=1,\cdots,q$。因此，我们得到通过确定性决策规则的频率求 $\mathrm{card}(\underline{CY_1})$ 的新方法。下面通过对实例1的决策表计算 C-近似精度来说明。

首先求各单属性对应的决策规则，如表2所示。从表2不难看出，属性集 $\{A_1\}$ 的近似精度由其所有确定性决策规则的频率之和(记为 $R(\{A_1\}).count$，也可记为 $R(\{A_1\}\mid\varnothing).count$)决定。不难看出，$R(\{A_1\}).count=R(\{A_3\}).count=3$，$R(\{A_2\}).count=R(\{A_4\}).count=2$。类似文[11]的思路，可选择属性 $A_1$ 作为最佳候选属性。按文[11]，下一个候选属性为 $A_3$，这显然是不合适的。由定理2可知，由 $\{A_1\}$ 的确定性规则 $R_{11}$ 诱导出的决策规则均为确定性决策规则，因而可不考虑 $R_{11}$ 所确定的所有对象。在不考虑表1中的对象1，3，4的情况下，得到约束属性集 $\{A1,A_2\}\mid\{A_1\}$，$\{A_1,A_3\}\mid\{A_1\}$，$\{A_1,A_4\}\mid\{A_1\}$ 对应的决策规则如表3所示，其中，$\{A_1,A_2\}\mid\{A_1\}$ 表示去掉 $\{A_1\}$ 的确定性规则 $R_{11}$ 所确定的所有对象后，可由 $\{A_1,$

$A_2\}$得到的决策规则。不难看出，$R(\{A_1,A_2\}|\{A_1\}).count=R(\{A_1,A_4\}|\{A_1\}).count=2$，$R(\{A_1,A_3\}|\{A_1\}).count=0$；可见，下一个最佳候选属性为$A_2$或$A_4$，且$R(\{A_1\}).count+R(\{A_1,A_2\}|\{A_1\}).count=5=card(\underline{CY_1})+card(\underline{CY_2})$，其中，$Y_1=[1]_{(d)}$，$Y_2=[2]_{(d)}$。因而，可得到决策表(表1)的一个属性的相对约简$\{A_1,A_2\}$。可以看出，表3的约束属性集的确定性规则的频度之和反映了候选属性与已选择属性集的相关度，避免了仅靠单属性的近似精度的大小来选择最佳属性的缺陷；同时，约束属性集无需考虑已选择属性集的确定性规则所确定的对象集合，显著地缩小了搜索空间，提高了计算速度(详细说明见第3.3节的算法分析)。

表2 各单属性对应的决策规则

| 属性集 | 决策规则 | 频度 | 置信度 |
|---|---|---|---|
| $\{A_1\}$ | $R_{11}:A_1=1\rightarrow d=1$ | 3 | 1 |
| | $R_{12}:A_1=2\rightarrow d=1$ | 2 | 0.5 |
| | $R_{13}:A_1=2\rightarrow d=2$ | 2 | 0.5 |
| $\{A_2\}$ | $R_{21}:A_2=2\rightarrow d=1$ | 2 | 1 |
| | $R_{22}:A_2=3\rightarrow d=1$ | 3 | 0.67 |
| | $R_{23}:A_2=3\rightarrow d=2$ | 3 | 0.33 |
| $\{A_3\}$ | $R_{31}:A_3=1\rightarrow d=1$ | 3 | 1 |
| | $R_{32}:A_3=3\rightarrow d=1$ | 2 | 0.5 |
| | $R_{33}:A_3=3\rightarrow d=2$ | 2 | 0.5 |
| $\{A_4\}$ | $R_{41}:A_4=2\rightarrow d=1$ | 2 | 1 |
| | $R_{42}:A_4=4\rightarrow d=1$ | 3 | 0.67 |
| | $R_{43}:A_4=4\rightarrow d=2$ | 3 | 0.33 |

表3 各约束属性集对应的决策规则

| 约束属性集 | 决策规则 | 频度 | 置信度 |
|---|---|---|---|
| $\{A_1,A_2\}|\{A_1\}$ | $R_{11}:A_1=2\wedge A_2=2\rightarrow d=1$ | 1 | 1 |
| | $R_{12}:A_1=2\wedge A_2=3\rightarrow d=2$ | 1 | 1 |
| $\{A_1,A_3\}|\{A_1\}$ | $R_{21}:A_1=2\wedge A_3=3\rightarrow d=1$ | 2 | 0.5 |
| | $R_{22}:A_1=2\wedge A_3=3\rightarrow d=2$ | 2 | 0.5 |
| $\{A_1,A_4\}|\{A_1\}$ | $R_{31}:A_1=2\wedge A_4=2\rightarrow d=1$ | 1 | 1 |
| | $R_{32}:A_1=2\wedge A_4=4\rightarrow d=2$ | 1 | 1 |

注1：在发现如表2和表3所示的决策规则时，仅需扫描决策表进行统计运算即可，无需进行等价类运算，但在叙述上，仍采用属性的等价类来描述。

**定理3** 若S为决策表DT的条件属性集的一个非空属性子集，则S-近似精度等于由S能完全确定的所有确定性决策规则的频率之和与card(U)之比。

证明：$E\in U/IND(S)$，且存在$Y_i\in\{Y_1,\cdots,Y_k\}$使得$E\subseteq Y_i$等价于决策规则$R:Des(E,S)\rightarrow Des(Y_i,D)$满足$R.conf=1$且$R.count=card(E)$。因而，S-近似精度等于由S能完全确定的所有确定性决策规则的频率之和与card(U)之比。证毕。

**定理4** 设C为决策表DT的条件属性集，则存在属性集序列$\{A_1\}\subset\{A_1,A_2\}\subset\cdots\subset\{A_1,A_2,\cdots,A_s\}\subset C$使得$R(\{A_1\}).count+R(\{A_1,A_2\}|\{A_1\}.count+\cdots+R(\{A_1,A_2,\cdots,A_s\}|\{A_1,A_2,\cdots,A_{s-1}\}).count=R(C).count$，其中，$\forall A\in C-\{A_1\}$有$R(\{A_1\}).count\geq R(\{A\}).count$；$\forall B\in C-\{A_1,A_2,\cdots,A_i\}$有$R(\{A_1,A_2,\cdots,A_i\}|\{A_1,A_2,\cdots,A_{i-1}\}).count\geq R(\{A_1,A_2,\cdots,B\}|\{A_1,A_2,\cdots,A_{i-1}\}).count$，$i=2,\cdots,s$；$s\leq card(C)$。

证明：不妨设由C的确定性规则所决定的对象集合组成的决策表为$DT'=(U',C\cup D,V,f)$，则$R(C).count=card(U')$。对决策表DT'，不难看出，若不存在$s<card(C)$的属性集序列$\{A_1\}\subset\{A_1,A_2\}\subset\cdots\subset\{A_1,A_2,\cdots,A_s\}\subset C$使得$R\{A_1\}.count+R(\{A_1,A_2\}|\{A_1\}).count+\cdots+R(\{A_1,A_2,\cdots,A_s\}|\{A_1,A_2,\cdots,A_{s-1}\}).count=card(U')$，其中，$\forall A\in C-\{A_1\}$有$R(\{A_1\}).count\geq R(\{A\}).count$；$\forall B\in C-\{A_1,A_2,\cdots,A_i\}$有$R(A_1,A_2,\cdots,A_i\}|\{A_1,A_2,\cdots,A_{i-1}\}).count\geq R(\{A_1,A_2,\cdots,B\}|\{A_1,A_2,\cdots,A_{i-1}\}).count$，$i=2,\cdots,s$，则当$s=card(C)$时，结论一定成立。证毕。

定理3提供了一种计算属性集的近似精度的手段，而由定理4可快速找到条件属性的属性子集，并保持近似精度不变。因而，由定理3和定理4为求决策表的属性约简提供了一种新的方法。

### 3.2 ARABR算法描述

定理4利用约束属性集的确定性规则的频度之和决定如何扩展候选属性，充分考虑了扩展属性和已选择属性集之间的相关度；同时，由定理3和定理4，可快速有效地得到决策表的属性的相对约简。于是，本文得到一种新的基于规则的属性约简算法(ARABR算法)，其基本步骤描述如下：

步骤1：令$Q=CORE(C)$，$P=Q$，$B=C\backslash P$；如果满足$\gamma_P=\gamma_C$，则停止，P为约简。

步骤2：对B中的每个属性$A_j$，$j=1,\cdots,card(B)$，计算$R(\{A_j\}\cup P|P).count$，计算$R(\{A_g\}\cup P|P).count=max(R(\{A_j\}\cup P|P).count:A_j\in B)$，置$P=P\cup\{A_g\}$，$B=B\backslash\{A_g\}$，转下一步继续执行。

步骤3：如果P满足$R(P).count=R(C).count$，转步骤4；否则，$R(P).count<R(C).count$，转步骤2。

步骤4：如果$P=CORE(P)$，则停止，P为一个属性约简(由文[9]改进的差别矩阵方法计算)；否则，转步骤5。

步骤5：任取$A_d\in P\backslash CORE(P)$，令$P=P-\{A_d\}$转步骤4。

由性质1可知，算法ARABR得到的P为一个属性约简。不难看出，ARABR算法中的步骤2求约束属性$\{A_j\}\cup P|P$的计算复杂度至多为$O(n^2)$，因而步骤2的总的计算复杂度至多为$O(mn^2)$，其中，$m=card(C)$，$n=card(U)$。而求核的计算复杂度为$O(mn^2)$，因而进入步骤4之前的计算复杂度至多为$O(m^2n^2)$。而步骤4的计算复杂度至多为$card(P)\cdot O(mn^2)=O(m^2n^2)$。故在最坏的情况下，ARABR算法的计算复杂度至多为$O(m^2n^2)$。

### 3.3 算法分析

在最坏的情况下，ARABR算法的计算复杂度至多为$O(m^2n^2)$，与算法A[11]一致，比Jelonek算法的计算复杂度($O(m^3n^2)$)降低了一个数量级。同时，ARABR算法还具有以下优点：①采用约束属性集的确定性规则的频度之和来度量候选属性与已选择属性集的相关度，使得与已选择属性集的相关度最大的属性得到扩展，克服了算法A的不足；②在进行属性扩展时，无需考虑已选择属性集的确定性决策规则所确定的对象集合，使得搜索空间显著减小；③由于合理地进行最佳属性的扩展，多数情况下能够直接得到属性约简，减少步骤5的计算次数。如：对实例1，由算法A得到$P=\{A_1,A_3,A_2\}$，还需多次进行步骤5运算，而由ARABR算法可直接得到属性约简$P=\{A_1,A_2\}$。理论分析表明，ARABR算法可使Jelonek算法的计算复杂度降低一个数量级，且克服了算法A存在的不足。

为了验证算法的有效性和执行效率，应用ARABR算法及算法A对UCI机器学习中的蘑菇数据库进行了计算。实验

结果表明，ARABR 算法的性能略优于算法 A，且多数情况下可获得最小属性约简，这与理论分析结论是一致的，因而 ARABR 算法是有效可行的。

**结束语** 本文在引入相容决策规则概念之后，提出了基于规则的属性约简算法 ARABR，算法 ARABR 不仅为决策表的最小属性约简提供了一种新的框架，而且还提供了一种度量候选属性与已选择属性集相关度的新方法，使相关度最大的候选属性得以扩展，并使搜索空间显著减小，克服了算法 A 单纯利用单属性的近似精度来扩展候选属性所存在的不足。理论分析表明，算法 ARABR 比 Jelonek 算法具有更低阶的计算复杂性，且可克服算法 A 存在的不足，因而是有效可行的。如何对属性约简集进行增量式维护为下一步的研究目标。

## 参 考 文 献


1 Pawlak Z. Rough sets. International Journal of Information and Computer Science, 1982, 11(5): 341～356

2 Pawlak Z. Rough Set-Theoretical Aspcts of Reasoning about Data. Dordrecht, Kluwer Academic Publishers, 1991

3 杨明，孙志挥，季小俊. 基于 Rough Set 的缺省加权规则挖掘算法. 东南大学学报，2002，32(1)：115～118

4 Pawlak Z, Slowinski R. Rough set approach to multiattribute decision analysis. Ivited Review. European Journal of Operational Research, 1994, 72: 443～459

5 Skowron A, Rauszer C. The discernibility matices and functions in information systems. In: Slowinski R, ed. Intelligent Decision Support—Handbook of Applications and Advances of the Rough Sets Theory. Dordrecht, Kluwer: Academic Publishers, 1992. 331～362

6 王珏，王任，苗夺谦，等. 基于 Rough Set 理论的"数据浓缩". 计算机学报，1998，21(5)：393～400

7 Wong S K M, Ziarko W. On optimal decision rules in decision tables. Bulletin of Polish Academy of Sciences, 1985, 33: 693～696

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

9 Hu X H, Cercone N. Learning in relational databases: a rough set approach. Computational Intelligence, 1995, 11(2): 323～338

10 Jelonek J, et al. Rough Set reduction of attributes and their domains for neural networks. Computational Intelligence, 1995, 11(2): 339～347

11 叶东毅. Jelonek 属性约简算法的一个改进. 电子学报，2000，12(12)：81～82


(上接第87页)

的知识结构，不同信息来源之间也可以建立联系，进行信息交换。网络存取的基本形态不再是链状的，可以分布表达和传递信息。由于信息的分布更为合理，并能根据网络的情况和用户的请求状况动态地调整，因此可以抵抗许多突发事件。大量实验结果证明了本模型的优越性。

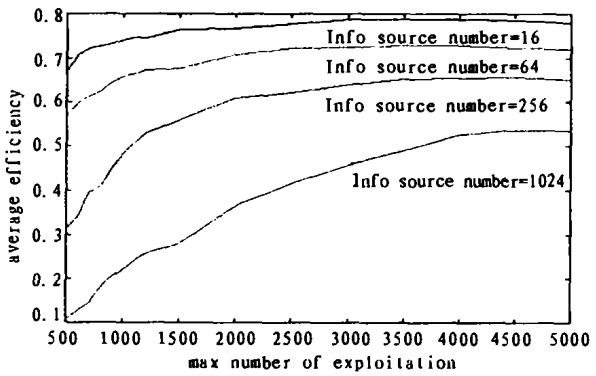


图6 网络信息自组织利用模式的平均响应效率曲线
(GCA 细胞数量10000；信息内容的类别数量100)

## 参 考 文 献


1 Kohavi R, Masand B, Spiliopoulou M. Web mining. Data Mining and Knowledge Discovery, 2002, 6 (1): 5～8

2 Tan P, Kumar V. Discovery of web robot sessions based on their navigational patterns. Data Mining and Knowledge Discovery, 2002, 6 (1): 9～35

3 Ferber, Jacques. Multi-agent systems: towards a collective intelligence. Reading, MA: Addison-Wesley, 1998

4 Benjaafar, Saifallah, et al. Cellular automata for traffic flow modeling, Univ. of Minnesota, Minneapolis, 1997

5 Shuai D X. A new parallel-by-cell approach to undistorted data compression based on cellular automaton and genetic algorithm. Journal of Computer Science and Technology, 1999, 14(6): 572～579

6 Shuai D X, Gu J. The faster high-order cellular automaton for hyper-parallel undistorted data compression. Journal of Computer Science and Technology, 2000, 15, (2): 126～135

7 Feldmann A, Greenberg A, et al. NetScope: traffic engineering for IP networks. IEEE Network, March/April, 2000. 11～19

8 Casetti C, Meo M. A new approach to model the stationary behavior of TCP connections . IEEE Infocom, 2000. 367～375

9 Liou C, Tai W. Conformality in the self-organization network. Artificial Intelligence, 2000, 116: 265～286

10 Liu C, Peek J, Jones R, et al. Managing Internet Information Service. O'Reilly Associates, Inc. 1994

11 Kumar G P, Venkataram P. Artificial Intelligence approaches to network management: recent advance and a survey. Computer Communication, 1997, 20(15): 1313～1322

12 Bernardo A, Rajan M. Social Dilemmas and Internet Congestion. Science, 1997, 277: 535～537

13 Chua L O, Roska T. The CNN paradigm. IEEE Trans. on Circuits and Systems, 1993, 40(3): 147～156

14 Trinh H, Aldeen M. Distributed state observer scheme for large-scale interconnected systems, IEE Pro. —Control Theory Appl., 1998, 145(3): 331～337