

基于双向 S-粗集的动态粗传递模型^{*})

刘纪芹

(山东财政学院统计与数理学院 济南 250014)

摘要 根据信息传递中我们需要准确可靠的信息这种实际情况,利用双向 S-粗集理论提出了一种动态粗传递模型,即双向 S-下近似动态粗传递模型。分析了该模型的特性,得到了信息保持不变、信息发生损失与知识的关系,给出了提高信息传递精确性的方法以及该模型的应用。

关键词 双向 S-粗集,动态粗传递,知识,信息损失,信息不变

A Dynamic Rough Communication Model Based on Two-direction S-rough Sets

LIU Ji-Qin

(Department of Statistics and Mathematics, Shandong Finance Institute, Jinan 250014)

Abstract According to the practical situation that more accurate information is required in the process of information communication, using two-direction S-rough sets theory, a new dynamic rough communication model is proposed, i. e. two-direction S-lower approximation dynamic rough communication model. Its properties are discussed, and the relationships between information invariability and knowledge, between information loss and knowledge are obtained. A method of improving dynamic rough communication accuracy is presented. In the end, an application of the model is given.

Keywords Two-direction S-rough sets, Dynamic rough communication, Knowledge, Information loss, Information invariability

人类认识、推理和做决策都是在大量的信息中进行的,人类所具有的分析问题和解决问题的知识是有粒度的^[1]。现实世界中存在很多事物,我们对其不具备充分的知识,即我们对其认识是不完全和不精确的,而且也很难做到对它们的认识是完全、精确的。因此,具有不同知识的人对同一事物会有不同的认识和理解。这时,在他们之间进行信息传递时,使信息传递具有粗糙性,这种传递称为粗传递^[2]。

粗糙集理论^[3,4]被认为是处理不精确、不确定、模糊性知识的新的数学工具。在粗糙集理论中,知识可由论域中的等价关系簇所表示,知识库中的知识体现为对论域中对象进行分类的能力,而知识的不确定性则体现为论域中对象的不可分辨性,即知识的粒度^[5]。因此,文[2]利用粗糙集来研究粗传递,这是粗糙集的一个新的应用。

在实际中进行信息传递时,常常要求传递过来的信息必须准确可靠。文[2]只是对这种情况下的静态粗传递进行了初步讨论,但没有给出这种静态粗传递的特性,也没有解决具有动态知识的个体对于动态信息的粗传递问题。而在实际中进行的信息传递一般都是动态的,这是因为:一方面,随着人们认识的深入,人们不断吸收新知识,去掉陈旧无用的知识;另一方面,信息源也在发展变化,信息源也在不断吸收一些有意义的信息,去掉一些无意义的信息。这种知识在变化,信息源也在变化的信息传递,称为动态粗传递。

针对具有动态特性的集合 $X \subset U$, U 为论域,文[6~8]提出双向 S-粗集,并对 S-粗集的诸多特性和应用给出了进一步的讨论, S-粗集对动态数据挖掘与动态知识发现提供了一种有用的数学工具。因此对上述情况下的动态粗传递,提出如下问题:双向 S-粗集能用来描述和研究这种动态粗传递吗?如果能,这种动态粗传递具有什么特性?

本文利用双向 S-粗集解决了上述问题,建立了一种动态

粗传递模型,即双向 S-下近似动态粗传递模型,分析了该模型的特性,给出了该模型的应用。通过分析可知,静态粗传递^[2]是动态粗传递的特殊情况,动态粗传递是双向 S-粗集的一种新的应用。

为了本文讨论的方便,把双向 S-粗集引入到本文中。

1 双向 S-粗集^[6~8]与双向 S-动态认识

设论域 U 为有限集, R 是 U 上的一族等价关系, $K = (U, R)$ 为一知识库, $[x]_R$ 为 U 上的 R -等价类。 $ind(R)$ ^[9,10] 为 R 上的不可分辨关系,且 $[x]_{ind(R)} = \bigcap_{R \in R} [x]_R$, $[x]_{ind(R)}$ 简记为 $[x]_R$, $U/ind(R)$ 为等价关系 $ind(R)$ 的所有等价类。

定义 1.^[6~8] 给定集合 $X \subset U$, U 为论域,称 $F = \{f_1, f_2, \dots, f_m\}$, $\bar{F} = \{\bar{f}_1, \bar{f}_2, \dots, \bar{f}_n\}$ 是定义在 U 上的元素迁移族,如果 $f_i \in F, \bar{f}_j \in \bar{F}$ 分别满足:

$$\exists u \in U, u \notin X \Rightarrow f_i(u) = z \in X, \quad (1.1)$$

$$\exists x \in X \Rightarrow \bar{f}_j(x) = y \notin X \quad (1.2)$$

$f_i \in F, \bar{f}_j \in \bar{F} \quad (i=1, 2, \dots, m; j=1, 2, \dots, n)$ 称作元素迁移。

(1.1)式直观意义:把 X 看作知识库中信息元素的集合,不在 X 中的信息元素 u ,在 $f_i \in F$ 的作用下, u 变为 $f_i(u) = z \in X$, $f_i(u)$ 进入 X 内,使 X 中的信息增加。

(1.2)式直观意义:在 X 中的信息元素 x ,在 $\bar{f}_j \in \bar{F}$ 的作用下,变为 $\bar{f}_j(x) = y$, $\bar{f}_j(x)$ 离开集合 X ,即在 $\bar{f}_j \in \bar{F}$ 的作用下,使 X 中的一部分信息元素丢失,使 X 中的信息减少。

元素迁移的概念也可应用到属性集 $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_k\}$ 中^[8],它用下面的式子表示:

$$\exists \beta \in V, \beta_i \in \alpha \Rightarrow f(\beta_i) = \alpha'_i \in \alpha$$

$$\exists \alpha_j \in \alpha \Rightarrow \bar{f}(\alpha_j) = \gamma_j \notin \alpha$$

显然,属性集 $\{\alpha_1, \alpha_2, \dots, \alpha_k\}$ 变为 $\{\alpha_1, \alpha_2, \dots, \alpha_{j-1}, \alpha_{j+1}, \dots, \alpha_k\}$,

^{*})国家自然科学基金项目(60364001)、山东省自然科学基金(Y2004A04)。刘纪芹 副教授,博士研究生,研究方向:粗系统理论与应用。

$\beta\}$, 其中 V 为属性论域。

由于元素迁移 $f \in F, \bar{f} \in \bar{F}$ 的存在, 属性集 α 中的属性得到补充和删除。如果把属性集 $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_k\}$ 看作 U 上的知识, 则说明由于 $f \in F$ 的存在, 使 U 上的知识增长即人们的知识增加。又由于 $\bar{f}$ 的存在, 使陈旧无用或错误的知识删除。

定义 1.2^[6~8] 若

$$X^* = X' \cup \{u | u \in U, u \in X, f(u) = x \in X\}$$

称 X^* 是 $X \subseteq U$ 的双向 S -集合 (Two-direction Singular Set), 这里 X' 称作 $X \subseteq U$ 的亏集 (loss set), 而且

$$X' = X \setminus \{x | x \in X, \bar{f}(x) = u \in X\}$$

定义 1.3^[6~8] 若

$$(R, \mathcal{F})_o(X^*) = \bigcup [x]_R = \{x | x \in U, [x]_R \subseteq X^*\}$$

称 $(R, \mathcal{F})_o(X^*)$ 是 $X^* \subseteq U$ 的下近似; 若

$$(R, \mathcal{F})^o(X^*) = \bigcup [x]_R = \{x | x \in U, [x]_R \cap X^* \neq \emptyset\}$$

称 $(R, \mathcal{F})^o(X^*)$ 是 $X^* \subseteq U$ 的上近似, 这里 $\mathcal{F} = F \cup \bar{F}, F \neq \emptyset, \bar{F} \neq \emptyset$ 。

定义 1.4 称

$$pos_{(R, \mathcal{F})}(X^*) = (R, \mathcal{F})_o(X^*)$$

为双向 S -集合 X^* 的正域。称

$$neg_{(R, \mathcal{F})}(X^*) = U - (R, \mathcal{F})^o(X^*) = \{x | x \in U, [x]_R \cap X^* = \emptyset\}$$

为双向 S -集合 X^* 的负域。称

$$bn_{(R, \mathcal{F})}(X^*) = (R, \mathcal{F})^o(X^*) - (R, \mathcal{F})_o(X^*)$$

为双向 S -集合 X^* 的边界域。

由定义 1.3、1.4 知, 若 $pos_{(R, \mathcal{F})}(X^*) \cup neg_{(R, \mathcal{F})}(X^*) = U$, 即 $bn_{(R, \mathcal{F})}(X^*) = \emptyset$, $(R, \mathcal{F})^o(X^*) = (R, \mathcal{F})_o(X^*)$, 则具有知识 R 的个体对 $X^* \subseteq U$ 的认识是精确的; 若 $pos_{(R, \mathcal{F})}(X^*) \cup neg_{(R, \mathcal{F})}(X^*) \neq U$, 即 $bn_{(R, \mathcal{F})}(X^*) \neq \emptyset$, $(R, \mathcal{F})^o(X^*) \subset (R, \mathcal{F})_o(X^*)$, 则具有知识 R 的个体对 $X^* \subseteq U$ 的认识是粗糙、近似的。因此由 $pos_{(R, \mathcal{F})}(X^*)$ 与 $neg_{(R, \mathcal{F})}(X^*)$ 构成的集合对 $(pos_{(R, \mathcal{F})}(X^*), neg_{(R, \mathcal{F})}(X^*))$ 可以作为具有知识 R 的个体对 $X^* \subseteq U$ 的认识, 由此给出如下定义。

定义 1.5 称集合对 $(pos_{(R, \mathcal{F})}(X^*), neg_{(R, \mathcal{F})}(X^*))$ 为具有知识 R 的个体对动态信息源 $X^* \subseteq U$ 的双向 S -动态认识。

对于这种集合对之间的关系, 为了后面讨论的方便, 下面给出集合对包含和相等的定义。

定义 1.6 设给定两个集合对 $(A_1, A_2), (B_1, B_2)$, 若

$$A_1 \subseteq B_1, A_2 \subseteq B_2$$

则称集合对 (B_1, B_2) 包含集合对 (A_1, A_2) , 记为 $(A_1, A_2) \subseteq (B_1, B_2)$ 。特别地, 若

$$A_1 = B_1, A_2 = B_2$$

则称集合对 (A_1, A_2) 与集合对 (B_1, B_2) 相等, 记为 $(A_1, A_2) = (B_1, B_2)$ 。

2 双向 S -下近似动态粗传递模型

2.1 两个个体之间的双向 S -下近似动态粗传递模型及其特性

约定 设 $(U, R_1), (U, R_2)$ 为两个知识库, A_1, A_2 分别为来自知识库 $(U, R_1), (U, R_2)$ 的个体, 则 R_1, R_2 分别为个体 A_1, A_2 的知识。假设在这两个个体之间对动态信息源 $X^* \subseteq U$ 进行信息传递, 集合对 $(pos_{(R_1, \mathcal{F})}(X^*), neg_{(R_1, \mathcal{F})}(X^*))$ 为 A_1 根据自己的知识 R_1 对 X^* 的双向 S -动态认识。为书写方便起见, 令 $G_1^+ = pos_{(R_1, \mathcal{F})}(X^*), G_1^- = neg_{(R_1, \mathcal{F})}(X^*)$ 。

在实际中进行信息传递时, 常常会遇到如下情况: 为了稳

妥起见, 我们需要准确可靠的信息, 信息不能是模糊的、不可靠的。针对这种情况下的信息传递, 下面利用双向 S -粗集建立这种信息粗传递模型, 先讨论信息在两个个体之间进行传递, 给出如下定义:

定义 2.1 设 (G_1^+, G_1^-) 为 A_1 对动态信息源 X^* 的双向 S -动态认识, 称 $\underline{apr}_{(R_2, \mathcal{F})}(G_1^+, G_1^-)$ 为 X^* 由个体 A_1 传递给 A_2 的双向 S -下近似动态粗传递。若

$$\underline{apr}_{(R_2, \mathcal{F})}(G_1^+, G_1^-) = (pos_{(R_2, \mathcal{F})}(G_1^+), pos_{(R_2, \mathcal{F})}(G_1^-))$$

其中 $pos_{(R_2, \mathcal{F})}(X^*), pos_{(R_2, \mathcal{F})}(G_1^-)$ 分别为 G_1^+, G_1^- 的 R_2 双向 S -下近似。称

$$|\underline{apr}_{(R_2, \mathcal{F})}(G_1^+, G_1^-)| = |(pos_{(R_2, \mathcal{F})}(G_1^+), pos_{(R_2, \mathcal{F})}(G_1^-))| = |pos_{(R_2, \mathcal{F})}(G_1^+)| + |pos_{(R_2, \mathcal{F})}(G_1^-)|$$

为 X^* 由 A_1 传递给 A_2 的双向 S -下近似动态粗传递信息量, 其中 $|X|$ 为集合 X 的基数。

由定义 1.3、1.4 知算子 $\underline{apr}_{(R_2, \mathcal{F})}(G_1^+, G_1^-)$ 的含义: 为了得到更准确的信息, A_2 根据自己的知识 R_2 只接受肯定包含在 (G_1^+, G_1^-) 中的信息, 对于不确定的信息则排除在外。

由于 $pos_{(R_2, \mathcal{F})}(G_1^+), pos_{(R_2, \mathcal{F})}(G_1^-)$ 分别为 G_1^+, G_1^- 的 R_2 双向 S -下近似, 由定义 1.3、1.4 及定义 1.6 知, $\underline{apr}_{(R_2, \mathcal{F})}(G_1^+, G_1^-) \subseteq (G_1^+, G_1^-)$, 则 $|\underline{apr}_{(R_2, \mathcal{F})}(G_1^+, G_1^-)| \leq |(G_1^+, G_1^-)|$ 或 $|\underline{apr}_{(R_2, \mathcal{F})}(G_1^+, G_1^-)| = |(G_1^+, G_1^-)|$ 。故 X^* 沿信息传递路径 $X^* \rightarrow A_1 \rightarrow A_2$ 进行下近似动态粗传递时, 由于个体本身具备的知识不同和知识的粒度, 信息传递存在两种结果, 即信息保持不变或发生损失。下面对信息发生变化的条件进行分析, 为此先给出如下定义。

定义 2.2 称量 $|(G_1^+, G_1^-) \setminus \underline{apr}_{(R_2, \mathcal{F})}(G_1^+, G_1^-)|$ 为 X^* 沿信息传递路径 $X^* \rightarrow A_1 \rightarrow A_2$ 传递时双向 S -下近似动态粗传递信息损失量, 记为 $Losq_{A_1 \rightarrow A_2}$, 即

$$\begin{aligned} Losq_{A_1 \rightarrow A_2} &= |(G_1^+, G_1^-) \setminus \underline{apr}_{(R_2, \mathcal{F})}(G_1^+, G_1^-)| \\ &= |(G_1^+ \setminus pos_{(R_2, \mathcal{F})}(G_1^+), G_1^- \setminus pos_{(R_2, \mathcal{F})}(G_1^-))| \\ &= |G_1^+ \setminus pos_{(R_2, \mathcal{F})}(G_1^+)| + |G_1^- \setminus pos_{(R_2, \mathcal{F})}(G_1^-)| \end{aligned}$$

定义 2.3 称

$$Losd_{A_1 \rightarrow A_2} = \begin{cases} \frac{Losq_{A_1 \rightarrow A_2}}{|(G_1^+, G_1^-)|}, & |(G_1^+, G_1^-)| \neq 0 \\ 0, & |(G_1^+, G_1^-)| = 0 \end{cases}$$

为 X^* 沿信息传递路径 $X^* \rightarrow A_1 \rightarrow A_2$ 传递时双向 S -下近似动态粗传递信息损失度。

利用以上定义可得传递过程中信息保持不变的条件, 有如下定理。

定理 2.1 (双向 S -下近似动态粗传递信息不变定理 1)

设 $(U, R_1), (U, R_2)$ 为两个知识库, 动态信息源 $X^* \subseteq U$, 信息传递路径为 $X^* \rightarrow A_1 \rightarrow A_2$ 。 R_1, R_2 分别为个体 A_1, A_2 的知识。若 A_1, A_2 的知识相同, 即 $U/ind(R_1) = U/ind(R_2)$, 则

(1) 信息由个体 A_1 传递给 A_2 时没有发生损失, 即 $\underline{apr}_{(R_2, \mathcal{F})}(G_1^+, G_1^-) = (G_1^+, G_1^-)$;

(2) $Losd_{A_1 \rightarrow A_2} = 0$ 。

由定义 2.1~2.3 易证, 略。

定理 2.2 (双向 S -下近似动态粗传递信息不变定理 2)

设 $(U, R_1), (U, R_2)$ 为两个知识库, $X^* \subseteq U$, 信息传递路径为 $X^* \rightarrow A_1 \rightarrow A_2$ 。 R_1, R_2 分别为个体 A_1, A_2 的知识。若 $U/ind(R_2) \subset U/ind(R_1)$, 则

(1) $\underline{apr}_{(R_2, \mathcal{F})}(G_1^+, G_1^-) = (G_1^+, G_1^-)$;

(2) $Losd_{A_1 \rightarrow A_2} = 0$ 。

证明 (1): 设 $U/ind(R_1) = \{X_1, X_2, \dots, X_m\}, U/ind(R_2) = \{Y_1, Y_2, \dots, Y_n\}$, 因为 $U/ind(R_2) \subset U/ind(R_1)$, 则 $m < n$,

而且存在 $\{1, 2, \dots, n\}$ 的一个划分 $\{D_1, D_2, \dots, D_m\}$, 使得 $X_i = \bigcup_{j \in D_i} Y_j$. 设 $G_1^+ = \text{pos}_{(R_1, \mathcal{F})}(X^*) = \{x \mid [x]_{R_1} \subseteq X^*\} = X_{i_1} \cup X_{i_2} \cup \dots \cup X_{i_t}$, 其中 $\{i_1, i_2, \dots, i_t\} \subseteq \{1, 2, \dots, m\}$. 故 $X_{i_l} \subseteq G_1^+$, 又 $X_i = \bigcup_{j \in D_i} Y_j$, 即 $X_{i_l} = \bigcup_{j \in D_{i_l}} Y_j \subseteq G_1^+$, $l=1, 2, \dots, t$. 从而

$$\text{pos}_{(R_2, \mathcal{F})}(G_1^+) = \{x \mid [x]_{R_2} \subseteq G_1^+\} = (\bigcup_{j \in D_{i_1}} Y_j) \cup (\bigcup_{j \in D_{i_2}} Y_j) \cup \dots \cup (\bigcup_{j \in D_{i_t}} Y_j) = X_{i_1} \cup X_{i_2} \cup \dots \cup X_{i_t} = G_1^+$$

同理可证 $\text{pos}_{(R_2, \mathcal{F})}(G_1^-) = G_1^-$. 故

$$\text{apr}_{(R_2, \mathcal{F})}(G_1^+, G_1^-) = (G_1^+, G_1^-)$$

(2) 由(1)及定义 2.3 知显然, 证明略。

在定理 2.1、2.2 中 A_1, A_2 的知识没有发生变化, 但随着对事物认识的深入, 若个体 A_1, A_2 均吸收新的知识, 删除陈旧无用的知识, 即它们的知识发生变化。下面利用元素迁移的概念来讨论 A_1, A_2 的知识发生变化对双向 S -动态粗传递的影响。先给出以下几个符号。

设 R_1, R_2 为个体 A_1, A_2 的知识, 若 A_1, A_2 均吸收新的知识, 删除冗余知识。设 A_1 的知识变化为 $R_1 \cup \{f_{11}(\alpha_{11}), f_{12}(\alpha_{12}), \dots, f_{1s}(\alpha_{1s})\} \setminus \{\bar{f}_{11}(\beta_{11}), \bar{f}_{12}(\beta_{12}), \dots, \bar{f}_{1l}(\beta_{1l})\}$. 为书写方便起见, 简记为 $R_1 \cup \alpha_1' \setminus \beta_1'$, 其中 $\alpha_1' = \{f_{11}(\alpha_{11}), f_{12}(\alpha_{12}), \dots, f_{1s}(\alpha_{1s})\}$ 为 A_1 吸收的新知识, $\beta_1' = \{\bar{f}_{11}(\beta_{11}), \bar{f}_{12}(\beta_{12}), \dots, \bar{f}_{1l}(\beta_{1l})\}$ 为 A_1 删除的冗余知识。这里 $\alpha_{1i} \in R_1$, 元素迁移 $f_{1i} \in F$ 将 α_{1i} 迁移进 R_1 中, $i=1, 2, \dots, s$; $\beta_{1j} \in R_1$, 元素迁移 $\bar{f}_{1j} \in F$ 将 β_{1j} 迁移出 R_1 中, $j=1, 2, \dots, l$. A_2 的知识变化为 $R_2 \cup \{f_{21}(\alpha_{21}), f_{22}(\alpha_{22}), \dots, f_{2h}(\alpha_{2h})\} \setminus \{\bar{f}_{21}(\beta_{21}), \bar{f}_{22}(\beta_{22}), \dots, \bar{f}_{2t}(\beta_{2t})\}$, 记为 $R_2 \cup \alpha_2' \setminus \beta_2'$, 其中 $\alpha_2' = \{f_{21}(\alpha_{21}), f_{22}(\alpha_{22}), \dots, f_{2h}(\alpha_{2h})\}$ 为 A_2 吸收的新知识, $\beta_2' = \{\bar{f}_{21}(\beta_{21}), \bar{f}_{22}(\beta_{22}), \dots, \bar{f}_{2t}(\beta_{2t})\}$ 为 A_2 删除的冗余知识。这里 $\alpha_{2i} \in R_2$, 元素迁移 $f_{2i} \in F$ 将 α_{2i} 迁移进 R_2 中, $i=1, 2, \dots, h$; $\beta_{2j} \in R_2$, 元素迁移 $\bar{f}_{2j} \in F$ 将 β_{2j} 迁移出 R_2 中, $j=1, 2, \dots, t$.

定理 2.3 (双向 S -下近似动态粗传递信息不变定理 3)

设 $(U, R_1), (U, R_2)$ 为两个知识库, $X^* \subseteq U$, 信息传递路径为 $X^* \rightarrow A_1 \rightarrow A_2$. R_1, R_2 分别为 A_1, A_2 的知识。若 A_1, A_2 的知识发生变化, 分别变化为 $R_1 \cup \alpha_1' \setminus \beta_1'$ 和 $R_2 \cup \alpha_2' \setminus \beta_2'$, 且满足 $U/\text{ind}\{R_2 \cup \alpha_2' \setminus \beta_2'\} \subseteq U/\text{ind}\{R_1 \cup \alpha_1' \setminus \beta_1'\}$, 则

(1) $\text{apr}_{(R_2 \cup \alpha_2' \setminus \beta_2', \mathcal{F})}(G_1^+, G_1^-) = (G_1^+, G_1^-)$, 其中 (G_1^+, G_1^-) 为具有知识 $R_1 \cup \alpha_1' \setminus \beta_1'$ 的个体对动态信息源 $X^* \subseteq U$ 的双向 S -动态认识;

(2) $\text{Losd}_{A_1 \rightarrow A_2} = 0$.

由定理 2.1、2.2 易证, 略。

定理 2.4 (双向 S -下近似动态粗传递信息损失定理 1)

设 $(U, R_1), (U, R_2)$ 为两个知识库, $X^* \subseteq U$, 信息传递路径为 $X^* \rightarrow A_1 \rightarrow A_2$. R_1, R_2 为个体 A_1, A_2 的知识。若 $U/\text{ind}(R_1) \subset U/\text{ind}(R_2)$, 则

(1) $\text{apr}_{(R_2, \mathcal{F})}(G_1^+, G_1^-) \subseteq (G_1^+, G_1^-)$;

(2) $\text{Losd}_{A_1 \rightarrow A_2} \geq 0$.

由定义 1.3 及定义 2.1~2.3 知显然成立, 证略。

说明: 定理 2.4 结论中的等号不能去掉, 它在下面特殊情况下成立, 先看如下例子:

例 2.1 设 $(U, R_1), (U, R_2)$ 为两个知识库, 论域 $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $X = \{1, 2, 3, 6, 7, 9\}$. 若元素 4, 5 被迁移进 X , 元素 6, 9 被迁移出 X , 从而得 $X^* = \{1, 2, 3, 4, 5, 7\}$. 设 $U/\text{ind}(R_1) = \{\{1, 2\}, \{3, 4\}, \{5, 6\}, \{7, 8\}, \{9\}, \{10\}\}$, $U/\text{ind}(R_2) = \{\{1, 2, 3, 4\}, \{5, 6, 7, 8\}, \{9, 10\}\}$. 显然, $U/\text{ind}(R_1) \subset U/\text{ind}(R_2)$. $G_1^+ = \text{pos}_{(R_1, \mathcal{F})}(X^*) = \{1, 2, 3, 4\}$,

$G_1^- = \text{neg}_{(R_1, \mathcal{F})}(X^*) = \{9, 10\}$. 由定义 2.1 知, $\text{apr}_{(R_2, \mathcal{F})}(G_1^+, G_1^-) = (\{1, 2, 3, 4\}, \{9, 10\}) = (G_1^+, G_1^-)$.

比较 $U/\text{ind}(R_1)$ 和 $U/\text{ind}(R_2)$ 可知, $U/\text{ind}(R_1)$ 被划分的部分恰好是 X^* 关于知识库 (U, R_1) 的正域与负域部分。这说明在定理 2.4 中, 因为 $U/\text{ind}(R_1) \subset U/\text{ind}(R_2)$, 若与 $U/\text{ind}(R_2)$ 相比, $U/\text{ind}(R_1)$ 被划分的部分恰好是 X^* 关于知识库 (U, R_1) 的正域与负域部分, 则结论中的等号成立, 即信息没有发生损失。排除这种情况, 即在一般情况下, 有 (1) $\text{apr}_{(R_2, \mathcal{F})}(G_1^+, G_1^-) \subset (G_1^+, G_1^-)$, (2) $\text{Losd}_{A_1 \rightarrow A_2} > 0$, 即信息发生损失。

推论 2.1 设 $(U, R_1), (U, R_2)$ 为两个知识库, $X^* \subseteq U$, 信息传递路径为 $X^* \rightarrow A_1 \rightarrow A_2$. R_1, R_2 分别为个体 A_1, A_2 的知识。若 $U/\text{ind}(R_1) \subset U/\text{ind}(R_2)$, 且与 $U/\text{ind}(R_2)$ 相比, $U/\text{ind}(R_1)$ 被划分的部分不是 X^* 关于知识库 (U, R_1) 的正域与负域部分, 则

(1) $\text{apr}_{(R_2, \mathcal{F})}(G_1^+, G_1^-) \subset (G_1^+, G_1^-)$;

(2) $\text{Losd}_{A_1 \rightarrow A_2} > 0$.

定理 2.5 (双向 S -下近似动态粗传递信息损失定理 2)

设 $(U, R_1), (U, R_2)$ 为两个知识库, $X^* \subseteq U$, 信息传递路径为 $X^* \rightarrow A_1 \rightarrow A_2$. R_1, R_2 分别为个体 A_1, A_2 的知识。若 A_1, A_2 的知识发生变化, 分别变化为 $R_1 \cup \alpha_1' \setminus \beta_1'$ 和 $R_2 \cup \alpha_2' \setminus \beta_2'$, 且满足 $U/\text{ind}\{R_1 \cup \alpha_1' \setminus \beta_1'\} \subset U/\text{ind}\{R_2 \cup \alpha_2' \setminus \beta_2'\}$, 则

(1) $\text{apr}_{(R_2 \cup \alpha_2' \setminus \beta_2', \mathcal{F})}(G_1^+, G_1^-) \subseteq (G_1^+, G_1^-)$, 其中 (G_1^+, G_1^-) 为具有知识 $R_1 \cup \alpha_1' \setminus \beta_1'$ 的个体对 $X^* \subseteq U$ 的双向 S -动态认识;

(2) $\text{Losd}_{A_1 \rightarrow A_2} \geq 0$.

推论 2.2 设 $(U, R_1), (U, R_2)$ 为两个知识库, $X^* \subseteq U$, 信息传递路径为 $X^* \rightarrow A_1 \rightarrow A_2$. R_1, R_2 分别为个体 A_1, A_2 的知识。若 A_1, A_2 的知识发生变化, 分别变化为 $R_1 \cup \alpha_1' \setminus \beta_1'$ 和 $R_2 \cup \alpha_2' \setminus \beta_2'$, 且满足 $U/\text{ind}\{R_1 \cup \alpha_1' \setminus \beta_1'\} \subset U/\text{ind}\{R_2 \cup \alpha_2' \setminus \beta_2'\}$, 若 $U/\text{ind}\{R_1 \cup \alpha_1' \setminus \beta_1'\}$ 与 $U/\text{ind}\{R_2 \cup \alpha_2' \setminus \beta_2'\}$ 相比, $U/\text{ind}\{R_1 \cup \alpha_1' \setminus \beta_1'\}$ 被划分的部分不是 X^* 关于知识库 $(U, R_1 \cup \alpha_1' \setminus \beta_1')$ 的正域与负域部分, 则

(1) $\text{apr}_{(R_2 \cup \alpha_2' \setminus \beta_2', \mathcal{F})}(G_1^+, G_1^-) \subset (G_1^+, G_1^-)$, 其中 (G_1^+, G_1^-) 为具有知识 $R_1 \cup \alpha_1' \setminus \beta_1'$ 的个体对 $X^* \subseteq U$ 的双向 S -动态认识;

(2) $\text{Losd}_{A_1 \rightarrow A_2} > 0$.

由定理 2.1~2.5, 可得提高动态粗传递精确性的方法。

定理 2.6 设 $(U, R_1), (U, R_2)$ 为两个知识库, $X^* \subseteq U$, 信息传递路径为 $X^* \rightarrow A_1 \rightarrow A_2$. R_1, R_2 为个体 A_1, A_2 的知识。若 $U/\text{ind}(R_1) \subset U/\text{ind}(R_2)$, 且与 $U/\text{ind}(R_2)$ 相比, $U/\text{ind}(R_1)$ 被划分的部分不是 X^* 关于知识库 (U, R_1) 的正域与负域部分, 则 A_2 通过丰富自己的知识, 可减少信息的损失及损失度。

证明: 由已知条件和推论 2.1 知, 按双向 S -下近似方式在个体 A_1, A_2 之间进行信息传递时信息发生损失。设 A_2 在知识 R_2 基础上又丰富自己的知识, 不妨设 A_2 的知识变化为 $R_2 \cup \alpha' \setminus \beta'$, 其中 α', β' 含义同定理 2.3. 为书写方便, $R_2 \cup \alpha' \setminus \beta'$ 记为 R_2' . 因为 A_2 的知识变丰富, 显然有 $U/\text{ind}(R_2') \subset U/\text{ind}(R_2)$, 则 $[x]_{R_2'} \subseteq [x]_{R_2}$.

设 $x \in \text{pos}_{(R_2, \mathcal{F})}(G_1^+) = \{x \in U \mid [x]_{R_2} \subseteq G_1^+\}$, 又 $[x]_{R_2'} \subseteq [x]_{R_2}$,

$$\Rightarrow x \in \{x \in U \mid [x]_{R_2'} \subseteq G_1^+\} = \text{pos}_{(R_2', \mathcal{F})}(G_1^+)$$

故 $\text{pos}_{(R_2, \mathcal{F})}(G_1^+) \subseteq \text{pos}_{(R_2', \mathcal{F})}(G_1^+)$

同理可证 $pos_{(R_2, \mathcal{F})}(G_1^-) \subseteq pos_{(R_2', \mathcal{F})}(G_1^-)$

则由定义 1.6, 有

$$(pos_{(R_2, \mathcal{F})}(G_1^+), pos_{(R_2, \mathcal{F})}(G_1^-)) \subseteq (pos_{(R_2', \mathcal{F})}(G_1^+), pos_{(R_2', \mathcal{F})}(G_1^-))$$

故 $apr_{(R_2, \mathcal{F})}(G_1^+, G_1^-) \subseteq apr_{(R_2', \mathcal{F})}(G_1^+, G_1^-)$

$$\begin{aligned} Losq_{A_1 \rightarrow A_2}^{R_2'} &= |(G_1^+, G_1^-) \setminus apr_{(R_2', \mathcal{F})}(G_1^+, G_1^-)| \\ &\leq |(G_1^+, G_1^-) \setminus apr_{(R_2, \mathcal{F})}(G_1^+, G_1^-)| = Losq_{A_1 \rightarrow A_2}^{R_2} \end{aligned}$$

故

$$Losd_{A_1 \rightarrow A_2}^{R_2'} \leq Losd_{A_1 \rightarrow A_2}^{R_2}$$

因此 A_2 通过丰富自己的知识, 获得的信息量增多, 从而减少了信息的损失, 使信息损失度变小, 使信息传递的精确度提高。

动态信息源 X^* 沿另一信息传递路径 $X^* \rightarrow A_2 \rightarrow A_1$ 进行信息传递时的情况可类似讨论, 略。

2.2 $n(n > 2)$ 个个体之间的双向 S-下近似动态粗传递模型及其特性

类似可讨论 n 个个体之间对同一动态信息源 $X^* \subseteq U$ 的双向 S-下近似动态粗传递, 下面只给出定义和结论, 证明过程略。

定义 2.4 设 $(U, R_1), (U, R_2), \dots, (U, R_n)$ 为 n 个知识库, $R_1, R_2, \dots, R_n$ 分别为个体 $A_1, A_2, \dots, A_n$ 的知识, 称

$$\begin{aligned} (G_1^+, G_1^-) &= (pos_{(R_1, \mathcal{F})}(X^*), neg_{(R_1, \mathcal{F})}(X^*)) \\ (G_i^+, G_i^-) &= apr_{(R_i, \mathcal{F})}(G_{i-1}^+, G_{i-1}^-) = (pos_{(R_i, \mathcal{F})}(G_{i-1}^+), pos_{(R_i, \mathcal{F})}(G_{i-1}^-)), i=2, 3, \dots, n \end{aligned}$$

为动态信息源 X^* 沿路径 $X^* \rightarrow A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_n$ 进行信息传递的双向 S-下近似动态粗传递。

类似地, 我们可以定义 X^* 沿其它路径进行信息传递时的双向 S-下近似动态粗传递。

设 $(U, R_1), (U, R_2), \dots, (U, R_n)$ 为 n 个知识库, $R_1, R_2, \dots, R_n$ 分别为 $A_1, A_2, \dots, A_n$ 的知识, 显然 X^* 将通过 $A_1, A_2, \dots, A_{n-1}$ 传递给 A_n 的路径数有 $(n-1)!$ 个。

定理 2.7 设 $(U, R_1), (U, R_2), \dots, (U, R_n)$ 为 n 个知识库, $R_1, R_2, \dots, R_n$ 分别为 $A_1, A_2, \dots, A_n$ 的知识, 信息传递路径为 $X^* \rightarrow A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_n$ 。若 $U/ind(R_1) \supseteq U/ind(R_2) \supseteq \dots \supseteq U/ind(R_n)$, 则在传递过程中信息保持不变, 即

$$(G_1^+, G_1^-) = (G_2^+, G_2^-) = \dots = (G_n^+, G_n^-)$$

定理 2.8 设 $(U, R_1), (U, R_2), \dots, (U, R_n)$ 为 n 个知识库, $R_1, R_2, \dots, R_n$ 分别为 $A_1, A_2, \dots, A_n$ 的知识, 信息传递路径为 $X^* \rightarrow A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_n$ 。若 $A_1, A_2, \dots, A_n$ 的知识均发生变化, A_i 的知识由 R_i 变化为 $R_i \cup \alpha_i' \setminus \beta_i'$, 其中 α_i', β_i' 含义同定理 2.3, $i=1, 2, \dots, n$, 且满足

$$U/ind(R_1 \cup \alpha_1' \setminus \beta_1') \supseteq U/ind(R_2 \cup \alpha_2' \setminus \beta_2') \supseteq \dots \supseteq U/ind(R_n \cup \alpha_n' \setminus \beta_n')$$

则在传递过程中, 信息保持不变, 即

$$(G_1^+, G_1^-) = (G_2^+, G_2^-) = \dots = (G_n^+, G_n^-)$$

其中 (G_i^+, G_i^-) 为个体 A_i 根据自己的知识 $R_i \cup \alpha_i' \setminus \beta_i'$ 得到的 A_{i-1} 传递过来的信息, $i=2, 3, \dots, n$ 。

定理 2.9 设 $(U, R_1), (U, R_2), \dots, (U, R_n)$ 为 n 个知识库, $R_1, R_2, \dots, R_n$ 分别为 $A_1, A_2, \dots, A_n$ 的知识, 信息传递路径为 $X^* \rightarrow A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_n$ 。若 $U/ind(R_1) \subset U/ind(R_2) \subset \dots \subset U/ind(R_n)$, 则

$$(1) (G_1^+, G_1^-) \supseteq (G_2^+, G_2^-) \supseteq \dots \supseteq (G_n^+, G_n^-)。$$

若与 $U/ind(R_i)$ 相比, $U/ind(R_{i-1})$ 被划分的部分不是 X^* 关于知识库 (U, R_{i-1}) 的正域与负域部分, $i=2, 3, \dots, n$, 则

$$(2) (G_1^+, G_1^-) \supset (G_2^+, G_2^-) \supset \dots \supset (G_n^+, G_n^-)。$$

定理 2.10 设 $(U, R_1), (U, R_2), \dots, (U, R_n)$ 为 n 个知识库, $R_1, R_2, \dots, R_n$ 分别为 $A_1, A_2, \dots, A_n$ 的知识, 信息传递路径为 $X^* \rightarrow A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_n$ 。若 $A_1, A_2, \dots, A_n$ 的知识均发生变化, A_i 的知识由 R_i 变化为 $R_i \cup \alpha_i' \setminus \beta_i'$, 其中 α_i', β_i' 含义同定理 2.3, $i=1, 2, \dots, n$, 且满足

$$U/ind(R_1 \cup \alpha_1' \setminus \beta_1') \subset U/ind(R_2 \cup \alpha_2' \setminus \beta_2') \subset \dots \subset U/ind(R_n \cup \alpha_n' \setminus \beta_n')$$

则

$$(1) (G_1^+, G_1^-) \supseteq (G_2^+, G_2^-) \supseteq \dots \supseteq (G_n^+, G_n^-)。$$

若与 $U/ind(R_i \cup \alpha_i' \setminus \beta_i')$ 相比, $U/ind(R_{i-1} \cup \alpha_{i-1}' \setminus \beta_{i-1}')$ 被划分的部分不是 X^* 关于知识库 $(U, R_{i-1} \cup \alpha_{i-1}' \setminus \beta_{i-1}')$ 的正域与负域部分, $i=2, 3, \dots, n$, 则

$$(2) (G_1^+, G_1^-) \supset (G_2^+, G_2^-) \supset \dots \supset (G_n^+, G_n^-)$$

其中 (G_i^+, G_i^-) 为个体 A_i 根据自己的知识 $R_i \cup \alpha_i' \setminus \beta_i'$ 得到的 A_{i-1} 传递过来的信息, $i=2, 3, \dots, n$ 。

3 应用示例

例 3.1 在案件的侦探工作中, 设 $U = \{u_1, u_2, \dots, u_{15}\}$ 是某案件中犯罪嫌疑人构成的论域, 取重大嫌疑人构成的集合 $X = \{u_2, u_3, u_6, u_8, u_{12}\}$ 。在案件的进一步侦探过程中, 发现 u_9, u_{10} 有重大作案嫌疑, 故将其迁移到 X 中, u_8 非重大嫌疑人, 将其迁出 X , 故得动态信息源 $X^* = \{u_2, u_3, u_6, u_9, u_{10}, u_{12}\}$ 。设三位刑警 A_1, A_2, A_3 , 通过对案件的调查和分析, 形成了对这 15 位犯罪嫌疑人的分类(即知识):

$$A_1 \text{ 的知识: } U/ind(R_1) = \{\{u_1\}, \{u_2, u_3\}, \{u_4, u_5, u_6\}, \{u_7, u_8\}, \{u_9, u_{10}\}, \{u_{11}, u_{12}, u_{13}\}, \{u_{14}, u_{15}\}\};$$

$$A_2 \text{ 的知识: } U/ind(R_2) = \{\{u_1\}, \{u_2, u_3\}, \{u_4\}, \{u_5, u_6\}, \{u_7, u_8, u_{11}\}, \{u_9, u_{10}\}, \{u_{12}, u_{13}\}, \{u_{14}, u_{15}\}\};$$

$$A_3 \text{ 的知识: } U/ind(R_3) = \{\{u_1, u_4\}, \{u_2, u_3, u_5\}, \{u_6, u_7, u_8\}, \{u_9, u_{10}\}, \{u_{11}, u_{12}, u_{13}, u_{14}, u_{15}\}\}。$$

为了使得到的信息准确可靠, 从中找到最重要的重大嫌疑人, 尽快侦破此案, 采取下近似粗传递方式。假设按信息传递路径 $X^* \rightarrow A_1 \rightarrow A_2 \rightarrow A_3$, 得

A_1 根据自己知识对 X^* 的认识为

$$(G_1^+, G_1^-) = \{\{u_2, u_3, u_9, u_{10}\}, \{u_1, u_7, u_8, u_{14}, u_{15}\}\}$$

A_2 根据自己的知识得到的 A_1 传递过来的信息为

$$(G_2^+, G_2^-) = \{\{u_2, u_3, u_9, u_{10}\}, \{u_1, u_{14}, u_{15}\}\}$$

设 A_3 对案件做了进一步调查和分析, 其知识 R_3 发生变化, 变为

$$U/ind(R_3 \cup \alpha_3' \setminus \beta_3') = \{\{u_1\}, \{u_2\}, \{u_3, u_4, u_5\}, \{u_6\}, \{u_7, u_8\}, \{u_9, u_{10}\}, \{u_{11}, u_{12}, u_{13}\}, \{u_{14}, u_{15}\}\}$$

这里 α_3', β_3' 含义同定理 2.3。则 A_3 根据自己新的知识得到的 A_2 传递过来的信息为

$$(G_3^+, G_3^-) = \{\{u_2, u_9, u_{10}\}, \{u_1, u_{14}, u_{15}\}\}$$

$$Losd_{A_1 \rightarrow A_2} = 2/9, Losd_{A_2 \rightarrow A_3} = 1/7$$

通过分析最终结果, 刑警们认为 u_2, u_9, u_{10} 为最重要的重大犯罪嫌疑人, 对他们进行重点审问, 从中找到重大线索和案件的突破口, 尽快侦破此案。

结束语 由于知识的局限性, 人们对现实中很多事物的认识是不完全、不精确的, 在这种情况下进行信息传递时会产生传递的粗糙性, 这种粗传递是静态的粗传递。事实上, 人们知识在不断变化。另一方面, 信息源也在不断变化, 这种情况下的信息传递即为动态粗传递。根据实际中要求传递过来的信息必须准确可靠这种情况, 结合双向 S-粗集的动态特性, 本文给出了一种动态粗传递模型, 即双向 S-下近似动态粗传

递模型,进一步分析了这种动态粗传递的特性,得到了动态粗传递信息不变定理、信息损失定理,及提高粗传递精确性的方法。最后给出了该模型的应用实例。通过研究得出:可以用双向 S 粗集来研究这种动态粗传递问题,动态粗传递是双向 S 粗集的一种新的应用。对于其它信息传递情况下的动态粗传递模型,限于篇幅,将另文讨论。

参考文献

- 1 耿志强,朱群雄,李芳. 知识粗糙性的粒度原理及其约简[J]. 系统工程与电子技术, 2004, 26(8):1112~1116
- 2 Mousavi A, Jabadar-Maralani P. Double-faced rough sets and rough communication [J]. Information Sciences, 2002(148): 41~53
- 3 Pawlak Z. Rough sets [J]. International Journal of Computer and Information Sciences, 1982(11):341~356
- 4 Pawlak Z. Rough sets: Theoretical Aspects of Reasoning about

- Data [M]. Boston, NA: Kluwer Academic Publishers, 1991
- 5 王瑜,胡运发,张凯. 基于粗集理论的知识含量度量研究[J]. 计算机研究与发展, 2004, 41(9):1500~1506
- 6 SHI Kaiquan. Two-direction S-rough sets [J]. International Journal of Fuzzy Mathematics, 2005, 13(2): 335~349
- 7 SHI Kaiquan. S-Rough sets and knowledge separation [J]. Journal of System Engineering and Electronics, 2005(2):32~37
- 8 SHI Kaiquan, CUI Yuquan. F-decomposition and $\bar{F}$ -reduction of S-rough sets [J]. An International Journal of Advances in Systems Sciences and Applications, 2004(4): 487~499
- 9 LU Changjing, SHI Kaiquan. Knowledge filter and its dependent reasoning discovery [J]. International Journal of Fuzzy Mathematics, 2005, 13(3): 613~626
- 10 Slizak D, Ziarko W. The investigation of the Bayesian rough set model [J]. International Journal of Approximate Reasoning, 2005 (40):81~91

(上接第 205 页)

运行 5 次,设定聚类的数目为 10,最大迭代次数为 80。

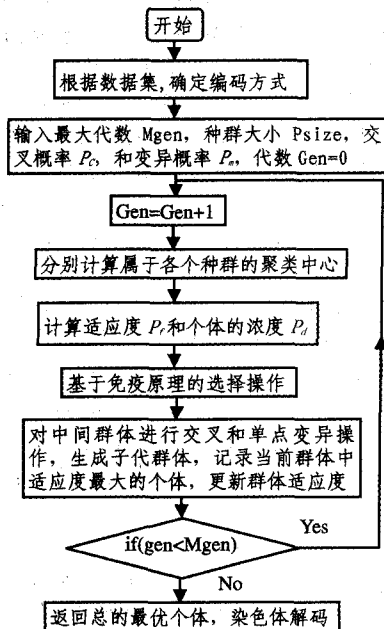


图 2 流程图

表 1 数据处理结果

运行次数	GA	MIGA
1	23	21
2	22	21
3	24	21
4	23	21
5	22	21

实验结果显示,本文提出的免疫遗传算法不论是在收敛速度上,还是聚类结果的稳定性上都优于传统的遗传算法。如果用该算法作用于较大的数据量,优越性更加明显。而将遗传算法与免疫算法相结合形成一种混合算法后,在保证准确性和提高执行效率的前提下,能更快地找到最有用的结果。

结论 本文将免疫原理引入遗传算法聚类中,通过对选择算子操作的优化,保证较高的准确率,又不至于陷入局部最优,克服了过早收敛的问题。MIGA 算法由于是降低了适应度的影响,因此会增加计算时间。如何通过对选择操作的参

速更好的设置,以及提升计算速度,是进一步研究的方向。

参考文献

- 1 Holland J H. Adaptation in Nature and Artificial System [M]. Cambridge: MIT Press, 1992
- 2 Krishma K, Murty M N. Genetic K-means Algorithm. IEEE Trans on System, Man, and Cybernetics, Part B, 1999, 29(3): 433~439
- 3 Yang Kiduk, Jacob E. A hybrid approach to generating and utilizing faceted vocabulary for knowledge discovery on the web [C]. In: Proc of the 2004 Joint ACM/IEEE Conference on Digital Libraries. New York: ACM Press, 2004. 395~395
- 4 高坚. 基于 C-均值和免疫遗传算法的聚类分析[J]. 计算机工程, 2003, 29(12): 65~66
- 5 李碧,雍正正. 一种改进的基于遗传算法的聚类分析方法[J]. 电路与系统学报, 2002, 7(3): 96~99
- 6 王小平,曹立明. 遗传算法[M]. 西安: 西安交通大学出版社, 2002. 1
- 7 郑日荣,毛宗源,等. 改进的人工免疫算法的分析研究[J]. 计算机工程与应用, 2003, 34: 35~37
- 8 Han Jiawei, Kamber M. Data Mining: concepts and techniques [M]. Jim Gray: Series Editor Morgan Kaufmann Publishers, 2000
- 9 Leclerc B. Minimum spanning trees for tree metrics: Abridgements and adjustments [J]. Journal of Classification, 1995, 12: 207~241
- 10 Selim S Z, Alsultan K. A Simulated Annealing Algorithm for the Clustering Problem [J]. Pattern Recognition, 1991, 24 (10): 1003~1008
- 11 Maulik U, Bandyopadhyay S. Genetic Algorithm-based Clustering Technique [J]. Patten Recognition, 2000, 33 (9): 1455~1465
- 12 行小帅,潘进,焦李成. 基于免疫规划的 K-means 聚类算法[J]. 计算机学报, 2003, 30(4): 150~152
- 13 傅景广,许刚,王裕国. 基于遗传算法的聚类分析[J]. 计算机工程, 2004, 30(4): 122~124
- 14 Fayyad U, Reina C, Bradley P S. Initialization of Iterative Refinement Clustering Algorithms [R]. [Microsoft Research Technical Report. MSR-TR-98-38]. June 1998. 1~5
- 15 王实,高文. 数据挖掘中的聚类算法[J]. 计算机科学, 2000, 27 (4): 42~45
- 16 杨淑莹. 图像模式识别[M]. 北京: 清华大学出版社, 北京交通大学出版社, 2005
- 17 陈莉,焦李成. 基于自适应聚类的数据预处理算法 I [J]. 计算机应用与软件, 2005, 22(3): 28~29
- 18 李春华,朱燕飞,毛宗源. 一种新型的自适应人工免疫算法[J]. 计算机工程与应用, 2004(22): 84~87
- 19 李洁,高新波,焦李成. 基于克隆算法的网络结构聚类新算法[J]. 电子学报, 2004, 32(7): 1195~1199