数据依赖与异常数据分离-应用

林宏康[1,2] 李豫颖[1,2] 阮群生[1]

(宁德师范学院计算机与信息工程系 宁德 352100)[1] (山东大学数学与系统科学学院 济南 250100)[2]

摘 要 数据在传递过程中,经常出现两类现象:一些被传递的数据在传递中发生部分数据元丢失;一些未知的数据元入侵到被传递的数据内。这两类现象使得被传递的数据出现"异常"。利用一个新的数学模型,给出两类现象的理论研究与应用。这个新的数学模型是P-集合(packet sets),P-集合是由内P-集合 X^F(internal packet set X^F)与外P-集合 X^F(outer packet set X^F)构成的集合对;或者,(X^F,X^F)是P-集合。给出数据的 $\overline{F}$-依赖、F-依赖的概念与特性,提出数据的依赖定理,给出异常数据被分离的应用。数据依赖是P-集合诸多应用特性之一。P-集合是研究动态数据系统的一个新理论与新方法。

关键词 P-集合,数据依赖,依赖度量,异常数据,数据分离,应用

中图法分类号 TP242 **文献标识码** A

Data Dependence and Separation-application of Abnormal Data

LIN Hong-kang[1,2] LI Yu-ying[1,2] RUAN Qun-Sheng[1]

(Department of Computer and Information Engineering, Ningde Normal University, Ningde 352100, China)[1]

(School of Mathematics and System Sciences, Shandong University, Jinan 250100, China)[2]

Abstract Two kinds of phenomenon often appear in data transference: some data elements are lost and some unknown data elements enter into original data. The data was changed into abnormal data due to such phenomena. By using a new mathematical model, theoretical research and application respect to two phenomena were given. This new model is called P-sets(packet sets). P-sets is a set pair which is composed by internal P-set X^F(internal packet set X^F) and outer P-set X^F(outer packet set X^F); or (X^F,X^F) is called P-sets. In this paper, the concepts and characteristics of $\overline{F}$-dependence and F-dependence were presented. The dependence theorem and the application of abnormal data separation were given. Data dependence is one of application characteristics and it is a new theory and method in dealing with dynamic data system.

Keywords P-sets, Data dependence, Dependence measure, Abnormal data, Data separation, Application

1 引言

利用数据传递网络 π,A 传递数据(x)(图像数据(x))给 B;因为网络 π 发生故障(或网络参数变化),B 接收到的数据出现两类现象:Ⅰ)被 B 接收到的数据是$(x)^F$;(x)与$(x)^F$满足$(x)^F\subseteq(x)$;换个说法,(x)内的部分数据元 $x_i\in(x)$在传递过程中丢失。Ⅱ)被 B 接收到的数据是$(x)^F$;(x)与$(x)^F$ 满足$(x)\subseteq(x)^F$;换个说法,(x)内增加了部分数据元 x_j(x_j 是一些(x)之外的未知数据元,x_j 入侵到(x)内,(x)变成$(x)^F$,$(x)\subseteq(x)^F$)。现象Ⅰ与Ⅱ在计算机应用领域与系统工程应用领域中经常遇到。$(x)^F$,$(x)^F$ 称作(x)的异常数据。(x)变成$(x)^F$等价于(x)依赖于$(x)^F$;或者,$(x)^F\Rightarrow(x)$。(x)变成$(x)^F$ 等价于$(x)^F$ 依赖于(x);或者$(x)\Rightarrow(x)^F$。名词"依赖"与符号"⇒"取自数理逻辑理论与推理理论。

显然,在现象Ⅰ与Ⅱ中,存在着一些有趣而重要的理论与应用,用什么方法能够把数据$(x)^F$,$(x)^F$ 从(x)中分离出来?数据$(x)^F$,$(x)^F$ 被分离成数据(x)具有什么特征?人们要求知道这些问题的答案。在能看到的P-集合应用文献中,没有人给出这些问题的理论研究与应用研究。

对于现象Ⅰ与Ⅱ,本文给出:A)人们怎样从数据(x)中发现数据$(x)^F$?或者,怎样把数据$(x)^F$从(x)中分离?B)人们怎样从数据(x)中发现数据$(x)^F$?或者,怎样把数据$(x)^F$ 从(x)中分离?本文利用一个新的数学模型P-集合(Packet sets)[1,2],给出问题 A 与 B 的讨论。这是因为P-集合的特性与问题 A,B 相似,P-集合具有依赖特性。P-集合是由内P-集合 X^F(internal packet set X^F)与外P-集合 X^F(outer packet set X^F)构成的集合对;或者(X^F,X^F)是P-集合。本文给出数据的依赖特性、依赖性定理以及异常数据的分离定理、异常数据的恢复定理,并给出应用。

为了便于讨论,保持本文的内容完整,容易接受本文给出

到稿日期:2010-06-11 返修日期:2010-09-25 本文受福建省自然科学基金(2009J01294),宁德师范学院科研资助重点项目(2008J002),宁德师范学院服务海西建设重点项目(2010H202)资助。

林宏康(1954—),男,副教授,主要研究方向为信息系统理论与应用,E-mail:lhk558@tom.com;**李豫颖**(1962—),女,副教授,主要研究方向为信息系统理论与应用;**阮群生**(1979—),男,硕士,讲师,主要研究方向为数据挖掘。

的结果，把 P-集合与它的结构简要地引入到本文的第 2 节中，作为本文的预备知识与理论准备。P-集合的更多概念与应用见文献[1-24]。

2 P-集合与它的特征

2008 年，文献[1,2]给出：

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

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

如果 X^F 的属性集合 α^F 满足

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

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

给定有限普通集合 $X=\{x_1,x_2,\cdots,x_m\}\subset U,\alpha=\{\alpha_1,\alpha_2,\cdots,\alpha_k\}\subset V$ 是 X 的属性集合；称 X^F 是 X 生成的外 P-集合(outer packet set)，简称 X^F 是外 P-集合，而且

$$X^F=X\cup X^+ \tag{4}$$

X^+ 称作 X 的 F-元素补充集合，而且

$$X^+=\{u\mid u\in U,u\overline{\in}X,f(u)=x'\in X,f\in F\} \tag{5}$$

如果 X^F 的属性集合 α^F 满足

$$\alpha^F=\alpha-\{\beta_i\mid\overline{f}(\alpha_i)=\beta_i\overline{\in}\alpha,\overline{f}\in\overline{F}\} \tag{6}$$

式中，$\alpha_i\in\alpha,\overline{f}\in\overline{F}$ 把 α_i 变成 $\overline{f}(\alpha_i)=\beta_i\overline{\in}\alpha,\alpha^F\neq\phi$。

其中，U 是非空有限元素论域，V 是非空有限属性论域。

由内 P-集合 X^F 与外 P-集合 X^F 构成的集合对称作普通集合 X 生成的 P-集合(Packet sets，P=Packet)，简称 P-集合，记作

$$(X^F,X^F) \tag{7}$$

普通集合 X 称作 (X^F,X^F) 的基集合(基础集，ground set)。

这里指出：

1)式(3)与计算机内存储器的结构 $T=T+1$ 相似。$T=T+1$ 具有动态性，式(3)也具有动态特性。如果用“静态”的观点认识 $T=T+1$，则 $T=T+1$ 不成立，因为 $2\neq2+1$。

2)式(2)，式(3)，式(5)，式(6)中的 $F=\{f_1,f_2,\cdots,f_m\}$，$\overline{F}=\{\overline{f}_1,\overline{f}_2,\cdots,\overline{f}_n\}$ 是元素迁移族[1-3]；$f\in F,\overline{f}\in\overline{F}$ 是元素迁移[1-3]，元素迁移 $f\in F,\overline{f}\in F$ 是一种给定的变换或者函数。

3)式(1)－式(3)中给出：X 内被删除部分元素，X 生成内 P-集合 X^F，等价于对 X 的属性集合 α 内补充新的属性，α 生成 α^F，$\alpha\subseteq\alpha^F$；或者，若 α_1^F,α_2^F 分别是 X_1^F,X_2^F 的属性集合，而且 $\alpha_1^F\subseteq\alpha_2^F$，则有 $X_2^F\subseteq X_1^F$。式(3)中的 $\{\alpha'\mid f(\beta)=\alpha'\in\alpha,f\in F\}$ 不是从 X 内被删除的元素构成的集合 X^- 的属性集合；或者 $\{\alpha'\mid f(\beta)=\alpha'\in\alpha,f\in F\}$ 不是 X^- 的属性集合。

由式(1)－式(7)给出的 P-集合结构，容易得到：

P-集合的集合对族结构

P-集合是由若干个集合对 (X_i^F,X_j^F) 构成的集合对族，而且

$$\{(X_i^F,X_j^F)\mid i\in I,j\in J\} \tag{8}$$

式中，I,J 为指标集(index set)。

特别指出：为了简单，又不失一般性，文献[1,2]及式(7)只利用了一个集合对 (X^F,X^F) 表示 P-集合。

P-集合的动态特性

给定有限普通集合 $X=\{x_1,x_2,\cdots,x_m\}\subset U,\alpha=\{\alpha_1,\alpha_2,\cdots,\alpha_k\}\subset V$ 是 X 的属性集合。如果在 α 内补充一些属性，同时在 α 内删除另外一些属性，则 α 分别变成 α_1^F 与 $\alpha_1^F,\alpha_1^F\neq\alpha_1^F$；$\alpha\subseteq\alpha_1^F,\alpha_1^F\subseteq\alpha$。由式(1)－式(7)得 P-集合 (X_1^F,X_1^F)。如果再在 α 内补充一些属性，同时又在 α 内删除另外一些属性，则 α 分别再变成 α_2^F 与 $\alpha_2^F,\alpha_2^F\neq\alpha_2^F;\alpha_1^F\subseteq\alpha_2^F,\alpha_2^F\subseteq\alpha_1^F$。由式(1)－式(7)得到 P-集合 (X_2^F,X_2^F)，如此等等。这些动态变化的一串集合对 (X_i^F,X_j^F) 构成式(8)。

P-集合的依赖特性

在式(1)－式(7)中，若 α_1^F,α_2^F 分别是 X_1^F,X_2^F 的属性集合，而且 $\alpha_1^F\subseteq\alpha_2^F$，则有 $X_2^F\subseteq X_1^F$，X_1^F 依赖于 X_2^F，而且

$$X_2^F\Rightarrow X_1^F \tag{9}$$

若 α_1^F,α_2^F 分别是 X_1^F,X_2^F 的属性集合，而且 $\alpha_2^F\subseteq\alpha_1^F$，则有 $X_1^F\subseteq X_2^F$，X_2^F 依赖于 X_1^F，而且

$$X_1^F\Rightarrow X_2^F \tag{10}$$

满足式(9)、式(10)的 P-集合 (X_2^F,X_2^F) 依赖于 P-集合 (X_1^F,X_1^F)，而且

$$(X_1^F,X_1^F)\Rightarrow(X_2^F,X_2^F) \tag{11}$$

式中，X_1^F,X_2^F 满足式(9)；X_1^F,X_2^F 满足式(10)；式(9)－式(11)中的“$\Rightarrow$”取自数理逻辑与推理理论。

显然，若 X_k^F 是 X 的内 P-集合，则 $X_k^F\Rightarrow X$；若 X_k^F 是 X 的外 P-集合，则 $X\Rightarrow X_k^F$。

再回到式(1)－式(7)中，容易得到：

定理 1(P-集合的还原定理) 给定 P-集合 (X^F,X^F) 与有限普通集合 X，若 $F=\overline{F}=\phi$，则

$$(X^F,X^F)_{F=\overline{F}=\phi}=X \tag{12}$$

事实上，若 $F=\phi$，则式(3)$\alpha^F=\alpha\cup\{\alpha'\mid f(\beta)=\alpha'\in\alpha,f\in F\}=\alpha$。这里 $\{\alpha'\mid f(\beta)=\alpha'\in\alpha,f\in F\}=\phi$；式(2)$X^-=\{x\mid x\in X,\overline{f}(x)=u\overline{\in}X,\overline{f}\in\overline{F}\}=\phi$，式(1)成为 $X^F=X-X^-=X$。若 $\overline{F}=\phi$，则式(6)$\alpha^F=\alpha-\{\beta_i\mid\overline{f}(\alpha_i)=\beta_i\overline{\in}\alpha,\overline{f}\in\overline{F}\}=\alpha$，这里 $\{\beta_i\mid\overline{f}(\alpha_i)=\beta_i\overline{\in}\alpha,\overline{f}\in\overline{F}\}=\phi$；式(5)$X^+=\{u\mid u\in U,u\overline{\in}X,f(u)=x'\in X,f\in F\}=\phi$，式(4)$X^F=X\cup X^+=X$。若 $F=\overline{F}=\phi$，则 $X^F=X,X^F=X$；式(12)成立。

定理 1 指出：在 $F=\overline{F}=\phi$ 的条件下，P-集合被还原成有限普通集合 X；P-集合 (X^F,X^F) 回到有限普通集合 X 的“原点”；换句话说，P-集合丢失了“动态特性”，P-集合 (X^F,X^F) 就是普通集合 X。

显然，若 $F=\overline{F}=\phi$，则式(8)变成

$$\{(X_i^F,X_j^F)\mid i\in I,j\in J\}_{F=\overline{F}=\phi}=X \tag{13}$$

式(13)指出：在 $\overline{F}=F=\phi$ 条件下，每个 X_i^F 与每个 X_j^F 都被还原成 X；$\{(X_i^F,X_j^F)\mid i\in I,j\in J\}$ 回到 X 的“原点”。

事实上，由定理 1，若 $\overline{F}=F=\phi$，则 $X^F=X^F;\forall i\in I,\forall j\in J$，若 $\overline{F}=F=\phi$，则 $X_i^F=X=X_j^F$，得到式(13)。

利用式(1)－式(13)的概念与给出的讨论，得到第 3 节。

3 数据依赖特性与依赖定理

约定 为了方便，第 1 节中的有限集合 X,X^F,X^F 分别用符号 $(x),(x)^F,(x)^F$ 表示，或者 $(x)=X,(x)^F=X^F,(x)^F=X^F$；符号 $(x),(x)^F,(x)^F$ 在第 3 节、第 4 节的讨论中使用。

定义 1 $(x)=\{x_1,x_2,\cdots,x_q\}\subset U$ 称作 U 上的一个数据，$x_i\in(x)$ 称作 (x) 的数据元，$i=1,2,\cdots,q$；如果 (x) 具有属

性集合 α，而且

$$\alpha=\{\alpha_1,\alpha_2,\cdots,\alpha_k\} \tag{14}$$

定义 2 $(x)^F=\{x_1,x_2,\cdots,x_p\}\subset U$ 称作(x)的一个 $\overline{F}$-数据，如果$(x)^F$的属性集α^F与(x)的属性集α满足

$$\alpha^F-\alpha\neq\phi \tag{15}$$

$(x)^F=\{x_1,x_2,\cdots,x_r\}\subset U$ 称作(x)的一个 F-数据，如果$(x)^F$的属性集α^F与(x)的属性集α满足

$$\alpha-\alpha^F\neq\phi \tag{16}$$

其中，定义 1、定义 2 中的 p,q,r 满足 $p\leqslant q\leqslant r,p,q,r\in\mathbb{N}^+$。

定义 3 称数据(x)单依赖于 $\overline{F}$-数据$(x)^F$，记作

$$(x)^F\Rightarrow(x) \tag{17}$$

称 F-数据$(x)^F$ 单依赖于数据(x)，记作

$$(x)\Rightarrow(x)^F \tag{18}$$

定义 4 称数据(x)双依赖于数据(x)，记作

$$(x)\Leftrightarrow(x) \tag{19}$$

式(17)－式(19)中的“⇒”与“⊆”等价；“⇔”与“=”等价。

由定义 1－定义 4 得到：

定理 2($\overline{F}$-数据单依赖定理) 若数据(x)单依赖于数据$(x)^*$，或者$(x)^*\Rightarrow(x)$，则

1)(x)的属性集α与$(x)^*$的属性集α^*满足

$$\alpha\Rightarrow\alpha^* \tag{20}$$

2)$(x)^*$是(x)的一个 $\overline{F}$-数据，而且$(x)^*=(x)^F$。

证明：1) 因为数据(x)单依赖数据$(x)^*$，或者$(x)^*\Rightarrow(x)$；或者$(x)^*=\{x_1,x_2,\cdots,x_p\}\subseteq\{x_1,x_2,\cdots,x_q\}=(x)$，由式(1)－式(3)，$(x)^*$的属性集$\alpha^*$与$(x)$的属性集$\alpha$满足$\alpha\subseteq\alpha^*$。由定义 3 得到$\alpha\Rightarrow\alpha^*$。2) 因为$\alpha\Rightarrow\alpha^*$，或者$\alpha\subseteq\alpha^*$，又因为$(x)$单依赖于$(x)^*$，或者$(x)^*\Rightarrow(x)$，或者$(x)^*\subseteq(x)$，由式(1)－式(3)，得到$(x)^*$是$(x)$的一个 $\overline{F}$-数据，$(x)^*=(x)^F$。

定理 3(F-数据单依赖定理) 若数据$(x)^\circ$单依赖数据(x)，或者$(x)\Rightarrow(x)^\circ$，则

1)(x)的属性集α与$(x)^\circ$的属性集α°满足

$$\alpha^\circ\Rightarrow\alpha \tag{21}$$

2)$(x)^\circ$是(x)的一个 F-数据，而且$(x)^\circ=(x)^F$。

证明与定理 2 类似，略。

定理 4($\overline{F}$-数据双依赖定理) $\overline{F}$-数据$(x)^F$双依赖于数据(x)的充分必要条件是$(x)^F$的属性集α^F与(x)的属性集α满足

$$\alpha^F-\{\alpha_i\mid\alpha_i\in\alpha^F,\overline{f}(\alpha_i)=\beta_i\overline{\in}\alpha^F,\overline{f}\in\overline{F}\}=\alpha \tag{22}$$

证明：1) 如果 $\overline{F}$-数据$(x)^F$双依赖于数据(x)，或者$(x)^F\Leftrightarrow(x)$，则$(x)^F$与(x)具有相同的属性集合。由式(1)－式(3)得到$(x)^F$与(x)的属性集α^F与α满足$\alpha\subseteq\alpha^F$。显然，存在属性差集$\nabla\alpha^F=\{\alpha_i\mid\alpha_i\in\alpha^F,\overline{f}(\alpha_i)=\beta_i\overline{\in}\alpha^F,\overline{f}\in\overline{F}\}$，把$\nabla\alpha^F$从$\alpha^F$内删除，或者$\alpha^F-\nabla\alpha^F=\alpha^F-\{\alpha_i\mid\alpha_i\in\alpha^F,\overline{f}(\alpha_i)=\beta_i\overline{\in}\alpha^F,\overline{f}\in\overline{F}\}=\alpha$。

2) 因为$(x)^F$的属性集α^F与(x)的属性集α满足$\alpha\subseteq\alpha^F$，或者α^F中存在属性集$\{\alpha_i\mid\alpha_i\in\alpha^F,\overline{f}(\alpha_i)=\beta_i\overline{\in}\alpha^F,\overline{f}\in\overline{F}\}=\nabla\alpha^F$。把$\nabla\alpha^F$从$\alpha^F$内删除，由式(22)$\alpha^F-\nabla\alpha^F=\alpha^F-\{\alpha_i\mid\alpha_i\in\alpha^F,\overline{f}(\alpha_i)=\beta_i\overline{\in}\alpha^F,\overline{f}\in\overline{F}\}=\alpha$。把$\nabla\alpha^F$从$\alpha^F$内删除，$(x)^F$与$(x)$具有相同的属性集合，$(x)^F\Leftrightarrow(x)$。

定理 5(F-数据双依赖定理) F-数据$(x)^F$双依赖于数据(x)的充分必要条件是$(x)^F$的属性集α^F与(x)的属性集α满足

$$\alpha^F\cup\{\beta_i\mid\beta_i\in V,\beta_i\overline{\in}\alpha^F,f(\beta_i)=\alpha_i\in\alpha^F,f\in F\}=\alpha \tag{23}$$

证明与定理 4 类似，略。

定理 6(数据单依赖的数据辨识定理) 若$(x)^F$，(x)，$(x)^F$满足

$$(x)^F\Rightarrow(x),(x)\Rightarrow(x)^F \tag{24}$$

则

$$\mathrm{IDE}\{(x)^F,(x),(x)^F\} \tag{25}$$

式中，IDE= identification[1,2]。

推论 1 若$(x)^F$，(x)，$(x)^F$满足$(x)^F\Leftrightarrow(x)\Leftrightarrow(x)^F$，则

$$\mathrm{UNI}\{(x)^F,(x),(x)^F\} \tag{26}$$

式中，UNI=unidentification[1,2]。

利用式(14)－式(26)给出第 4 节。

4 数据依赖度量与异常数据分离

定义 5 称θ^F是数据(x)单依赖于 $\overline{F}$-数据$(x)^F$的 $\overline{F}$-依赖度量，简称θ^F是$(x)^F$的 $\overline{F}$-依赖度量，而且

$$\theta^F=\mathrm{card}(\alpha^F)/\mathrm{card}(\alpha) \tag{27}$$

式中，(x)与$(x)^F$满足$(x)^F\Rightarrow(x)$；α,α^F分别是(x)与$(x)^F$的属性集，card=cardinal number。

定义 6 称θ^F是 F-数据$(x)^F$单依赖于数据(x)的 F-依赖度量，简称θ^F是$(x)^F$的 F-依赖度量，而且

$$\theta^F=\mathrm{card}(\alpha^F)/\mathrm{card}(\alpha) \tag{28}$$

式中，$(x)^F$与(x)满足$(x)\Rightarrow(x)^F$；α^F,α分别是$(x)^F$与(x)的属性集。

定义 7 称ρ是数据$(x)=\{x_1,x_2,\cdots,x_q\}\subset U$的模，而且

$$\rho=\|y\|/\|y\| \tag{29}$$

式中，$\|y\|=(y_1^2+y_2^2+\cdots+y_q^2)^{1/2}$是向量$y=(y_1,y_2,\cdots,y_q)^T$的 2-范数；$y=(y_1,y_2,\cdots,y_q)^T$是$x_i\in(x)$的特征值$y_i$($x_i$的数值)生成的向量，$i=1,2,\cdots,q$；$y=\{y_1,y_2,\cdots,y_q\}$是$(x)$的特征值集合。

定义 8 称ρ^F是数据$(x)^F=\{x_1,x_2,\cdots,x_p\}\subset U$的模，而且

$$\rho^F=\|y^F\|/\|y\| \tag{30}$$

称ρ^F是数据$(x)^F=\{x_1,x_2,\cdots,x_r\}\subset U$的模，而且

$$\rho^F=\|y^F\|/\|y\| \tag{31}$$

式中，$\|y^F\|=(y_1^2+y_2^2+\cdots+y_p^2)^{1/2}$是向量$y^F=(y_1,y_2,\cdots,y_p)^T$的 2-范数；$y^F=(y_1,y_2,\cdots,y_p)^T$是$x_j\in(x)^F$的特征值$y_j$生成的向量，$j=1,2,\cdots,p$。$\|y^F\|=(y_1^2+y_2^2+\cdots+y_r^2)^{1/2}$是向量$y^F=(y_1,y_2,\cdots,y_r)^T$的 2-范数；$y^F=(y_1,y_2,\cdots,y_r)^T$是$x_k\in(x)^F$的特征值$y_k$生成的向量，$k=1,2,\cdots,r$。$y^F=\{y_1,y_2,\cdots,y_p\}$是$(x)^F$的特征值集合；$y^F=\{y_1,y_2,\cdots,y_r\}$是$(x)^F$的特征值集合。

由定义 5－定义 8 直接得到：

命题 1 数据(x)的依赖度量θ与数据$(x)^*$的依赖度量θ^*满足$\theta^*-\theta\geqslant0$，$(x)^*$一定是$(x)$的一个 $\overline{F}$-数据，$(x)^*=(x)^F$；反之亦真。

命题 2 数据(x)的依赖度量θ与数据$(x)^\circ$的依赖度量θ°满足$\theta^\circ-\theta\leqslant0$，$(x)^\circ$一定是$(x)$的一个 F-数据，$(x)^\circ=(x)^F$；反之亦真。其中，$\theta=\mathrm{card}(\alpha)/\mathrm{card}(\alpha)$是$(x)$的依赖度量，$\alpha$是$(x)$的属性集。

定理 7(异常数据的第一分离定理) 若数据$(x)^*$的模ρ^*与数据(x)的模ρ满足

$$\rho^* - \rho \leqslant 0 \tag{32}$$

则

$$\mathrm{GRD}((x)^*) - \mathrm{GRD}((x)) \leqslant 0 \tag{33}$$

$(x)^*$从(x)内被分离-发现，$(x)^*$是(x)的一个$\bar{F}$-数据。其中，$\mathrm{GRD}((x)^*)=\mathrm{card}((x)^*)/\mathrm{card}((x))$是$(x)^*$的颗粒度，$\mathrm{GRD}((x))\doteq\mathrm{card}((x))/\mathrm{card}((x))$是$(x)$的颗粒度，GRD=granulation degree。

证明：因为ρ^*，ρ分别是数据$(x)^*$，(x)的模，而且满足$\rho^* - \rho \leqslant 0$，由式(29)、式(30)得到$y^* = \{y_1, y_2, \cdots, y_p\} \subseteq \{y_1, y_2, \cdots, y_q\} = y$；$y^*$，$y$分别是$(x)^*$，$(x)$的特征值集合。容易得到$(x)^* = \{x_1, x_2, \cdots, x_p\} \subseteq \{x_1, x_2, \cdots, x_q\} = (x)$，$\mathrm{GRD}((x)^*)=\mathrm{card}((x)^*)/\mathrm{card}((x)) \leqslant \mathrm{card}((x))/\mathrm{card}((x)) = \mathrm{GRD}((x))$，或者$\mathrm{GRD}((x)^*) - \mathrm{GRD}((x)) \leqslant 0$，$(x)^*$在$(x)$内被分离-发现。因为$(x)^* \subseteq (x)$，数据$(x)$单依赖于数据$(x)^*$，由定理2得到$(x)^*$是$(x)$的一个$\bar{F}$-数据，而且$(x)^* = (x)^F$。

定理7给出一个事实：因为(x)丢失了部分数据元x_i，(x)变小(或者$\mathrm{card}((x))$减少)，(x)变成$(x)^*$，$(x)^*$是(x)的一个异常数据(或者$(x)^F$是(x)的异常数据)，$(x)^*$潜藏在(x)内(或者$(x)^* \subset (x)$)，$(x)^*$从(x)内被分离。

定理8(异常数据的第二分离定理)　若数据$(x)^\circ$的模ρ°与数据(x)的模ρ满足

$$\rho^\circ - \rho \geqslant 0 \tag{34}$$

则

$$\mathrm{GRD}((x)^\circ) - \mathrm{GRD}((x)) \geqslant 0 \tag{35}$$

$(x)^\circ$从(x)外被分离-发现，$(x)^\circ$是(x)的一个F-数据。

定理8的证明与定理7的证明类似，略。

定理8给出一个事实：因为部分数据元x_j入侵到数据(x)内，(x)变大(或者$\mathrm{card}((x))$增大)，(x)变成$(x)^\circ$，$(x)^\circ$是(x)的一个异常数据(或者$(x)^F$是(x)的异常数据)，$(x)^\circ$潜藏在(x)外(或者$(x) \subset (x)^\circ$)，$(x)^\circ$从(x)外被分离。

由定理7、定理8直接得到：

定理9(异常数据的第一恢复定理)　异常数据$(x)^*$被恢复成数据(x)的充分必要条件是$(x)^*$的属性集α^F与数据(x)的属性集α满足

$$\alpha^F - \alpha = \phi \tag{36}$$

事实上，由式(1)－式(3)得到$(x)^*$的属性集$\alpha^F = \alpha \cup \{\alpha' \mid f(\beta) = \alpha' \in \alpha, f \in F\}$；或者$(x)^*$的属性集$\alpha^F$与$(x)$的属性集$\alpha$满足$\alpha \subseteq \alpha^F$。因此，若$\{\alpha' \mid f(\beta) = \alpha' \in \alpha, f \in F\} = \phi$，则$\alpha^F = \alpha$，或者$\alpha^F - \alpha = \phi$，$(x)^*$与$(x)$具有相同的属性集。由式(29)、式(30)得到$(x)^*$与$(x)$具有相等的模$\rho = \rho^*$；或者$(x) = (x)^*$，$(x)^*$被恢复成$(x)$。定理的证明略。

定理10(异常数据的第二恢复定理)　异常数据$(x)^\circ$被恢复成数据(x)的充分必要条件是$(x)^\circ$的属性集α^F与数据(x)的属性集α满足

$$\alpha - \alpha^F = \phi \tag{37}$$

证明由式(4)－式(6)与式(29)、式(31)得到，略。

定理9给出一个工程应用事实：在引言中的数据传输网络π(计算机视觉-识别系统的数据传输网络)中，不存在网络器件老化与器件失效现象。

定理10给出一个工程应用事实：在引言中的数据传输网络π(计算机视觉-识别系统的数据传输网络)中，存在数据之间的耦合(数据入侵)现象；若消除这个现象，则需要修正网络中的“滤波”参数。

其中，定理9与定理10给出的工程应用事实，在我们的实验中得到了认证。

利用第2节中的概念与第3，4节中给出的讨论与结果，给出第5节。

5　数据依赖与异常数据分离的应用

约定　$y = \{y_1, y_2, \cdots, y_q\}$，$y^F = \{y_1, y_2, \cdots, y_p\}$，$y^F = \{y_1, y_2, \cdots, y_r\}$在本节中分别称作$(x)$，$(x)^F$，$(x)^F$的特征值集合。为了简单，又不产生误解，$y$，$y^F$，$y^F$在这一节中称作数据。

本节的例子取自计算机视觉-识别系统。为了简单，又不失一般性，取这个系统的一个子系统，图1给出了这个子系统的简化框图。

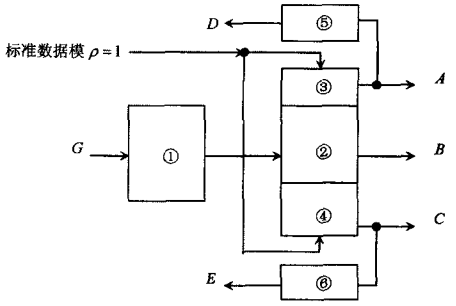

图1　计算机视觉-识别系统的一个子系统

子系统图1的工作过程：若子系统处于正常工作状态，系统的数据$y = \{y_1, y_2, \cdots, y_q\}$由$G$端进入①，在①中$y$生成$\rho = \|y\| / \|y\| = 1$；$\rho$与来自标准的$\rho$比较，数据$y$进入②，②给出的输出$B$=“1”；$A$=“1”，$D$=“0”，$C$=“1”，$E$=“0”。$B$=“1”表示输出$B$给出标准数据$y = \{y_1, y_2, \cdots, y_q\}$。系统的异常状态(数据被丢失状态)Ⅰ)因为$G$端进入①的数据$y$存在数据元$y_i$丢失，标准数据$y = \{y_1, y_2, \cdots, y_q\}$变成$y^F = \{y_1, y_2, \cdots, y_p\}$，$p \leqslant q$；$y^F$进入①生成$\rho^F = \|y^F\| / \|y\| \leqslant \|y\| / \|y\| = \rho = 1$，或者$\rho^F \leqslant \rho = 1$，③输出“0”；⑤输出$D$=“1”，给出预警；输出$A$=“0”而且输出$B$=“0”；$B$=“0”表示输出$B$给出异常数据$y^F$。系统的异常状态(外来数据入侵状态)Ⅱ)因为$G$端进入①的数据$y$存在数据$y_k'$的插入(外来数据$y_k'$入侵)，标准数据$y = \{y_1, y_2, \cdots, y_q\}$变成$y^F = \{y_1, y_2, \cdots, y_r\}$，$q \leqslant r$；$y^F$进入①生成$\rho^F = \|y^F\| / \|y\| \geqslant \|y\| / \|y\| = \rho = 1$，或者$\rho^F \geqslant \rho = 1$；④输出“0”，⑥输出$E$=“1”，给出预警；输出$C$=“0”而且输出$B$=“0”；$B$=“0”表示输出$B$给出异常数据$y^F$。状态Ⅰ与状态Ⅱ表示系统的异常状态(数据$y^F$或$y^F$)从正常状态(标准数据$y$)中被分离-辨识。这里$G$是系统的输入，$B$是子系统的输出；“0”，“1”是$B$的状态“逻辑值”。

图1中，①是数据模ρ^F，ρ^F生成模块；②是数据模ρ^F，ρ^F比较模块；③，④是标准数据模$\rho = 1$存储模块；⑤是数据模ρ^F的预警模块；⑥是数据模ρ^F的预警模块。

为方便例子的讨论，这里给出：

异常数据分离-辨识准则

数据y^*的模ρ^*($\rho^* = \rho^F$或$\rho^* = \rho^F$)与数据y的模ρ满足

$$\mathrm{IDE}(\rho^*, \rho) \tag{38}$$

系统输出数据被分离，系统输出“逻辑值”=“0”。其中，IDE=identification。

取 2009-10-21 某计算机视觉-辨识系统的实验数据，这些数据列入表 1 中。

表 1　计算机视觉-辨识系统输出的标准数据

y_1	y_2	y_3	y_4	y_5	y_6	y_7
1.83	1.92	1.67	1.04	1.16	1.52	1.43

表 1 中，$y_1, y_2, y_3, y_4, y_5, y_6, y_7$ 是图 1 中 B 的输出数据 $y=\{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$。表 1 给出的是实验数据经过技术方法后的数据，它不影响本节中例子的分析与异常数据分离-辨识的讨论。表 1 中的数据对应的视觉-辨识系统给出的图像符合要求(失真度=0.03)，视觉-辨识系统能正确地辨识运动着的图像(图像监视)；或者，系统输出 $B=$"1"(见图 1)，满足式(29)：$\rho=1$；图中，$A=$"1"，$D=$"0"，$C=$"1"，$E=$"0"。$B=$"1"表示 B 给出标准数据 $y=\{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$，如表 1 所列。

为了把这个系统应用于某案件并准确地侦破，考核这个系统的实用性与稳定性，我们给出一个实验(状态Ⅰ)：调整图 1 中模块①的参数，使表 1 中的数据丢失，得到表 2。

表 2　存在数据丢失的计算机视觉-辨识系统数据

y_1	y_2	y_3	y_4	y_5	y_6	y_7
1.83	1.92	—	1.04	1.16	1.52	—

表 2 中的"—"表示零数据。表 2 中的数据对应的视觉-辨识系统给出的图像发生变形(失真度=0.87)，系统输出 $B=$"0"(见图 1)，满足式(32)：$\rho^*=3.43/4.08=0.84\leqslant 1$。图 1 中 $A=$"0"，$C=$"0"，$D=$"1"，$E=$"1"，$B=$"0"表示输出 B 给出异常数据 $\{y_1, y_2, y_4, y_5, y_6\}=y^F$，如表 2 所列。

这里特别说明：调整图 1 中模块①的参数，在实际系统中是增加了一个"滤波器"。增加"滤波器"等价于式(14)α 内补充属性。显然，若式(14)α 是 $y=\{y_1, y_2, \cdots, y_q\}$ 的属性集，α^F 是 $y^F=\{y_1, y_2, \cdots, y_p\}$ 的属性集，$\alpha\subseteq\alpha^F$，则 $y^F\subseteq y$；α^F 是由 α 被补充属性得到的。

利用表 1、表 2 得到分析结论：由式(27)得到 $\theta^F=\mathrm{card}(\alpha^F)/\mathrm{card}(\alpha)\geqslant 1$，由定理 2 得到 $y^F\Rightarrow y$；由定理 7 知 y^F 是 y 的异常数据，而且 $\rho^F\leqslant\rho$；利用式(38)得到异常数据 y^F 从标准数据 y 中被分离-辨识。

另外一个实验(状态Ⅱ)略。

系统的实际认证

本文给出的计算机视觉-辨识系统(包括图 1 给出的子系统)在某侦察系统中得到了实际应用；在侦破某案件中，这个系统提供了准确的图像定位信息。

结束语　异常数据(系统输出数据的无规则变化)在信息系统、计算机控制系统与计算机信息识别系统中经常遇到，人们对异常数据只给出感性认识，对它的理性认识却少见。在计算机数据传输系统中，被传输的数据是 $(x)=\{x_1, x_2, \cdots, x_q\}$，人们得到的数据却是 $(x)^F=\{x_1, x_2, \cdots, x_q\}$，$p\leqslant q$；或者 $(x)^F=\{x_1, x_2, \cdots, x_r\}$，$q\leqslant r$；$p, q, r\in\mathbb{N}^+$。事实上，$(x)$ 变成 $(x)^F$，$(x)^F\subseteq(x)$，(x) 变成 $(x)^F$，$(x)\subseteq(x)^F$ 等是由 (x) 的属性集 α 的变化引起的，这个事实却被人们长期忽略，未引起注意。本文把被人们忽略的事实找回来，并以 P-集合为依托，对异常数据($(x)^F$ 或 $(x)^F$)给出讨论，给出应用。应用例子取自计算机视觉-辨识系统。本文给出的讨论与结果，可以扩展到计算机数据库系统及计算机控制系统中，能够得到一些新的研究。本文给出的数据依赖与异常数据分离在计算机视觉-辨识系统的应用，仅是 P-集合在动态信息系统的众多应用之一。P-集合是研究动态信息系统的一个新的数学理论与数学方法。

参 考 文 献

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

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

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

[4] Shi Kaiquan, Li Xiuhong. Camouflaged information identification and its applications[J]. An International Journal Advances in Systems Science and Applications, 2010, 10(2): 157-167

[5] 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

[6] 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

[7] 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

[8] 张冠宇，周厚勇，史开泉. P-集合与双 P-数据恢复-辨识[J]. 系统工程与电子技术, 2010, 32(9): 1233-1238

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

[10] Li Yuying, Zhang Li, Shi Kaiquan. Generation and recovery of compressed data and redundant data[J]. Quantitative Logic and Soft Computing, 2010, 2(1): 661-671

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

[12] Qui Yufeng, Chen Baohui. f-model generated by P-sets [J]. Quanti tative Logic and Soft Computing, 2010, 2(1): 613-620

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

[14] Lin Hongkang, Li Yuying. P-sets and its P-separation theorems [J]. An International Journal Advances in Systems Science and Applications, 2010, 10(2): 209-215

[15] Huang Shunliang, Wang Wei, Geng Dianyou. P-sets and its internal P-memory characteristics[J]. An International Journal Advances in Systems Science and Applications, 2010, 10(2): 216-222

[16] 周玉华，张冠宇，史开泉. P-集合与双信息规律生成[J]. 数学的实践与认识, 2010, 40(13): 71-80

[17] 周玉华，张冠宇，张丽. 内-外数据圆与动态数据恢复[J]. 山东大学学报：理学版, 2010, 45(8): 21-26

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

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

[20] Zhang Li, Cui Yuquan. Outer P-sets and data internal-recovery [J]. An International Journal Advances in Systems Science and Applications, 2010, 10(2): 189-199

[21] 张飞，陈萍，张丽. P-集合的 P-分离与应用[J]. 山东大学学报：理学版, 2010, 45(3): 71-75

[22] 于秀清. P-集合的识别与筛选[J]. 山东大学学报：理学版, 2010, 45(1): 94-98

[23] 汤积华，陈保会，史开泉. P-集合与$(\bar{F}, F)$-数据生成-辨识[J]. 山东大学学报：理学版, 2009, 44(11): 83-92

[24] 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