半P-集合($X^{\bar{F}},X$)与噪声数据剔除-应用

李豫颖

(宁德师范学院计算机与信息工程系 宁德 352100) (山东大学数学与系统科学学院 济南 250100)

摘 要 半P-集合(half packet sets)是由内P-集合$X^{\bar{F}}$(internal packet set $X^{\bar{F}}$)与有限普通集合X构成的集合对,或者($X^{\bar{F}},X$)是半P-集合,它具有内-动态特性。为了剔除噪声数据,获得目标数据,利用半P-集合提出了基于属性补充的递推-剔除噪声数据的方法。提出了噪声数据、噪声数据集成与$\bar{F}$-数据核概念;给出了噪声数据与$\bar{F}$-数据生成的递推方法与递推结构、噪声数据集成与$\bar{F}$-数据核关系定理、$\bar{F}$-数据依赖与辨识定理、噪声数据递推-剔除定理、噪声数据辨识准则与噪声数据递推-剔除准则,以及噪声数据递推-剔除应用。半P-集合是P-集合理论与应用的一个新的研究分支,是研究具有内-动态信息系统的一个新的数学方法。

关键词 半P-集合,噪声数据,噪声数据集成,$\bar{F}$-数据核,递推-剔除准则,应用

中图法分类号 TP274 **文献标识码** A

Half P-sets ($X^{\bar{F}},X$) and Noise Data Rejection-application

LI Yu-ying

(Department of Computer and Information Engineering, Ningde Normal University, Ningde 352100, China)

(School of Mathematics and System Sciences, Shandong University, Jinan 250100, China)

Abstract Half P-sets (half packet sets) are a set pair composed of internal P-set $X^{\bar{F}}$ (internal packet set $X^{\bar{F}}$) and finite general set X, or ($X^{\bar{F}},X$) is Half P-sets. It has internal dynamic characteristic. Using Half P-sets for rejecting noise data and getting target data, the recursion-rejection method of noise data by supplementing attributes was put forward in this paper. Some concepts were presented such as the noise data, the noise data integration, $\bar{F}$-data core. The recursion method and the recuision structure about the generation of the noise data and the generation of $\bar{F}$-data were given. The relation theorem between the noise data integration and $\bar{F}$-data core was given as well as the dependence and identification theorems of $\bar{F}$-data, the recursion-rejection theorems of the noise data. The identification criterion and the recursion-rejection criterion for the noise data were provided as well as the applicatin of recursion-rejection for the noise data. Half P-sets are not only a new study branch of theory and application about P-set but also a new mathematical method for researching information systems which have internal dynamic characteristic.

Keywords Half P-sets, Noise data, Noise data integration, $\bar{F}$-data core, Recursion-rejection criterion, Application

1 引言

文献[1]改进并简化P-集合(packet sets),提出半P-集合($X^{\bar{F}},X$)(half packet sets),给出半P-集合的结构,半P-集合具有内-动态特性。半P-集合是由内P-集合$X^{\bar{F}}$(internal packet set $X^{\bar{F}}$)与有限普通集合X构成的集合对,或者,($X^{\bar{F}}$, X)是半P-集合。半P-集合提出的背景是:在信息系统中,人们收集到信息$(x)=\{x_1,x_2,\cdots,x_m\}$,$\alpha=\{\alpha_1,\alpha_2,\cdots,\alpha_k\}$是$(x)$的属性集。在$(x)$中,人们要找到包含在$(x)$内的部分信息$(x)'=\{x_1,x_2,\cdots,x_n\}$,$n\leqslant m$,必须对属性集$\alpha$给予属性补充;或者,若$(x)'\subseteq(x)$,则$\alpha\subseteq\alpha^F$。其中:$\alpha^F$是$(x)'$的属性集,它是$\alpha$被补充部分属性之后生成的属性集。事实上,在$(x)$内找到$(x)'$是一种信息搜索的方法。显然,半P-集合($X^{\bar{F}},X$)是P-集合($X^{\bar{F}},X^F$)的一个部分。P-集合的结构、特征以及在信息系统多个领域中的应用,见文献[2-18]。

利用半P-集合,本文给出这样的讨论:设$(x)=\{x_1,x_2,\cdots,x_m\}$是被采集得到的数据,$\forall x_i\in(x)$是$(x)$的一个数据元。怎样把暂时无用的数据从$(x)$内剔除?换一个说法,数据$(x)=\{x_1,x_2,\cdots,x_m\}$内包含多个噪声(噪声元),用什么方法把这些噪声从$(x)$内剔除?如果$(x)=\{x_1,x_2,\cdots,x_m\}$中的$x_{p+1},x_{p+2},\cdots,x_m$被剔除,得到噪声数据$(x)^-=\{x_{p+1},x_{p+2},\cdots,x_m\}$与$\bar{F}$-数据$(x)^{\bar{F}}=\{x_1,x_2,\cdots,x_p\}$,满足$(x)^-=(x)-(x)^{\bar{F}}$,$(x)^{\bar{F}}\subseteq(x)$。显然,无用的数据,或者噪声数据从$(x)$内被剔除与半P-集合的内-动态特性相同。本文利用半P-集合的内-动态特性,给出信息系统的噪声数据剔除的研究,给出剔除定理、剔除方法与应用。本文利用新的数学模型:半P-集合,给出一种噪声数据剔除的新方法。半P-集合是P-集合理论与应用的一个新的研究分支。

为了方便讨论,这里把半P-集合的结构引入到第2节中,作为讨论的预备知识。

到稿日期:2010-09-11 返修日期:2010-12-09 本文受福建省自然科学基金(2009J01294),宁德师范学院科研重点项目(2008J002)资助。

李豫颖(1962-),女,副教授,CCF会员,主要研究方向为信息系统理论与应用,E-mail:asszlyy@163.com。

2 半 P-集合(X^F,X)的结构[1]

约定 U 是有限元素论域;V 是有限属性论域;X 是 U 上的有限非空普通集合;$F=\{f_1,f_2,\cdots,f_m\}$,$\overline{F}=\{\overline{f}_1,\overline{f}_2,\cdots,\overline{f}_n\}$是元素迁移族[2,3];$f\in F,\overline{f}\in\overline{F}$ 是元素迁移[2,3]。

给定有限普通集合 $X=\{x_1,x_2,\cdots,x_q\}\subset U$,$\alpha=\{\alpha_1,\alpha_2,\cdots,\alpha_k\}\subset V$ 是 X 的属性集合,称 X^F是 X 生成的内 P-集合[2,3](internal packet set X^F),简称 X^F是内 P-集合,而且

$$X^F=X-X^- \tag{1}$$

X^- 称作 X 的 $\overline{F}$-元素删除集合,而且

$$X^-=\{x|x\in X,\overline{f}(x)=u\overline{\in}X,\overline{f}\in\overline{F}\} \tag{2}$$

如果 X^F 的属性集合 α^F 满足

$$\alpha^F=\alpha\cup\{\alpha'|f(\beta)=\alpha'\in\alpha,f\in F\} \tag{3}$$

式中,$X^F\neq\varnothing$;$\beta\in V,\beta\overline{\in}\alpha,f\in F$ 把 β 变成 $f(\beta)=\alpha'\in\alpha$。

由内 P-集合与有限普通集合 X 构成的集合对,而且

$$(X^F,X) \tag{4}$$

称作 X 生成的半 P-集合(X^F,X)(half packet sets),有限普通集合 X 称作半 P-集合(X^F,X)的基集合(基础集,ground set)。

若在有限普通集合 X 的属性集 α 内,不断地补充属性;或者

$$\alpha\subseteq\alpha_1^F\subseteq\alpha_2^F\subseteq\cdots\subseteq\alpha_{n-1}^F\subseteq\alpha_n^F \tag{5}$$

则有

$$X_n^F\subseteq X_{n-1}^F\subseteq\cdots\subseteq X_2^F\subseteq X_1^F\subseteq X \tag{6}$$

由式(5),式(6)得到半 P-集合(X^F,X)的集合对族形式

$$\{(X_i^F,X)|i\in I\} \tag{7}$$

式中,I 是指标集。

3 噪声数据与 $\overline{F}$-数据的生成-辨识

约定 第 2 节中的 X^F,X,X^- 在第 3—5 节中,分别记作$(x)^F,(x),(x)^-$;或者$(x)^F=X^F,(x)=X,(x)^-=X^-$;而且,$(x)^F$在第 3—5 节的讨论中称作$(x)$的 $\overline{F}$-数据;$(x)=\{x_1,x_2,\cdots,x_q\}\subset U$,$\alpha=\{\alpha_1,\alpha_2,\cdots,\alpha_k\}\subset V$ 是(x)的属性集合,$\forall x_i\in(x)$称作(x)的数据元,$i=1,2,\cdots,q$。

定义 1 称$(x)^-=\{x_{p+1},x_{p+2},\cdots,x_q\}\subset U$ 是一个关于属性集 α 被补充属性 α_i'生成的噪声数据;称 $x_j\in(x)^-$是(x)的噪声元,简称噪声,$j=p+1,p+2,\cdots,q$,而且

$$(x)^-=(x)-(x)^F \tag{8}$$

如果 α 被补充属性 α'_i,生成的属性集 α^F 是 $\overline{F}$-数据$(x)^F$的属性集,满足

$$\alpha^F=\alpha\cup\{\alpha_i'|\beta_i\overline{\in}\alpha,\beta_i\in V,f(\beta_i)=\alpha_i'\in\alpha,f\in F\} \tag{9}$$

式中,$(x)^F=\{x_1,x_2,\cdots,x_p\}$,$p\leqslant q;p,q\in\mathbb{N}^+$;而且存在元素迁移 $\overline{f}\in\overline{F}$,使得

$$(x)^-=\{x|x\in(x),\overline{f}(x)=x'\overline{\in}(x),\overline{f}\in\overline{F}\} \tag{10}$$

定义 2 称$(x)_k^-$ 是(x)的 k 阶噪声数据,$k=1,2,\cdots,n$,$(x)_k^F$ 是数据(x)的 k 阶 $\overline{F}$-数据;如果$(x)_k^F$ 是$(x)_{k-1}^F$的 $\overline{F}$-数据,而且

$$\begin{cases}(x)_k^-=(x)_{k-1}^F-(x)_k^F, & k\neq1\\(x)_1^-=(x)-(x)_1^F, & k=1\end{cases} \tag{11}$$

定义 3 称$(x)_k^{-,*}$ 是噪声数据集$\{(x)_i^-|i=1,2,\cdots,k\}$关于$(x)$的 k 阶噪声数据集成,简称$(x)_k^{-,*}$ 是(x)的 k 阶噪声数据集成,如果

$$(x)_k^{-,*}=\bigcup_{i=1}^{k}(x)_i^- \tag{12}$$

定义 4 称$(x)_k^{F,*}$ 是 $\overline{F}$-数据集$\{(x)_i^F|i=1,2,\cdots,k\}$关于$(x)$的 k 阶 $\overline{F}$-数据核,简称$(x)_k^{F,*}$ 是(x)的 k 阶 $\overline{F}$-数据核,如果

$$(x)_k^{F,*}=\bigcap_{i=1}^{k}(x)_i^F \tag{13}$$

式中,$(x)_k^{F,*}\neq\varnothing$。

由定义 1—4 容易得到:

命题 1 补充的属性越多,被剔除的噪声数据越大,生成的 $\overline{F}$-数据越小。

命题 2 k 阶噪声数据$(x)_k^-$ 与 k 阶 $\overline{F}$-数据$(x)_k^F$ 是(x)的递推生成数据。

命题 3 噪声数据$(x)_k^-=\varnothing$,必有 $\mathrm{UNI}\{(x)_k^F,(x)_{k-1}^F\}$。

其中,UNI=unidentification。

定理 1($\overline{F}$-数据的序定理) 若$(x)_k^F$ 是(x)的 k 阶 $\overline{F}$-数据,则

$$(x)_k^F\subseteq(x)_{k-1}^F\subseteq\cdots\subseteq(x)_2^F\subseteq(x)_1^F\subseteq(x) \tag{14}$$

证明由定义 2 直接得到,证明略。

推论 1 k 阶 $\overline{F}$-数据$(x)_k^F$ 满足

$$\mathrm{IDE}\{(x)_k^F,(x)_{k-1}^F,\cdots,(x)_1^F\} \tag{15}$$

式中,IDE=identification。

推论 2 k 阶 $\overline{F}$-数据$(x)_k^F$ 满足

$$(x)_k^F\Rightarrow(x)_{k-1}^F\Rightarrow\cdots\Rightarrow(x)_2^F\Rightarrow(x)_1^F \tag{16}$$

式中,符号"$\Rightarrow$"是单依赖,"$\Rightarrow$"与"$\subseteq$"等价;$(x)_k^F\Rightarrow(x)_{k-1}^F$表示$(x)_{k-1}^F$单依赖于$(x)_k^F$。

定理 2(噪声数据集成与 $\overline{F}$-数据核的关系定理) 若$(x)_k^{-,*}$ 是(x)的 k 阶噪声数据集成,$(x)_k^{F,*}$ 是(x)的 k 阶 $\overline{F}$-数据核,则

$$(x)_k^{-,*}\cup(x)_k^{F,*}=(x) \tag{17}$$

证明:由定理 1 得到$(x)_k^{F,*}=\bigcap_{i=1}^{k}(x)_i^F=(x)_k^F$;又由式(11)得到:$(x)_{k-1}^F=(x)_k^-\cup(x)_k^F$,$(x)=(x)_1^-\cup(x)_1^F$;则

$$\begin{aligned}&(x)_k^{-,*}\cup(x)_k^{F,*}\\&=(\bigcup_{i=1}^{k}(x)_i^-)\cup(\bigcap_{i=1}^{k}(x)_i^F)=(\bigcup_{i=1}^{k}(x)_i^-)\cup(x)_k^F=(\bigcup_{i=1}^{k-1}(x)_i^-)\cup((x)_k^-\cup(x)_k^F)\\&=(\bigcup_{i=1}^{k-1}(x)_i^-)\cup(x)_{k-1}^F=(\bigcup_{i=1}^{k-2}(x)_i^-)\cup((x)_{k-1}^-\cup(x)_{k-1}^F)=(\bigcup_{i=1}^{k-2}(x)_i^-)\cup(x)_{k-2}^F\\&=\cdots=(\bigcup_{i=1}^{2}(x)_i^-)\cup(x)_2^F=(x)_1^-\cup((x)_2^-\cup(x)_2^F)=(x)_1^-\cup(x)_1^F=(x)\end{aligned}$$

定理 3($\overline{F}$-数据的属性单依赖定理) 若 $\overline{F}$-数据$(x)_i^F$,$(x)_j^F$,$(x)_k^F$ 的属性集分别是 $\alpha_i^F,\alpha_j^F,\alpha_k^F$,满足

$$\alpha_i^F\Rightarrow\alpha_j^F\Rightarrow\alpha_k^F \tag{18}$$

则
$$(x)_k^F\Rightarrow(x)_j^F\Rightarrow(x)_i^F \tag{19}$$

证明:$(x)_i^F,(x)_j^F,(x)_k^F$ 的属性集 $\alpha_i^F,\alpha_j^F,\alpha_k^F$ 满足 $\alpha_i^F\Rightarrow\alpha_j^F\Rightarrow\alpha_k^F$,由"$\Rightarrow$"与"$\subseteq$"等价,则 $\alpha_i^F\subseteq\alpha_j^F\subseteq\alpha_k^F$,所以,$(x)_k^F\subseteq(x)_j^F\subseteq(x)_i^F$,则$(x)_k^F\Rightarrow(x)_j^F\Rightarrow(x)_i^F$。

推论 3 若 $\overline{F}$-数据$(x)_i^F,(x)_j^F,(x)_k^F$ 的属性集 $\alpha_i^F,\alpha_j^F,\alpha_k^F$,满足 $\alpha_i^F\Rightarrow\alpha_j^F\Rightarrow\alpha_k^F$,则

$$\mathrm{IDE}\{(x)_i^F,(x)_j^F,(x)_k^F\} \tag{20}$$

定理 4($\overline{F}$-数据的属性双依赖辨识定理) 若 $\overline{F}$-数据

$(x)_i^F,(x)_j^F,(x)_k^F$ 的属性集分别是 $\alpha_i^F,\alpha_j^F,\alpha_k^F$，满足

$$\alpha_i^F \Leftrightarrow \alpha_j^F \Leftrightarrow \alpha_k^F \tag{21}$$

则 $$\mathrm{UNI}\{(x)_i^F,(x)_j^F,(x)_k^F\} \tag{22}$$

式中，符号“$\Leftrightarrow$”是双依赖，“$\Leftrightarrow$”与“$=$”等价；$\alpha_i^F \Leftrightarrow \alpha_j^F$ 表示，α_j^F 双依赖于 α_i^F。

证明与定理 3 类似，证明略。

定理 1—4 与推论 1—3 给出一个事实：递推-剔除噪声数据的过程，是一个属性依赖与数据依赖的过程；当在 $(x)_{k-1}^F$ 的属性集 α_{k-1}^F 内补充属性，得到属性集 α_k^F，满足 $\alpha_{k-1}^F \Leftrightarrow \alpha_k^F$ 时，则噪声数据 $(x)_{k-1}^-$ 与 $(x)_k^-$ 不可辨识，或者递推-剔除噪声数据的操作无效，需要重新补充属性。

噪声数据的辨识准则

若 $\bar{F}$-数据 $(x)_k^F$ 的属性集 α_k^F 与 $(x)_{k-1}^F$ 的属性集 α_{k-1}^F 满足

$$\alpha_{k-1}^F \Leftrightarrow \alpha_k^F \tag{23}$$

则 $$\mathrm{UNI}\{(x)_{k-1}^-,(x)_k^-\} \tag{24}$$

4 噪声数据度量与数据剔除-还原

定义 5 称 μ_k^- 是噪声数据 $(x)_k^-$ 关于 $(x)_{k-1}^F$ 的相对剔除度，简称 μ_k^- 是 $(x)_k^-$ 的相对剔除度，而且

$$\mu_k^- = \mathrm{card}((x)_k^-)/\mathrm{card}((x)_{k-1}^F) \tag{25}$$

定义 6 称 δ_k^- 是噪声数据集成 $(x)_k^{-,*}$ 关于 (x) 的绝对剔除度，简称 δ_k^- 是 $(x)_k^{-,*}$ 的绝对剔除度，而且

$$\delta_k^- = \mathrm{card}((x)_k^{-,*})/\mathrm{card}((x)) \tag{26}$$

由定义 5，定义 6 得到：

命题 4 噪声数据 $(x)_k^-$ 愈大，相对剔除度愈大；反之亦真。

命题 5 噪声数据集成愈大，绝对剔除度也愈大；反之亦真。

定理 5(相对剔除度单位离散区间内点定理) 若 μ_k^- 是 $(x)_k^-$ 的相对剔除度，$(x)_k^- \neq \varnothing$，则 μ_k^- 是单位离散区间 $(0,1]$ 的内点，或者

$$\mu_k^- \in (0,1) \tag{27}$$

式中，$(0,1]$ 是由数值 0 与自剔除度 $\mu=\mathrm{card}(x)/\mathrm{card}(x)=1$ 构成的单位离散区间。

证明：因为 μ_k^- 是 $(x)_{k-1}^F$ 的相对剔除度，由定义 5 得到 $\mu_k^- = \mathrm{card}((x)_k^-)/\mathrm{card}((x)_{k-1}^F)$；由 $(x)_k^- = (x)_{k-1}^F - (x)_k^F$，则 $(x)_k^- \subseteq (x)_{k-1}^F$；又 $(x)_k^- \neq \varnothing$，则 $(x)_k^- \subset (x)_{k-1}^F$。所以，$0<\mu_k^-<1$，或者 $\mu_k^- \in (0,1)$。

定理 6(绝对剔除度的序定理) 若 δ_i^- 是 $(x)_i^{-,*}$ 的绝对剔除度，$i=1,2,\cdots,k$，则

$$\delta_1^- \leqslant \delta_2^- \leqslant \cdots \leqslant \delta_{k-1}^- \leqslant \delta_k^- \tag{28}$$

证明由定义 3 与定义 6 直接得到，证明略。

定理 7(噪声数据属性剔除定理) (x) 的噪声数据 $(x)^-$ 从 (x) 内被剔除的充分必要条件是：(x) 的 $\bar{F}$-数据 $(x)^F$ 的属性集 α^F 与 (x) 的属性集 α 满足

$$\alpha^F = \alpha \cup \{\alpha'_i \mid \beta_i \overline{\in} \alpha, \beta_i \in V, f(\beta_i)=\alpha'_i \in \alpha, f \in F\} \tag{29}$$

证明：1°. 若 $\bar{F}$-数据 $(x)^F$ 的属性集 α^F 与 (x) 的属性集 α 满足 $\alpha^F = \alpha \cup \{\alpha'_i \mid \beta_i \overline{\in} \alpha, \beta_i \in V, f(\beta_i)=\alpha'_i \in \alpha, f \in F\}$，由定义 1 得到：$(x)^F=\{x_1,x_2,\cdots,x_p\}$，$(x)^-=\{x_{p+1},x_{p+2},\cdots,x_q\}$，满足 $(x)^-=(x)-(x)^F$，而且 $(x)^-=\{x \mid x\in(x), \bar{f}(x)=x' \overline{\in} (x), \bar{f}\in\bar{F}\}$ 表示数据 (x) 的噪声元通过元素迁移 $\bar{f}\in\bar{F}$ 从 (x) 内被剔除。2°. $(x)^-$ 从 (x) 内被剔除，则存在元素迁移 $\bar{f}\in\bar{F}$，使得 $(x)^-=\{x \mid x\in(x), \bar{f}(x)=x' \overline{\in} (x), \bar{f}\in\bar{F}\}$，不妨设 $(x)^-=\{x_{p+1},x_{p+2},\cdots,x_q\}$，则 $(x)^F=\{x_1,x_2,\cdots,x_p\}$，满足 $(x)^-=(x)-(x)^F$，由定义 1，$(x)^F$ 的属性集 α^F 与 (x) 的属性集 α 满足 $\alpha^F=\alpha\cup\{\alpha_i' \mid \beta_i \overline{\in} \alpha, \beta_i\in V, f(\beta_i)=\alpha'_i\in\alpha, f\in F\}$。

定理 8(噪声数据递推-剔除第一定理) $(x)_{k-1}^F$ 的噪声数据 $(x)_k^-$ 从 $(x)_{k-1}^F$ 内被剔除的充分必要条件是：$(x)_{k-1}^F$ 的 $\bar{F}$-数据 $(x)_k^F$ 的属性集 α_k^F 与 $(x)_{k-1}^F$ 的属性集 α_{k-1}^F 满足

$$\alpha_k^F = \alpha_{k-1}^F \cup \{\alpha'_j \mid \beta_j \overline{\in} \alpha_{k-1}^F, f(\beta_j)=\alpha'_j \in \alpha_{k-1}^F, f\in F\} \tag{30}$$

证明：1°. 若 $(x)_{k-1}^F$ 的 $\bar{F}$-数据 $(x)_k^F$ 的属性集 α_k^F 与 $(x)_{k-1}^F$ 的属性集 α_{k-1}^F 满足 $\alpha_k^F = \alpha_{k-1}^F \cup \{\alpha'_j \mid \beta_j \overline{\in} \alpha_{k-1}^F, f(\beta_j)=\alpha'_j \in \alpha_{k-1}^F, f\in F\}$，则 $(x)_k^F \subseteq (x)_{k-1}^F$；设 $(x)_k^- = (x)_{k-1}^F - (x)_k^F$，则 $(x)_k^-$ 从 $(x)_{k-1}^F$ 内被剔除。2°. 若 $(x)_k^-$ 从 $(x)_{k-1}^F$ 内被剔除，则 $(x)_k^- = (x)_{k-1}^F - (x)_k^F$，或者 $(x)_k^F \subseteq (x)_{k-1}^F$，则 $(x)_{k-1}^F$ 的 $\bar{F}$-数据 $(x)_k^F$ 的属性集 α_k^F 与 $(x)_{k-1}^F$ 的属性集 α_{k-1}^F 满足 $\alpha_{k-1}^F \subseteq \alpha_k^F$，或者 α_k^F 与 α_{k-1}^F 之间存在属性差集 $\{\alpha'_j \mid \beta_j \overline{\in} \alpha_{k-1}^F, f(\beta_j)=\alpha'_j \in \alpha_{k-1}^F, f\in F\}$，将其补充到 α_{k-1}^F 内，满足式(30)。

定理 9(噪声数据递推-剔除第二定理) 若 k 阶噪声数据集成 $(x)_k^{-,*}$ 与 $k-1$ 阶噪声数据集成 $(x)_{k-1}^{-,*}$ 满足

$$(x)_k^{-,*} - (x)_{k-1}^{-,*} \neq \varnothing \tag{31}$$

则噪声数据 $(x)_k^-$ 从 $(x)_{k-1}^F$ 内被剔除。

证明：若 k 阶噪声数据集成 $(x)_k^{-,*}$ 与 $k-1$ 阶噪声数据集成 $(x)_{k-1}^{-,*}$ 满足 $(x)_k^{-,*} - (x)_{k-1}^{-,*} \neq \varnothing$，则 $(x)_k^{-,*} - (x)_{k-1}^{-,*} = \bigcup_{i=1}^{k}(x)_i^- - \bigcup_{i=1}^{k-1}(x)_i^- = (x)_k^- \neq \varnothing$；由式(11)得到 $(x)_k^- = (x)_{k-1}^F - (x)_k^F \neq \varnothing$，则噪声数据 $(x)_k^-$ 从 $(x)_{k-1}^F$ 内被剔除。

定理 10($\bar{F}$-数据还原定理) 若 $(x)_k^F$ 是 k 阶 $\bar{F}$-数据，则 $(x)_k^F$ 被还原成 (x) 的充分必要条件是

$$(x)_k^F \cup \{x_i \mid x_i \in (x)_k^{-,*}, x_i \overline{\in} (x)_k^F, f(x_i)=x'_i \in (x)_k^F, f\in F\} = (x) \tag{32}$$

证明：1°. 若式(32)成立，$\{x_i \mid x_i \in (x)_k^{-,*}, x_i \overline{\in} (x)_k^F, f(x_i)=x_i' \in (x)_k^F, f\in F\}$ 表示从噪声数据集成 $(x)_k^{-,*}$ 中不断补充噪声 x_j 到 $(x)_k^F$ 中，直到 $(x)_k^F$ 与 (x) 相等，则 $(x)_k^F$ 被还原成 (x)。2°. 若 $(x)_k^F$ 被还原成 (x)，由式(17)有 $(x)_k^{-,*} \cup (x)_k^{F,*} = (x)$；又 $(x)_k^{F,*} = \bigcap_{i=1}^{k}(x)_i^F = (x)_k^F$，则 $(x)_k^{-,*} \cup (x)_k^F = (x)$，或者式(32)成立。

推论 4 若 $(x)_k^F$ 与 $(x)_{k-1}^F$ 满足

$$(x)_k^F \cup \{x_j \mid x_j \in (x)_k^-, x_j \overline{\in} (x)_k^F, f(x_j)=x'_j \in (x)_k^F, f\in F\} = (x)_{k-1}^F \tag{33}$$

则 $(x)_k^F$ 被还原成 $(x)_{k-1}^F$。

定理 10 与推论 4 给出了一种数据还原方法，事实上，定理 2 是数据还原的另一种方式。

噪声数据递推-剔除原理

在 (x) 的属性集 α 内补充属性，得到属性集 α_1^F，通过元素迁移 $\bar{f}\in\bar{F}$，噪声数据 $(x)_1^-$ 被剔除，生成 $\bar{F}$-数据 $(x)_1^F$；在 $(x)_1^F$ 的属性集 α_1^F 内补充属性，得到属性集 α_2^F，通过元素迁移 $\bar{f}\in\bar{F}$，噪声数据 $(x)_2^-$ 被剔除，生成 $\bar{F}$-数据 $(x)_2^F$；如此不断地操作，则噪声数据 $(x)_k^-$ 从 $(x)_{k-1}^F$ 内被剔除，得到 k 阶 $\bar{F}$-数据 $(x)_k^F$。

噪声数据递推-剔除准则

根据噪声数据递推-剔除原理，不断地进行递推-剔除噪声数据操作，直到绝对剔除度 $\delta_k^- \geqslant \varphi$ 为止。

$\varphi(0<\varphi<1)$是预设的噪声数据剔除阈值,噪声数据剔除操作终止时,得到(x)的$\overline{F}$-数据核$(x)^{\overline{F},*}$,将其称作(x)的目标数据。

5 噪声数据递推-剔除的应用

本节例子取自台风预警搜救系统P的实验。为了提高图像传输质量与显示效果,研究基于特征的图像处理,剔除图像中背景噪声,重点保留图像的动态信息。表1-表3是手机终端传输图像Q到系统控制中心,控制中心对图像Q进行噪声剔除,获得目标图像数据,并在监控中心的显示屏上显示。其中"-"表示空数据。

为了简单又不失一般性,取图像Q的样本数据$(x)=\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9\}$,$(x)$具有参数集$\alpha=\{\alpha_1,\alpha_2,\alpha_3\}$,见表1。这里:$x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9$的数值与$\alpha_1,\alpha_2,\alpha_3$的名称,略。

表1 系统P的Q图像数据(x)

α	α_1	α_2	α_3						
(x)	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9

表2中,系统检测到图像数据(x)内的x_3,x_5是加性噪声,则在(x)的参数$\alpha=\{\alpha_1,\alpha_2,\alpha_3\}$内补充参数$\alpha_1'$,得到参数集$\alpha_1^F=\{\alpha_1,\alpha_2,\alpha_3,\alpha_1'\}$,应用元素迁移$\overline{f}_1\in\overline{F}$(启动均值滤波模块),则$x_3,x_5$从$(x)$内被剔除,生成噪声数据$(x)_1^-=\{x_3,x_5\}$与$\overline{F}$-数据$(x)_1^F=\{x_1,x_2,x_4,x_6,x_7,x_8,x_9\}$,满足$(x)_1^-=(x)-(x)_1^F$。

表2 系统P的Q图像噪声数据$(x)_1^-$与$\overline{F}$-数据$(x)_1^F$

α_1^F	α_1	α_2	α_3	α_1'					
$(x)_1^-$	—	—	x_3	—	x_5	—	—	—	—
$(x)_1^F$	x_1	x_2	—	x_4	—	x_6	x_7	x_8	x_9

表3中,系统检测到图像数据(x)内的数据元x_6,x_7是椒盐噪声,则在$(x)_1^F$的参数集$\alpha_1^F=\{\alpha_1,\alpha_2,\alpha_3,\alpha_1'\}$内补充参数$\alpha_2'$,得到参数集$\alpha_2^F=\{\alpha_1,\alpha_2,\alpha_3,\alpha_1',\alpha_2'\}$,应用元素迁移$\overline{f}_2\in\overline{F}$(启动中值滤波模块),则$x_6,x_7$从$\overline{F}$-数据$(x)_1^F$内被剔除,生成噪声数据$(x)_2^-=\{x_6,x_7\}$与$\overline{F}$-数据$(x)_2^F=\{x_1,x_2,x_4,x_8,x_9\}$,满$(x)_2^-=(x)_1^F-(x)_2^F$。

表3 系统P的Q图像噪声数据$(x)_2^-$与$\overline{F}$-数据$(x)_2^F$

α_2^F	α_1	α_2	α_3	α_1'	α_2'				
$(x)_2^-$	—	—	—	—	—	x_6	x_7	—	—
$(x)_2^F$	x_1	x_2	—	x_4	—	—	—	x_8	x_9

因为$\alpha\Rightarrow\alpha_1^F\Rightarrow\alpha_2^F$,由定理3,则$(x)_2^F\Rightarrow(x)_1^F\Rightarrow(x)$;由推论3,IDE$\{(x),(x)_1^F,(x)_2^F\}$。根据噪声数据的辨识准则,$(x)_1^-$与$(x)_2^-$可辨识。由定义3得到:$(x)$的1阶与2阶噪声数据集成分别为$(x)_1^{-,*}=\{x_3,x_5\}$与$(x)_2^{-,*}=\{x_3,x_5,x_6,x_7\}$,由定义4,$(x)$的2阶$\overline{F}$-数据核$(x)_2^{\overline{F},*}=\{x_1,x_2,x_4,x_8,x_9\}$。由定义5,定义6得到:$(x)_1^-$与$(x)_2^-$的相对剔除度分别是$\mu_1^{\overline{F}}=0.22$,$\mu_2^{\overline{F}}=0.29$;噪声数据集成$(x)_1^{-,*}$,$(x)_2^{-,*}$的绝对剔除度分别是$\delta_1^-=0.22$,$\delta_2^-=0.44$。

若取噪声剔除阈值$\varphi=0.35$,由噪声数据递推-剔除准则,对图像Q第一次剔除噪声时,$\delta_1^-<\varphi$,继续剔除噪声;$\delta_2^->\varphi$,则停止剔除操作,得到$(x)_2^{\overline{F},*}$是系统需要的目标数据。

利用定理2,把噪声集成数据$(x)_2^{-,*}$补充到$\overline{F}$-数据核$(x)_2^{\overline{F},*}$内,满足$(x)_2^{\overline{F},*}\cup(x)_2^{-,*}=(x)$,则显示在屏幕上的图像数据被还原成初始数据$(x)$。利用定理10的推论4,将$(x)_2^-$中的噪声$x_6,x_7$补充到$(x)_2^F$中,则$(x)_2^F$被还原成$(x)_1^F$,再将$(x)_1^-$中的噪声$x_3,x_5$补充到$(x)_1^F$中,则$(x)_1^F$被还原成$(x)$。

上面的数据分析在实验中得到验证,$(x)_2^{\overline{F},*}$是系统P中图像Q的数据(x)经过递推-剔除噪声数据(噪声数据剔除阈值$\varphi=0.35$)后的图像数据。

结束语 半P-集合$(X^{\overline{F}},X)$是简化P-集合$(X^{\overline{F}},X^F)$得到的一个新的数学模型与数学结构,它的存在依赖于这样的事实:给定信息$(x)=\{x_1,x_2,\cdots,x_m\}$,$\alpha=\{\alpha_1,\alpha_2,\cdots,\alpha_k\}$是$(x)$的属性集;对$\alpha$内不断地补充属性,得到$\alpha_1^F\subseteq\alpha_2^F\subseteq\cdots\subseteq\alpha_{n-1}^F\subseteq\alpha_n^F$,$(x)_n^{\overline{F}}\subseteq(x)_{n-1}^{\overline{F}}\subseteq\cdots\subseteq(x)_2^{\overline{F}}\subseteq(x)_1^{\overline{F}}$;具有这个事实的信息系统被人们常常遇到。这个事实中包含的一个现象是:(x)的属性集不断变大,(x)不断变小;换句话说,依据属性集不断变大的过程,使人们能够在(x)内找到需要的信息$(x)^*$,$(x)^*\subseteq(x)$。如果把$(x)^o=(x)-(x)^*$定义成是在(x)内剔除的信息,显然,半P-集合能够应用到信息系统中噪声信息(数据)剔除研究中,在这个思想导引下,本文给出半P-集合与噪声数据剔除研究,给出应用。本文给出的讨论,仅是半P-集合的应用之一。人们容易看到:半P-集合具有内-动态特性;或者,$X_n^{\overline{F}}\subseteq X_{n-1}^{\overline{F}}\subseteq\cdots\subseteq X_2^{\overline{F}}\subseteq X_1^{\overline{F}}$。利用这个特性,在$(x)$内找到更小的$(x)'$,$(x)'$是$(x)$的核信息。本文给出的研究扩展,半P-集合还能应用于下列的应用研究领域中,这些领域的研究都是由半P-集合的内-动态特性得到的。

- 信息的内-搜索
- 信息的内-分类
- 信息的内-滤波

"滤波"概念来自通信系统与控制系统;若$p(x)$是由多个高次与低次谐波迭加生成的波,通过技术方法,从$p(x)$中滤掉高次谐波$p(x)'$,得到低次谐波$p(x)''$;或者$p(x)''$是通过对$p(x)$的滤波得到的。其中:$p(x)=p(x)'\oplus p(x)''$,"$\oplus$"表示迭加运算符。

显然,若给定信息(x),$\overline{F}$-信息$(x)^{\overline{F}}$,$(x)=\{x_1,x_2,\cdots,x_m\}$,$(x)^{\overline{F}}=\{x_1,x_2,\cdots,x_n\}$,$n\leqslant m$;$y=\{y_1,y_2,\cdots,y_m\}$是$(x)$的信息值集合,$y^{\overline{F}}=\{y_1,y_2,\cdots,y_n\}$是$(x)^{\overline{F}}$的信息值集合;利用$y$与$y^{\overline{F}}$,得到数据点:$(1,y_1),(2,y_2),\cdots,(m,y_m)$;$(1,y_1),(2,y_2),\cdots,(n,y_n)$。由

$$p(x)=\sum_{j=1}^{r}y_j\prod_{\substack{i,j=1\\i\neq j}}^{r}\frac{x-x_i}{x_j-x_i}\tag{34}$$

容易得到式(35),式(36),而且

$$p(x)=a_{m-1}x^{m-1}+a_{m-2}x^{m-2}+\cdots+a_1x+a_0\tag{35}$$

$$p(x)^{\overline{F}}=b_{n-1}x^{n-1}+b_{n-2}x^{n-2}+\cdots+b_1x+b_0\tag{36}$$

式(34)中,$r=m$,$r=n$。

把$p(x)$,$p(x)^{\overline{F}}$定义成两种波(函数);显然,对(x)的属性集α给予属性补充,$p(x)^{\overline{F}}$从$p(x)$内被过滤出来,或$p(x)^{\overline{F}}$通过"滤波"的方式,从$p(x)$中获得。

- 信息内-发现
- 信息内-隐藏
- 信息内-隐写

利用本文给出的简单讨论,人们或许能够看到,半P-集合$(X^{\overline{F}},X)$是研究具有内-动态特性信息系统的一个新的数学

方法。

参考文献

[1] 李豫颖,史开泉. 半 P-集合(X^F,X)与信息的内-真度环特征[J]. 计算机科学,2011,38(4):239-248

[2] Shi Kai-quan. P-sets and its applications[J]. An International Journal Advances in Systems Science and Applications,2009,9(2):209-219

[3] 史开泉. P-集合[J]. 山东大学学报:理学版,2008,43(11):77-84

[4] 史开泉. P-集合与它的应用特征[J]. 计算机科学,2010,37(8):1-8

[5] Shi Kai-quan, Li Xiu-hong. Camouflaged information identification and its applications[J]. International Journal Advances in Systems Science and Applications,2010,10(2):157-167

[6] 史开泉,张丽. P-集合与数据外-恢复[J]. 山东大学学报:理学版,2009,44(4):8-14

[7] 李豫颖,谢维奇,史开泉. $\bar{F}$-残缺数据的辨识与恢复[J]. 山东大学学报:理学版,2010,45(9):57-64

[8] 李豫颖. F-畸变数据的生成与修复[J]. 吉首大学学报:自然科学版,2010,31(3):59-72

[9] 于秀清. $P_{(\rho,\sigma)}$-集合与它的随机特性[J]. 计算机科学,2010,37(9):218-221

[10] 张丽,崔玉泉,史开泉. 外 P-集合与数据内-恢复[J]. 系统工程与电子技术,2010,32(6):1919-1924

[11] Li Yu-ying, Zhang Li, Shi Kai-quan. Generation and recovery of compressed data and redundant data[J]. Quantitative Logic and Soft Computing,2010,2(1):661-671

[12] Zhang Ling, Ren Xue-fang. P-sets and its $(f,\bar{f})$-heredity[J]. Quantitative Logic and Soft Computing,2010,2(1):735-742

[13] Qiu Yu-feng, Chen Bao-hui. f-Model generated by P-sets[J]. Quantitative Logic and Soft Computing,2010,2(1):613-620

[14] Xiu Ming, Shi Kai-quan, Zhang Li. P-sets and $\bar{F}$-data selection-discovery[J]. Quantitative Logic and Soft Computing,2010,2(1):791-799

[15] Zhang Li, Xiu Ming, Shi Kai-quani. P-sets and applications of power circle[J]. Quantitative Logic and Soft Computing,2010,2(1):581-591

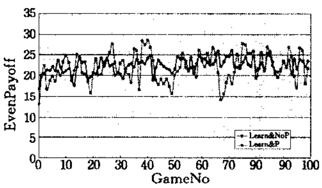
[16] Lin Hong-kang, Li Yu-ying. P-sets and its P-separation theorems[J]. An International Journal Advances in Systems Science and Applications,2010,10(2):209-215

[17] Huang Shun-liang, Wang Wei, Geng Dian-you. P-sets and its internal P-memory characteristics[J]. International Journal Advances in Systems Science and Applications,2010,10(2):216-222

[18] Wang Yang, Geng Hong-qin, Shi Kai-quan. The mining of dynamic information based on P-sets and its applications[J]. International Journal Advances in Systems Science and Applications,2010,10(2):234-240

[19] Zhang Guan-yu, Li En-zhong. Information gene and identification of its information Knock-out/Knock-in[J]. International Journal Advances in Systems Science and Applications,2010,10(2):308-315

(上接第 200 页)



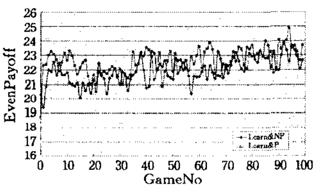


图 7 情绪对个体收益的影响 图 8 情绪对群体平均收益的影响

图 8 中表现的是情绪对所有个体的平均收益(群体平均收益)的影响。我们发现,从总体上看,情绪对群体平均收益的影响不大。群体平均收益波动略微增大,且在博弈后期总体呈上升趋势。从博弈多轮的平均收益来看,未引入情绪影响时群体平均收益为 22.03,引入情绪影响时平均收益为 22.24,相差较少。另外,从第 70 轮博弈开始,不论是引入情绪机制还是未引入情绪机制,群体的平均收益呈上升趋势,这说明个体学习机制开始发挥效用。

结束语 本文就情绪对具有学习特征的个体连续博弈过程进行了研究,得出情绪对个体博弈过程影响较大,而对群体博弈过程影响较小的结论。同时,引入学习特征的个体经过学习后,其收益情况也发生了较大变化,TFT 类型和 Tol 类型不再具有绝对优势,其它类型的个体收益也会逐渐提高,最终实现了群体收益的增长。但本文对情绪的分类较为简单,没有考虑不同情绪对个体的影响效果。另外,个体学习机制采取了较为单一的邻域学习,没有考虑个体之间的亲密度及合作的成功率。由于情绪和学习机制对个体博弈的影响研究是一个较新的研究领域[11,12],我们将在后期进一步展开研究。

参考文献

[1] Bernoulli D. Specimen Theoriae Novae de Measura Sortis(Exposition of a New Theory on the Measurement of Risk)[J]. Econometrica,1954,22(1):23-26

[2] 吴昊,杨梅英,陈良猷. 合作竞争博弈中的复杂性与演化均衡的稳定性分析 [J]. 系统工程理论与实践,2004(2):90-94

[3] 傅玉颖,潘晓弘,王正肖. 模糊合作博弈下的供应链多目标优化[J]. 浙江大学学报:工学版,2009(9):1644-1648

[4] Mellers B A, Schwartz A, Ritov I. Emotion-based Choice [J]. Journal of Experimental Psychology:General,1999,128:332-345

[5] Hastie R. Problems for Judgement and Decision Making [J]. Annual Review of Psychology,2001,52:653-683

[6] Cao Wen-ming, He Tian-cheng. The Multi-weight Neuron with Geometry Algorithm and Its Application[J]. Chinese of Journal Electronics,2008,17(2):261-264

[7] 周昌乐. 心脑计算举要[M]. 北京:清华大学出版社,2003

[8] Tankersley D, Stowe J C, Huettel S A. Altruism is associated with an increased neural response to agency [J]. Nature Neuroscience,2007(10):150-151

[9] Tetsushi O, Takao T. Cooperation in the Prisoner's Dilemma Game Based on the Second-Best Decision [J]. Journal of Artificial Societies and Social Simulation,2009,12(4):7

[10] 李静,陈蜀宇,吴长泽. 网格中一种小世界网络的分布式构造方法 [J]. 电子学报,2008,36(2). 413-416

[11] Meng Qing-mei, Wu Wei-guo. Artificial emotional model based on finite state machine [J]. Journal of Central South University of Technology,2008,15(5):694-699

[12] Nie Yong-you, Shan Xiao-wen, Bai Tao, et al. Evolutionary game analysis between the government and the waste producer in the venous industry [J]. Journal of Shanghai University (English Edition),2010,14(2):116-121