第 42 卷 第 8 期
Vol.42 No.8

山 东 大 学 学 报 (理 学 版)
JOURNAL OF SHANDONG UNIVERSITY

2007 年 8 月
Aug. 2007

文章编号:1671-9352(2007)08-0046-05

单向 S-粗集与数据筛选 – 过滤

蔡成闻[1],赵俊恺[1],史开泉[2]

(1. 山东大学 控制科学与工程学院,山东 济南 250061;
2. 山东大学 数学与系统科学学院,山东 济南 250100)

摘要:利用单向 S-粗集,给出了数据属性,数据筛选 – 过滤概念,提出了数据筛选 – 过滤序定理,合成数据筛选 – 过滤定理,数据筛选 – 过滤准则,利用这些结果,给出其应用.

关键词:单向 S-粗集;数据筛选 – 过滤;筛选 – 过滤度;筛选 – 过滤定理

中图分类号:O159;TP18 **文献标识码**:A

One direction S-rough sets and data sieve-filtration

CAI Cheng-wen[1], ZHAO Jun-kai[1] and SHI Kai-quan[2]

(1. School of Control Science and Engineering, Shandong University, Jinan 250061, Shandong, China;
2. School of Mathematics and System Science, Shandong University, Jinan 250100, Shandong, China)

Abstract: By using the one direction S-rough sets, the data attribute and the concept of data sieve-filtration are given. Furthermore, the data sieve-filtration order theorem, composite- data sieve-filtration theorem and data sieve-filtration rule are presented. An application is given based on the results.

Key words: one direction S-rough sets; data sieve-filtration; sieve-filtration degree; data sieve-filtration theorem

0 引言

在经济,通讯,投资等诸多系统中,人们分析处理着大量的数据,从这些数据中寻找所需要的数据(例如,通讯故障数据,投资风险数据,等).给定数据 $w=\{w_1,w_2,\cdots w_n\}$,w 具有属性集(特征集)$\alpha=\{\alpha_1,\alpha_2,\cdots \alpha_\lambda\}$;换句话说,因为 $w_1,w_2,\cdots w_n$ 都具有属性集 α,则 $w_1,w_2,\cdots w_n$ 构成 w;用数学的语言表达,则 $w_1,w_2,\cdots w_n$ 构成关于 α 的等价类 $[w]=\{w_1,w_2,\cdots w_n\}$.事实上,$[w]$ 能够分成多个子类 $[w]_i\subset[w]$,$[w]_i$ 具有属性集 α_i,这是把数据 w,给出等价类的认识.一个事实:若子类 $[w]_i=\{w_{i,1},w_{i,2},\cdots w_{i,p}\}$ 具有属性集 $\alpha_i=\{\alpha_{i,1},\alpha_{i,2},\cdots\alpha_{i,t}\}$,对 α_i 进行属性补充,α_i 变成 $\alpha'_i=\{\alpha_{i,1},\alpha_{i,2},\cdots\alpha_{i,t},\alpha_{i,t+1},\cdots\alpha_{i,t+k}\}$,则子类 $[w]_i$ 变成 $[w]'_i=\{w_{i,1},w_{i,2},\cdots w_{i,q}\}$,$q<p$,或者 $[w]'_i\subseteq[w]_i$.这个事实告诉我们:子数据 $[w]'_i$ 潜藏在 $[w]_i$ 中,$[w]'_i$ 从 $[w]_i$ 中被挖据出来.文献[1~5]改进了 Z.Pawlak 粗集,提出 S-粗集,S-粗集具有三类形式:单向 S-粗集,单向 S-粗集对偶与双向 S-粗集;S-粗集具有动态特性(单向动态特性,双向动态特性),S-粗集的动态特性与本文给出的数据筛选 – 过滤相吻合,而且它们之间存在着密切的联系;利用单向 S-粗集这一新的数学工具,本文给出数据筛选 – 过滤的讨论,给出应用.

为了便于讨论,又能容易接收本文给出的结果,把单向 S-粗集与它的结构[1~5]引入到这里:

收稿日期:2007-02-09
基金项目:山东省自然科学基金资助项目(Y2004A04);福建省自然科学基金资助项目(200650391)
作者简介:蔡成闻(1980-),男,硕士,研究方向:系统识别.E-mail: miccw101@163.com

给定元素论域 U，F 是定义在 U 上的元素迁移族，$F=\{f_1,f_2,\cdots f_n\}$，称 $X^{\circ}\subset U$ 是 U 上的一个单向 S-集合，而且

$$X^{\circ}=X\cup\{u\mid u\in U,u\bar{\in}X,f(u)=x\in X\};$$

称 X^f 是 $X\subset U$ 的 f-扩张，而且

$$X^f=\{u\mid u\in U,u\bar{\in}X,f(u)=x\in X\}.$$

其中，X 是 Z. Pawlak 粗集（$R_-(X),R^-(X)$）中的集合，$X\subset U$.

设 X° 是 U 上的单向 S-集合，$X^{\circ}\subset U$，称 $(R,F)_{\circ}(X^{\circ})$ 是 $X^{\circ}\subset U$ 的下近似，如果

$$(R,F)_{\circ}(X^{\circ})=\bigcup[x]=\{x\mid x\in U,[x]\subseteq X^{\circ}\};$$

称 $(R,F)^{\circ}(X^{\circ})$ 是 $X^{\circ}\subset U$ 的上近似，如果

$$(R,F)^{\circ}(X^{\circ})=\bigcup[x]=\{x\mid x\in U,[x]\cap X^{\circ}\neq\phi\}.$$

其中 $F\neq\phi$.

由 $(R,F)_{\circ}(X^{\circ})$，$(R,F)^{\circ}(X^{\circ})$ 构成的集合对，称作 $X^{\circ}\subset U$ 的单向 S-粗集，而且

$$((R,F)_{\circ}(X^{\circ}),(R,F)^{\circ}(X^{\circ}));$$

称 $\mathrm{Bn}_R(X^{\circ})$ 是 $X^{\circ}\subset U$ 的 R-边界，而且

$$\mathrm{Bn}_R(X^{\circ})=(R,F)_{\circ}(X^{\circ})-(R,F)^{\circ}(X^{\circ});$$

称 $\mathrm{As}(X^{\circ})$ 是单向 S-粗集生成的副集，而且

$$\mathrm{As}(X^{\circ})=\{x\mid u\in U,u\bar{\in}X,f(u)=x\,\widetilde{\in}\,X\}.$$

这里："$\widetilde{\in}$"是一个特别的记号，它的意义见[1～5].

从单向 S-粗集[1~5]特性中，能够直接得到下面的命题：

命题 1 给定 R-元素等价类 $[x]$，α 是 $[x]$ 的属性集，若 $\exists$ 属性 $\beta\in V$，$B\bar{\in}\alpha$，$f(\beta)=\alpha'\in\alpha$；具有属性集 $\alpha^f=\alpha\cup\{f(\beta)=\alpha'\}$ 的 R-元素等价类 $[x]^f$ 满足

$$[x]^f\subseteq[x].$$

命题 2 若 $[x]_-=(R,F)_{\circ}(X^{\circ})=\bigcup[x]$，$[x]_F=(R,F)_{\circ}(X^{\circ})_F=\bigcup[x]^f$，则

$$[x]_F\subseteq[x]_-;$$

若 $[x]^-=(R,F)^{\circ}(X^{\circ})=\bigcup[x]$，$[x]^F=(R,F)^{\circ}(X^{\circ})^F=\bigcup[x]^f$，则

$$[x]^F\subseteq[x]^-.$$

1 数据的属性特征

设 W 是一个数据，W 具有 m 个子数据 $w_1,w_2,\cdots w_m$；子数据 w_i 由若干个元素构成，或者 $w_1=\{x_{1,1},x_{1,2},\cdots x_{1,t}\}$，$w_2=\{x_{2,1},x_{2,2},\cdots x_{2,s}\}$，$\cdots w_m=\{x_{m,1},x_{m,2},\cdots x_{m,r}\}$；$\alpha_1,\alpha_2,\cdots\alpha_m$ 分别是子数据 $w_1,w_2,\cdots w_m$ 的属性集，而且 $\alpha_1=\{\alpha_{1,1},\alpha_{1,2},\cdots\alpha_{1,p}\}$，$\alpha_2=\{\alpha_{2,1},\alpha_{2,2},\cdots\alpha_{2,q}\}$，$\cdots\alpha_m=\{\alpha_{m,1},\alpha_{m,2},\cdots\alpha_{m,k}\}$；$\alpha_1\neq\alpha_2\neq\cdots\neq\alpha_m$. 显然，$w_1=\{x_{1,1},x_{1,2},\cdots x_{1,t}\}$，$w_2=\{x_{2,1},x_{2,2},\cdots x_{2,s}\}$，$\cdots w_m=\{x_{m,1},x_{m,2},\cdots x_{m,r}\}$ 是分别关于属性集 $\alpha_1,\alpha_2,\cdots\alpha_m$ 的等价类 $[w]_1,[w]_2,\cdots[w]_m$. 利用单向 S-粗集，容易在 w_j 中寻找到 $w'_j\subseteq w_j$，如果存在属性集 α'_j 满足 $\alpha_j\subseteq\alpha'_j$，因此，得到命题 1～4.

命题 1 子数据 w_j,w'_j，而且 $w'_j\subseteq w_j$；w_j,w'_j 的属性集 α_j,α'_j 满足

$$\mathrm{card}(\alpha_j)\leq\mathrm{card}(\alpha'_j).\tag{1.1}$$

命题 2 w_-,w_F 分别是子数据 w_i 合成构成的数据，$w_F\subseteq w_-$，w_- 的属性集 α_- 与 w_F 的属性集 α_F 满足

$$\mathrm{card}(\alpha_-)\leq\mathrm{card}(\alpha_F).\tag{1.2}$$

这里：$w_-=\bigcup[w]_i$，$w_F=\bigcup[w]_{F,i}$，$[w]_{F,i}\subseteq[w]_i$.

命题 3 w^-,w^F 分别是子数据 w_j 合成构成的数据，$w^F\subseteq w^-$，w^- 的属性集 α^- 与 w^F 的属性集 α^F 满足

$$\mathrm{card}(\alpha^-)\le \mathrm{card}(\alpha^F).\tag{1.3}$$

这里:$w^-=\bigcup[w]_j$,$w^F=\bigcup[w]_j^F$,$[w]_j^F\subseteq[w]_j$.

命题 1,2,3 的证明,由单向 S-粗集的动态特性得到,证明略.

为了便于理解,给出一个实际的例子,因特网已成为人们日常工作,生活中不可或缺的一部分,有很多人通过因特网获取有用的信息.事实上,网页[6,7]可以看作是一个数据 $w=\{w_1,w_2,\cdots w_n\}$,网址可以看作是数据 w 对应的属性集 α,换句话说,因为子数据 $w_1,w_2,\cdots w_n$ 都具有属性集 α,所以 $w_1,w_2,\cdots w_n$ 共同构成了网页数据 w;用数学的语言表达,$w_1,w_2,\cdots w_n$ 构成了关于 α 的等价类$[w]$.从另外一个角度讲,查看网页就是通过属性集 α 发现等价类$[w]$的过程.有这样一个事实,$[w]$能够分成多个子类$[w]_i\subset[w]$,$[w]_i$ 具有属性集 α_i,因此得到这样一个结论:若子类$[w]_i=\{w_{i,1},w_{i,2},\cdots w_{i,p}\}$具有属性集 $\alpha_i=\{\alpha_{i,1},\alpha_{i,2},\cdots\alpha_{i,\lambda}\}$,对属性集 α_i 进行属性扩充,α_i 变成 $\alpha_i^f=\{\alpha_{i,1},\alpha_{i,2},\cdots\alpha_{i,\lambda},\alpha_{i,\lambda+1},\cdots\alpha_{i,\lambda+k}\}$,则子类$[w]_i$ 变成$[w]_i^f=\{w_{i,1},w_{i,2},\cdots w_{i,q}\}$,$q<p$;换一个角度考虑,通过对网页属性集的扩充,可以实现从现有网页到一个内容有所收缩网页的转变.这个事实告诉我们:子数据$[w]_j^f$ 隐藏在$[w]_i$ 中,通过对属性集 α_i 的扩充,使$[w]_j^f$ 从$[w]_i$ 中被筛选 - 过滤出来.

由前面的讨论可知:如果属性集 α 中的属性增加,子数据 w 中的数据个数将减少,即子数据收缩;如果属性集 α 中的属性减少,子数据 w 中的数据个数将增加,即子数据扩张.显然,属性集 α 的变化将引起子数据 w 的变化.因此,子数据 w 是具有颗粒特征的.

2　数据的筛选 - 过滤

约定:$W=\{w_1,w_2,\cdots w_m\}$是数据,$\alpha_1,\alpha_2,\cdots\alpha_m$ 分别是子数据 $w_1,w_2,\cdots w_m$ 的属性集.

定义 2.1　称 $\mathrm{GRD}(w_i)$是子数据 $w_i\in W$ 的粒度,如果

$$\mathrm{GRD}(w_i)=\mathrm{card}(w_i)/\mathrm{card}(W).\tag{2.1}$$

其中,$\mathrm{card}(w_i)$是子数据 w_i 的基数.

定义 2.2　设 $\mathrm{GRD}(w_i)$是子数据 $w_i\in W$ 的粒度,$\mathrm{SFD}(w_i)$称作子数据 w_i 关于 W 的筛选 - 过滤度,如果

$$\mathrm{SFD}(w_i)=\mathrm{SUR}(w_i)/\mathrm{card}(W)=1-\mathrm{GRD}(w_i);\tag{2.2}$$

称 $\mathrm{SUR}(w_i)$是子数据 w_i 关于 W 的筛选 - 过滤剩余,如果

$$\mathrm{SUR}(w_i)=\mathrm{card}(W)-\mathrm{card}(w_i).\tag{2.3}$$

定义 2.3　f-数据 w^f 称作 w 的 f-筛选 - 过滤,如果 w^f 具有关于 w 的 α^f-筛选 - 过滤度,而且

$$\mathrm{SFD}(w^f)=\mathrm{SUR}(w^f)/\mathrm{card}(w)=1-\mathrm{GRD}(w^f).\tag{2.4}$$

其中,$\mathrm{SUR}(w^f)=\mathrm{card}(w)-\mathrm{card}(w^f)$,$\mathrm{GRD}(w^f)$是 w^f 关于 w 的粒度.

定义 2.4　F-下合成数据 w_F 称作 w_- 的 α_F-筛选 - 过滤,如果 w^F 关于 w_- 具有 α^F-筛选 - 过滤度,而且

$$\mathrm{SFD}(w_F)=\mathrm{SUR}(w_F)/\mathrm{card}(w_-)=1-\mathrm{GRD}(w_F).\tag{2.5}$$

定义 2.5　F-上合成数据 w^F 称作 w^- 的 α^F-筛选 - 过滤,如果 w^F 关于 w^- 具有 α^F-筛选 - 过滤度,而且

$$\mathrm{SFD}(w^F)=\mathrm{SUR}(w^F)/\mathrm{card}(w^-)=1-\mathrm{GRD}(w^F).\tag{2.6}$$

其中,$\mathrm{SUR}(w_F)=\mathrm{card}(w_-)-\mathrm{card}(w_F)$,$\mathrm{SUR}(w^F)=\mathrm{card}(w^-)-\mathrm{card}(w^F)$,$\mathrm{GRD}(w_F)$,$\mathrm{GRD}(w^F)$分别是关于 w_-,w^- 的粒度.

由定义 2.1 得到下面的事实:

若 w° 是最小子数据,而且 $\mathrm{card}(w^\circ)=1$,则

$$\mathrm{GRD}(w^\circ)=1/\mathrm{card}(W);$$

若 w^* 是最大子数据,则

$$\mathrm{GRD}(w^*)=1;$$

若 w_i 是子数据,而且 $w^\circ\subseteq w_i\subseteq w^*$,则 w_i 满足

$$\mathrm{GRD}(w^\circ)\le\mathrm{GRD}(w_i)\le\mathrm{GRD}(w^*).$$

由定义2.2得到下面的事实:

若 w° 是最小子数据,而且 $\mathrm{card}(w^{\circ})=1$,则

$$\mathrm{SFD}(w^{\circ})=1-1/\mathrm{card}(W);$$

若 w^{*} 是最大子数据,则

$$\mathrm{SFD}(w^{*})=0;$$

若 w_i 子数据,而且 $w^{\circ}\subseteq w_i\subseteq w^{*}$,则 w_i 满足

$$\mathrm{SFD}(w^{*})\leq \mathrm{SFD}(w_i)\leq \mathrm{SFD}(w^{\circ}).$$

定理 2.1(数据筛选-过滤粒度序定理) 设 $\alpha_1,\alpha_2,\alpha_3$ 分别是子数据 w_1,w_2,w_3 的属性集,若存在 $\mathrm{card}(\alpha_3)\leq \mathrm{card}(\alpha_2)\leq \mathrm{card}(\alpha_1)$,则 w_1,w_2,w_3 的粒度满足

$$\mathrm{GRD}(w_1)\leq \mathrm{GRD}(w_2)\leq \mathrm{GRD}(w_3). \tag{2.7}$$

证明 因为 $\mathrm{card}(\alpha_3)\leq \mathrm{card}(\alpha_2)\leq \mathrm{card}(\alpha_1)$,由命题1,得到:$w_1\subseteq w_2\subseteq w_3$;或者 $\mathrm{card}(w_1)\leq \mathrm{card}(w_2)\leq \mathrm{card}(w_3)$,因此 $\mathrm{GRD}(w_1)\leq \mathrm{GRD}(w_2)\leq \mathrm{GRD}(w_3)$.

推论 1 若 α 是 W 的属性集,对于任意一个属性集 α',而且满足 $\mathrm{card}(\alpha)\leq \mathrm{card}(\alpha')$,$\alpha'$ 对应子数据 w' 的粒度满足

$$\mathrm{GRD}(w')\leq \mathrm{GRD}(W). \tag{2.8}$$

推论 2 若 α 是 W 的属性集,对于任意一个属性集 α'',而且满足 $\mathrm{card}(\alpha)\geq \mathrm{card}(\alpha'')$,$\alpha''$ 对应子数据 w'' 的粒度满足

$$\mathrm{GRD}(W)\leq \mathrm{GRD}(w''). \tag{2.9}$$

定理 2.2(数据筛选-过滤度序定理) 设 $\alpha_1,\alpha_2,\alpha_3$ 分别是子数据 w_1,w_2,w_3 的属性集,若存在 $\mathrm{card}(\alpha_3)\leq \mathrm{card}(\alpha_2)\leq \mathrm{card}(\alpha_1)$,则 w_1,w_2,w_3 的筛选-过滤度满足

$$\mathrm{SFD}(w_3)\leq \mathrm{SFD}(w_2)\leq \mathrm{SFD}(w_1). \tag{2.10}$$

证明 与定理2.1类似,证明略.

推论 3 若 α 是 W 的属性集,对于任意一个属性集 α',而且满足 $\mathrm{card}(\alpha)\leq \mathrm{card}(\alpha')$,$\alpha'$ 对应子数据 w' 的筛选-过滤度满足

$$\mathrm{SFD}(w')\geq \mathrm{SFD}(W). \tag{2.11}$$

推论 4 若 α 是 W 的属性集,对于任意一个属性集 α'',而且满足 $\mathrm{card}(\alpha)\geq \mathrm{card}(\alpha'')$,$\alpha''$ 对应子数据 w'' 的筛选-过滤度满足

$$\mathrm{SFD}(w'')\leq \mathrm{SFD}(W). \tag{2.12}$$

定理 2.3(f-数据筛选-过滤定理) 给定数据 w_i^f,w_j^f,w_k^f 关于 w 的 α^f-筛选-过滤度 $\mathrm{SFD}(w_i^f)$,$\mathrm{SFD}(w_j^f)$,$\mathrm{SFD}(w_k^f)$,如果

$$\mathrm{SFD}(w_k^f)\leq \mathrm{SFD}(w_j^f)\leq \mathrm{SFD}(w_i^f) \tag{2.13}$$

则 w 依 $\mathrm{SFD}(w_p^f)$,$p=1,2,3$ 的顺序被筛选-过滤为 w_i^f,w_j^f,w_k^f,而且

$$w_i^f\subseteq w_j^f\subseteq w_k^f. \tag{2.14}$$

证明 若 w_i^f,w_j^f,w_k^f 关于 w 的 α^f-筛选-过滤度满足 $\mathrm{SFD}(w_k^f)\leq \mathrm{SFD}(w_j^f)\leq \mathrm{SFD}(w_i^f)$,则 $\mathrm{GRD}(w_i^f)\leq \mathrm{GRD}(w_j^f)\leq \mathrm{GRD}(w_k^f)$,得到:$\mathrm{card}(w_i^f)\leq \mathrm{card}(w_j^f)\leq \mathrm{card}(w_k^f)$,因此,$w_i^f\subseteq w_j^f\subseteq w_k^f$.

定理 2.4(F-下合成数据筛选-过滤定理) 给定 F-下合成数据 $w_{F,i},w_{F,j},w_{F,k}$ 关于 w_- 的 α_F-筛选-过滤度 $\mathrm{SFD}(w_{F,i})$,$\mathrm{SFD}(w_{F,j})$,$\mathrm{SFD}(w_{F,k})$,如果

$$\mathrm{SFD}(w_{F,k})\leq \mathrm{SFD}(w_{F,j})\leq \mathrm{SFD}(w_{F,i}), \tag{2.15}$$

则 w_- 依 $\mathrm{SFD}(w_{F,q})$,$q=1,2,3$ 的顺序被筛选-过滤为 $w_{F,i},w_{F,j},w_{F,k}$,而且

$$w_{F,i}\subseteq w_{F,j}\subseteq w_{F,k}. \tag{2.16}$$

定理 2.5(F-上合成数据筛选-过滤定理) 给定 F-上合成数据 w_i^F,w_j^F,w_k^F 关于 w^- 的 α^F-筛选-过滤度 $\mathrm{SFD}(w_i^F)$,$\mathrm{SFD}(w_j^F)$,$\mathrm{SFD}(w_k^F)$,如果

$$\mathrm{SFD}(w_k^F)\leq \mathrm{SFD}(w_j^F)\leq \mathrm{SFD}(w_i^F), \tag{2.17}$$

则 w_- 依 $\mathrm{SFD}(w_t^F)$, $t=1,2,3$ 的顺序被筛选－过滤为 w_i^F, w_j^F, w_k^F,而且

$$w_i^F \subseteq w_j^F \subseteq w_k^F. \tag{2.18}$$

定理 2.4～2.5 的证明与定理 2.3 类似,证明略.

数据的单向 S-粗集筛选－过滤粒度准则

在包含 m 个子数据 w_p 的 (W,α) 的数据系统中,数据 $w_i \in W$ 的粒度 $\mathrm{GRD}(w_i)$ 满足

$$\mathrm{GRD}(w_1) \leq \mathrm{GRD}(w_2) \leq \cdots \leq \mathrm{GRD}(w_m);$$

被最先筛选－过滤的数据 w_k 满足

$$\mathrm{GRD}(w_k) = \min_{i=1}^{m}(\mathrm{GRD}(w_i)).$$

数据的单向 S-粗集筛选－过滤属性准则

在包含 m 个子数据 w_q 的 (W,α) 的数据系统中,被最先筛选－过滤的数据 $w_k \in W$ 的属性集满足

$$\alpha_k = \max_{j=1}^{m}(\alpha_1, \alpha_2, \cdots \alpha_m).$$

3　数据筛选－过滤的一个应用

下面是一个网站数据库[6,7]的例子.这是子数据库筛选关于动态数据库的一个具体应用.

在一个网络服务器数据库中保存有大量的信息,在用户浏览时需要快速有效的从数据库中筛选出用户最需要的信息.在这里,本文将数据库看作一类特殊的数据,同时,数据库的大小限定在一个较小的范围内,以便于说明,数据库的大小不影响推导结果.

设 $W=\{w_1,w_2,w_3,w_4,w_5,w_6,w_7,w_8,w_9,w_{10},w_{11},w_{12}\}$ 是数据库中的一个子数据库,具有属性集 $\alpha=\{\alpha_1,\alpha_2\}$, α_1 代表“国内网站”, α_2 代表“网站名称”.对于属性集 $\alpha=\{\alpha_1,\alpha_2\}$, $w_1,w_2,w_3,w_4,w_5,w_6,w_7,w_8,w_9,w_{10},w_{11},w_{12}$ 是不可分辨的,记作 $\mathrm{IND}(W)$.在一个网站中往往包含很多个栏目,这里定义栏目的属性包括 $\beta_1,\beta_2,\beta_3,\beta_4$,其中 β_1 代表“国内”, β_2 代表“国际”, β_3 代表“军事”, β_4 代表“体育”.根据栏目属性 $\beta_1,\beta_2,\beta_3,\beta_4$ 得到 W 中元素的状态表,如表 1.表 1 中的符号“*”表示元素具有属性 A, $A\in\{\beta_1,\beta_2,\beta_3,\beta_4\}$.

表 1　W 上的元素的状态表

Table 1　State table of element in W

栏目属性	W 中的元素											
	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
β_1	*			*	*		*		*		*	
β_2		*	*			*		*		*		*
β_3	*						*	*			*	
β_4		*		*					*	*		*

当用户点击某一栏目,计算机在后台向数据库发出指令,调用这一栏目的网页,例如当用户点击“国内”这一栏目时,属性集 α 变成 $\alpha'=\alpha\cup\{\beta_1\}=\{\alpha_1,\alpha_2,\beta_1\}$,子数据库 W 收缩为 $W_1=\{w_1,w_4,w_5,w_7,w_9,w_{11}\}$,这样,用户可以更快,更方便地访问子数据库 W_1.

因为在属性集 α 中填充属性 β_1 得到 α',而且 $\alpha\subseteq\alpha'$,由定理 2.2 得到

$$\mathrm{SFD}(W_1) \geq \mathrm{SFD}(W). \tag{3.1}$$

式(3.1)表明: W_1 的筛选－过滤度大于 W 的筛选－过滤度, W_1 从 W 中被筛选－过滤出来.利用筛选－过滤,子数据 W_1 从子数据 W 中被发现.

同样能得到关于其他的栏目子数据库 W_2, W_3, W_4 的讨论,这些讨论略.

4　讨论

本文依赖单向 S-粗集理论,给出数据筛选－过滤研究,给出数据筛选-过滤方法,准则与(下转第 54 页)

(上接第 50 页) 应用.单向 S-粗集是在数据库中挖掘未知数据的一个新的,重要的数学工具;特别是对具有动态特性的未知数据的寻找,单向 S-粗集给予了直接的理论支持.单向 S-粗集,给人们寻找具有动态特性的未知数据提供了新的思路.

参考文献:

[1] SHI Kaiquan. S-rough sets and its applications in diagnosis-recognition for disease[J]. IEEE Proceedings of the First International Conference on Machine Learning and Cybernetics, 2002, (1):50-54.

[2] SHI Kaiquan, CUI Yuquan. F-decomposition and $\bar{F}$-reduction of S-rough sets[J]. An International Journal Advances in System Science and Applications, 2004, (4):487-499.

[3] SHI Kaiquan, CHANG Tingcheng. One direction S-rough sets[J]. International Journal of Fuzzy Mathematics, 2005, (2):239-334.

[4] SHI Kaiquan. Two direction S-rough sets[J]. International Journal of Fuzzy Mathematics, 2005, (2):335-349.

[5] 史开泉, 崔玉泉. S-粗集与粗决策[M]. 北京: 科学出版社, 2006.101-126.

[6] 张永仁, 黄科军, 李德孝. 基与数据库的文件管理[J]. 计算机工程与设计, 2006, (6):2 044-2 045.

[7] 曹泰钧. 粗集理论在数据处理中的研究与应用[J]. 河北理工学院学报, 2003, (4):67-72.

(编辑:李晓红)