

Journal of Software Engineering

ISSN 1819-4311



www.academicjournals.com

Journal of Software Engineering

ISSN 1819-4311 DOI: 10.3923/jse.2017.32.39



Research Article Dependent Characteristics of Data and Separation of (F, F)-data

¹Li Yuying, ²Lin Qifa, ³Xie Weiqi and ⁴Li Jintong

¹Department of Computer, Ningde Normal University, 352100 Ningde City, China ²Department of Mathematics, Ningde Normal University, 352100 Ningde City, China ³School of Computer Engineering, Jinling Institute of Technology, 211169 Nanjing City, China ⁴Department of Library, Ningde Normal University, 352100 Ningde City, China

Abstract

Background: In data transmission, system often appears abnormity as outputting data of the system is changed. **Materials and Methods:** In order to obtain the method of identification or separation for abnormal data, the paper proposes (\overline{F} , F) data and data dependence through using P-sets that is a new mathematic tool to research the dynamic system. The (\overline{F} , F)-data consists of \overline{F} -data and F-data. The \overline{F} -data, F-data and (\overline{F} , F)-data are all abnormal. Also this paper gives dependence characteristics as well as dependence-identification theorems of (\overline{F} , F)-data. **Results:** By defining the module of data, the paper provides the method of separating-recovery (\overline{F} , F)-data and the criterion of separating abnormal data. Finally, the application of separating abnormal (\overline{F} , F)-data is provided. **Conclusion:** The proposed method of separating abnormal data can be applied to data mining, data filtration and data discovery, fault detection and ruling out and so on.

Key words: (F, F)-data, data dependence characteristics, identification, separation, P-sets

Received: May 13, 2016

Accepted: September 12, 2016

Published: December 15, 2016

Citation: Li Yuying, Lin Qifa, Xie Weiqi and Li Jintong, 2017. Dependent characteristics of data and separation of (F, F)-data. J. Software Eng., 11: 32-39.

Corresponding Author: Li Yuying, Department of Computer, Ningde Normal University, 352100 Ningde City, China

Copyright: © 2017 Li Yuying *et al.* This is an open access article distributed under the terms of the creative commons attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

In data transmission of a computer network Ω , terminal A passes data $(x) = \{x_1, x_2, ..., x_q\}$ to another terminal B. Because the external environment or the system parameter about the network is changed, the data received for the terminal B appears 2 types of phenomena: (I) The terminal B receives data $(x)^{\overline{F}}$ at time t_1 . $(x)^{\overline{F}}$ and (x) satisfy $(x)^{\overline{F}} \subseteq (x)$. Namely, some data elements $x_i \in (x)$ are lost in transmission. (II) The terminal B receives data $(x)^{F}$ at time t_{2} . (x) and $(x)^{F}$ satisfy $(x) \subseteq (x)^{F}$. Namely, some data elements $x_i \in (x)$ are added into (x) or (x) is invaded by x_i . The (x) changes into $(x)^F$ and $(x)\subseteq (x)^F$. Phenomena I and II often appear in the fields of computer applications and engineering applications. That (x) is changed into $(x)^{\overline{F}}$ is equivalent to that (x) depends on $(x)^{\overline{F}}$ or $(x)^{\overline{F}} \rightarrow (x)$. That (x) is changed into $(x)^{F}$ is equivalent to that $(x)^{F}$ depends on (x) or $(x) \rightarrow (x)^F$. Here, "Dependence" symbol " \rightarrow " comes from mathematical logic theory and reasoning theory.

For the phenomena of I and II, there are some interesting and important theories and applications. The research studies the method of discovering or separating data $(x)^{\overline{r}}$ or $(x)^{F}$ from (x) and studies the characteristics that $(x)^{\overline{r}}$ and $(x)^{F}$ are separated from (x).

This study uses a new mathematical model P-sets^{1,2} to discuss above problems because the characteristics of P-sets is similar to phenomena and P-sets has dependent characteristics. The P-sets are a set pair that comprises internal P-set $X^{\overline{F}}$ and outer P-set $x^{\overline{F}}$. P-sets has dynamic characteristics¹⁻⁸. P-sets achieved many applications³⁻²⁸. The paper proposes (\overline{F} , \overline{F})-data and data dependence and gives dependence characteristics as well as dependence and identification theorems of (\overline{F} , \overline{F})-data. By defining the module of data, the study provides the method of separating and recovery (\overline{F} , \overline{F})-data and the application of separating of abnormal (\overline{F} , \overline{F})-data.

To discuss conveniently and easily accept following discussion of the study, the structure of P-sets is introduced in section 2.

MATERIALS AND METHODS

Structure and characteristics of P-sets^{1,2} Structure of P-sets

Assumption 1: U is a finite element universe and V is a finite attribute universe. $X = (x_1, x_2, ..., x_m) \subset U$ is a given general set and its attribute set is $\alpha = (\alpha_1, \alpha_2, ..., \alpha_k) \subset V$. The $X^{\overline{F}}$ is called an internal packet set of x (called an internal P-set for short) and X^- is called an \overline{F} -element deleted set of X and:

$$\mathbf{X}^{\overline{\mathbf{F}}} = \mathbf{X} \cdot \mathbf{X}^{-} \tag{1}$$

 X^- is called an \overline{F} -element deleted set of X and:

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

If attribute set α^{F} of X^{F} and attribute set α of X satisfy:

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

where, $\beta \in V$, $\beta \in \alpha$, $f \in F$, turns β into $f(\beta) = \alpha' \in \alpha$.

 $X = \{x_1, x_2, ..., x_m\} \subset U$ is a given general set and its attribute set is $\alpha = \{\alpha_1, \alpha_2, ..., \alpha_k\} \subset V$. The X^F is called an outer packet set of X (called an outer P-set for short) and X^+ is called an F-element supplemented set and:

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

X⁺ is called an F-element supplemented set and:

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

If the attribute set $\alpha^{\overline{F}}$ of X^F satisfies:

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

The set pair composed of the internal P-set X^F and the outer P-set X^F is called packet sets (called P-sets for short), denoted by:

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

where, X is base set of $(X^{\overline{F}}, X^{\overline{F}})$.

The descriptions about P-sets are as follows:

- The characteristics of the Eq. 3 is similar to T = T+1 of computer memory. The T = T+1 has iterative and dynamic characteristics
- In Eq. 2, 3, 5, 6 and F = {f₁, f₂,..., f_n} are element transfer families¹⁻⁸, f∈F and f∈F are element transfers¹⁻⁸. The f∈F and f∈F are given functions. The characteristics of f∈F is that for u∈U and u∈X, f∈F changes u into f(u) = x'∈X. Or for β∈V and β∈α, f∈F changes β into f(β) = α'∈α. The characteristics of f∈F is that for x∈X, f∈F changes x into f(x) = u∈X. Or for α_i∈α, f∈F changes α_i into f(α_i) = β_i∈α
- In the Eq. 3, $\{\alpha'|f(\beta) = \alpha' \in \alpha, f \in F\}$ is attributes set. It consists of new attributes elements replenished to α . It and α satisfy that $\{\alpha'|f(\beta) = \alpha' \in \alpha, f \in F\} \cap \alpha = \emptyset$. For example, $\alpha = \{\alpha_1, \alpha_2, \alpha_3\}, \{\alpha'|f(\beta) = \alpha' \in \alpha, f \in F\} = \{\alpha'_1, \alpha'_2, \alpha'_3\}, \alpha_k \neq \alpha'_k, k = 1, 2, 3$. Apparently α and $\{\alpha'|f(\beta) = \alpha' \in \alpha, f \in F\}$ have no common elements (attributes)

• The Eq. 1-3 indicate that some elements are deleted from X at same time the internal P-sets $X^{\overline{F}}$ generated by X. Namely, the attribute set α of X is added by new attribute elements. So, α generates α^{F} and $\alpha \subseteq \alpha^{F}$. Namely, if α_{1}^{F} and α_{2}^{F} are attribute sets of $X_{1}^{\overline{F}}$ and $X_{2}^{\overline{F}}$, respectively and $\alpha_{1}^{F} \subseteq \alpha_{2}^{F}$ then $X_{2}^{\overline{F}} \subseteq X_{1}^{\overline{F}}$

Because P-sets has dynamic characteristics, it can general a cluster P-sets. The cluster P-sets is denoted by:

$$\{(\mathbf{X}_{i}^{\overline{F}}, \mathbf{X}_{j}^{F})|i \in \mathbf{I}, j \in \mathbf{J}\}$$
(8)

where, I and J are index sets.

P-sets usually is represented as one set pair $(X^{\overline{F}},\,X^{F})$ for simplicity.

Dynamic characteristics of P-sets: If the attribute set α is supplemented by some attribute elements and other attribute elements are deleted from the attribute set α , then α is changed into α_1^F and $\alpha_1^{\overline{F}}$, respectively, $\alpha_1^{F} \neq \alpha_1^{\overline{F}}$; $\alpha \subseteq \alpha_1^F$, $\alpha_1^{\overline{F}} \subseteq \alpha$. By the Eq. 1-7, P-sets (X₁^{\overline{F}}, X₁^{\overline{F}}) can be gotten. If going on this procedure, then α changes into α_2^F and $\alpha_2^{\overline{F}}$, respectively. The $\alpha_2^{F} \neq \alpha_2^{\overline{F}}$; $\alpha_1^{F} \subseteq \alpha_2^{F}$, $\alpha_2^{\overline{F}} \subseteq \alpha_1^{\overline{F}}$. By the Eq. 1-7, P-sets (X₂^{\overline{F}}, X₂^{\overline{F}}) can be gotten and so on. Equation 8 is composed of the strings of set pairs (X₁^{\overline{F}}, X₁^{\overline{F}}).

Dependent characteristics of P-sets: In the Eq. 1-7, if α_1^F , α_2^F are the attribute sets of $X_1^{\overline{F}}$, $X_2^{\overline{F}}$, respectively and $\alpha_1^{F} \subseteq \alpha_2^F$, then there is $X_2^{\overline{F}} \subseteq X_1^{\overline{F}}$. Namely $X_1^{\overline{F}}$ depends on $X_2^{\overline{F}}$, denoted as:

$$X_2^{\bar{F}} \rightarrow X_1^{\bar{F}} \tag{9}$$

If α_1^{F} , α_2^{F} are the attribute sets of X_1^{F} , X_2^{F} , respectively and $\alpha_2^{\overline{F}} \subseteq \alpha_1^{\overline{F}}$, then there is $X_1^{F} \subseteq X_2^{F}$. Namely X_2^{F} depends on X_1^{F} , denoted as:

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

and P-sets $(X_2^{\bar{F}}, X_2^{F})$ depends on P-sets $(X_1^{\bar{F}}, X_1^{F})$, denoted as:

$$(X_1^{\overline{F}}, X_1^{\overline{F}}) \Longrightarrow (X_2^{\overline{F}}, X_2^{\overline{F}})$$
(11)

where, X_1^F , X_2^F in the Eq. 11 satisfy 9 and X_1^F , X_2^F satisfy the Eq. 10. The " \rightarrow " in the Eq. 9-11 comes from mathematical logic and reasoning.

If $X_k^{\overline{F}}$ is internal P-sets of X, then $X_k^{\overline{F}} \rightarrow X$. If $X_k^{\overline{F}}$ is outer P-sets of X, then $X \rightarrow X_k^{\overline{F}}$.

By the Eq. 1-7, the following theorem is gotten easily.

Theorem 1: (The recovery theorem of P-sets) If $(X^{\overline{F}}, X^{\overline{F}})$ is P-sets of X and $F = \overline{F} = \emptyset$, then:

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

In fact, if $F=\emptyset,$ then there is $\alpha^F=\alpha\cup\{\alpha'\big|f(\beta)=\alpha'\in\alpha,\ f\in F=\alpha.$

where, $\{\alpha'|f(\beta) = \alpha' \in \alpha, f \in F\} = \emptyset$. Also there is $X^- = \{x|x \in X, \overline{f} \in X\}$ = $u \in X, \overline{f} \in \overline{F}\} = \emptyset$ in the Eq. 2. Equation 1 changes into $X^{\overline{F}} = X - X^-$ = X. If $\overline{F} = \emptyset$, then there is the Eq. 6: $\alpha^{\overline{F}} = \alpha - \{\beta_i | \overline{f}(\alpha_i) = \beta_i \in \alpha, \overline{f} \in \overline{F}\}$ = α , where $\{\beta_i | \overline{f}(\alpha_i) = \beta_i \in \alpha, \overline{f} \in \overline{F} = \emptyset$. In the Eq. 5, there is $X^+ = \{u|u \in U, u \in X, f(u)\} = x' \in X, f \in F\} = \emptyset$. Equation 4 changes into $X^F = X \cup X^+ = x$. If $F = \overline{F} = \emptyset$, then $X^{\overline{F}} = X, X^F = X$. Thus there is the Eq. 12.

Theorem 1 indicates that P-sets are restored to the finite general set X under $F = \overline{F} = \emptyset$. P-sets (X^{\overline{F}}, X^F) comes back to the base set X. Namely, P-sets have lost dynamic characteristics. Actually, P-sets (X^{\overline{F}}, X^F) is the base set X under $F = \overline{F} = \emptyset$.

Apparently, if $F = \overline{F} = \emptyset$, the Eq. 8 changes into:

$$(X_i^{\overline{F}}, X_i^{F})|i \in I, j \in J\}_{\overline{F} = F = \emptyset} = X$$
(13)

Equation 13 indicates that every $X_i^{\overline{F}}$, X_j^{F} are restored to X under $\overline{F} = F = \emptyset$. The $(X_i^{\overline{F}}, X_j^{F})|i \in I, j \in J$ comes back to the base set X.

In fact according to the theorem 1, if $\overline{F} = F = \emptyset$, then $X^{\overline{F}} = X = X^{F}$, $\forall i \in I$, $\forall j \in J$. If $\overline{F} = F = \emptyset$, then $X_{i}^{\overline{F}} = X = X_{j}^{F}$. So, there is the Eq. 13.

DEPENDENT CHARACTERISTICS AND DEPENDENCE THEOREM FOR DATA

Assumption 2: The finite general sets X, $X^{\overline{F}}$, $X^{\overline{F}}$ above are indicated by (x), $(x)^{\overline{F}}$, $(x)^{F}$, respectively, namely (x) = X, $(x)^{\overline{F}} = X^{\overline{F}}$, $(x)^{F} = X^{F}$. Namely (x), $(x)^{\overline{F}}$, $(x)^{F}$ indicate data for following discussion.

Definition 1: (x) = {x₁, x₂,..., x_q} \subset U is called data of U and x_i \in (x) is called data element of (x) (i = 1, 2,..., q), if (x) has attribute set α that is denoted by:

$$\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_k\}$$
(14)

Definition 2: $(X)^{\overline{F}} = \{x_1, x_2, ..., x_p\} \subset U$ is called \overline{F} -data of (x), if attribute set α^F of $(x)^{\overline{F}}$ and α of (x) satisfy:

$$\alpha^{\rm F} - \alpha \neq \emptyset \tag{15}$$

 $(x)^F = \{x_1, x_2, ..., x_r\} \subset U$ is called F-data of (x), if attribute set $\alpha^{\overline{F}}$ of $(x)^F$ and attribute set α of (x) satisfy:

$$\alpha - \alpha^{\overline{F}} \neq \emptyset \tag{16}$$

The data pair $((x)^{\overline{r}}, (x)^{F})$ is called (\overline{F}, F) -data of (x), if $(x)^{\overline{F}}$ is \overline{F} -data of (x) and $(x)^{F}$ is F-data of (x). Where, p, q, r satisfy that $p \le q \le r. p, q, r \in \mathbb{N}^+$.

By the definition 3, the following proposition is gotten.

Proposition 1: \overline{F} -data $(x)^{\overline{F}}$, \overline{F} -data $(x)^{\overline{F}}$ and $(\overline{F}, \overline{F})$ -data $((x)^{\overline{F}}, (x)^{\overline{F}})$ are all abnormal data for (x).

Definition 3: Data (x) is called one-directional dependence on \overline{F} -data $(x)^{\overline{F}}$ and it is denoted as:

$$(\mathbf{x})^{\overline{\mathbf{r}}} \rightarrow (\mathbf{x}) \tag{17}$$

F-data $(x)^F$ is called one-directional dependence on (x) and it is denoted as:

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

Definition 4: Data (x) is called two-directional depends on (x) and it is denoted as:

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

where, in the Eq. 17-19 " \Rightarrow " is equivalent to " \subseteq " and " \Leftrightarrow " is equivalent to "=".

Theorem 2: \overline{F} -data one-directional dependence theorem. If data (x) one-directionally depends on (x)*, namely (x)* \rightarrow (x), then:

 The attribute set α of (x) and the attribute set α* of (x)* satisfy following dependent relation:

 $\alpha \Rightarrow \alpha^*$ (20)

(x)* is an F-data of (x), namely (x)* = (x)^F

Proof 1: As data (x) one-directionally depends on (x)*, namely, (x)* \rightarrow (x), (x)* = {x₁, x₂,..., x_p} \subseteq {x₁, x₂,..., x_q} = (x). By the Eq. 1-3, the attribute set α * of (x)* and the attribute set α of (x) satisfy that $\alpha \subseteq \alpha^*$. By definition 3, there is $\alpha \Rightarrow \alpha^*$. Because (x)* \Rightarrow (x), namely (x)* \subseteq (x), then (x)* is an F-data of (x) by the Eq. 1-3. Namely (x)* = (x)^F.

Theorem 3: F-data one-directional dependence theorem. If data $(x)^{\circ}$ one-directionally depends on (x), namely $(x) \rightarrow (x)^{\circ}$, then:

 The attribute set α of (x) and the attribute set α° of (x)° following dependent relation:

$$\alpha^{\circ} \Rightarrow \alpha$$
 (21)

(x)° is an F-data of (x), namely (x)° = (x)^F

The proof of the theorem 3 is similar to the proof of the theorem 2. Here, it is omitted.

Theorem 4: \overline{F} -data two-directional dependence theorem. The necessary and sufficient condition of the \overline{F} -data (x)^{\overline{F}} depending on data (x) two-directionally is that the attribute set α^{F} of (x)^{\overline{F}} and the attribute set α of (x) satisfy following dependent relation:

$$\alpha^{\mathrm{F}} \cdot \{\alpha_{i} | \alpha_{i} \in \alpha^{\mathrm{F}}, \overline{f}(\alpha_{i}) = \beta_{i} \overline{\in} \alpha^{\mathrm{F}}, \overline{f} \in \overline{\mathrm{F}}\} = \alpha$$
(22)

Proof 1: If \overline{F} -data $(x)^{\overline{F}}$ two-directionally depends on (x), namely $(x)^{\overline{F}} \leftrightarrow (x)$, then $(x)^{\overline{F}}$ has the same attribute set with (x). According to the Eq. 1-3, α^{F} and α of (x) satisfy $\alpha \subseteq \alpha^{F}$. Apparently, there is the attribute difference set $\nabla \alpha^{F} = \{\alpha_{i} | \alpha_{i} \in \alpha^{F}, \overline{f}(\alpha_{i}) = \beta_{i} \overline{e} \alpha^{F}, \overline{f} \in \overline{F}\}$. The $\nabla \alpha^{F}$ is deleted from α^{F} . Namely, $\alpha^{F} - \nabla \alpha^{F} = \alpha^{F} - \{\alpha_{i} | \alpha_{i} \in \alpha^{F}, \overline{f}(\alpha_{i}) = \beta_{i} \overline{e} \alpha^{F}, \overline{f} \in \overline{F}\} = \alpha$. As α^{F} and α satisfy $\alpha \subseteq \alpha^{F}$, or there is the attribute difference set $\{\alpha_{i} | \alpha_{i} \in \alpha^{F}, \overline{f}(\alpha_{i}) = \beta_{i} \overline{e} \alpha^{F}, \overline{f}(\alpha_{i}) = \beta_{i}$

Theorem 5: F-data two-directional dependence theorem. The necessary and sufficient condition of F-data $(x)^F$ depending on data (x) two-directionally is that the attribute set α^F of $(x)^F$ and the attribute set α of (x) satisfy following dependent relation:

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

The proof of the theorem 5 is similar to the proof of the theorem 4. Here, it is omitted.

Theorem 6: The data identification theorem of data one-directional dependence. If $(x)^{\overline{F}}$, (x), $(x)^{F}$ satisfy that:

$$(x)^{\overline{F}} \rightarrow (x), (x) \rightarrow (x)^{\overline{F}}$$
 (24)

Then:

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

Where: $IDE = Identifiable^{1,2}$

Corollary 1: If $(x)^{\overline{F}}$, (x), $(x)^{F}$ satisfy $(x)^{\overline{F}} \Leftrightarrow (x)$, $(x) \Leftrightarrow (x)^{F}$ then:

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

Where: $UNI = Unidentifiable^{1,2}$

DEPENDENT CHARACTERISTICS AND DEPENDENCE THEOREM FOR DATA

Definition 5: $\theta^{\overline{F}}$ is called \overline{F} -dependence measure for data (x) one-directionally depending on \overline{F} -data (x) $^{\overline{F}}$, called \overline{F} -dependence measure of (x) $^{\overline{F}}$ for short, moreover:

$$\theta^{\overline{F}} = \operatorname{card}(\alpha^{\overline{F}})/\operatorname{card}(\alpha)$$
 (27)

where, (x) and $(x)^{\overline{F}}$ satisfy $(x)^{\overline{F}} \rightarrow (x)$. The α and α^{F} are the attribute set of (x) and $(x)^{\overline{F}}$ and card is cardinal number.

Definition 6: θ^F is called F-dependence measure for data (x) one-directionally depending on F-data (x)^F, called F-dependence measure of (x)^F for short, moreover:

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

where, $(x)^{F}$ and (x) satisfy $(x) \rightarrow (x)^{F}$, $\alpha^{\overline{F}}$ and α are the attribute set of $(x)^{F}$ and (x).

Definition 7: ρ is called module of data (x) = {x₁, x₂,..., x_q} \subset U, moreover:

$$\rho = ||y||/||y||$$
(29)

where, $||y|| = (y_1^2+y_2^2+...+y_q^2)^{\frac{1}{2}}$ is 2-norm of vectory $y = (y_1, y_2,..., y_q)^T$, $y = (y_1, y_2,..., y_q)^T$ is vector generated by

characteristics value y_i of $x_i \in (x)$, i = 1, 2, ..., q. $y = (y_1, y_2, ..., y_q)$ is characteristics value set of (x).

Definition 8: $\rho^{\overline{F}}$ is called module of data $(x)^{\overline{F}} = (x_1, x_2, ..., x_p) \subset U$, moreover:

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

The $\rho^{\scriptscriptstyle F}$ is called module of data $(x)^{\scriptscriptstyle F}=(x_1,\ x_2,...,\ x_r){\scriptscriptstyle \subset} U,$ moreover:

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

where, $||y_{\bar{r}}|| = (y_1^2 + y_2^2 + ... + y_p^2)^{y_2}$ is 2-norm of vector $y_{\bar{r}} = (y_1, y_2, ..., y_p)^T$, $y_{\bar{r}}^{\bar{r}} = (y_1 + y_2, ..., y_p)^T$ is vector generated by the characteristics value y_j of $x_j \in (x)^{\bar{r}}$, j = 1, 2, ..., p. $||y_{\bar{r}}|| = (y_1^2 + y_2^2 + ... + y_r^2)^{1/2}$ is 2-norm of $y_{\bar{r}}^F = (y_1, y_2, ..., y_r)^T = (y_1, y_2, ..., y_r)^T$ is vector generated by the characteristics value y_k of $x_k \in (x)^F$, k = 1, 2, ..., r. $y_{\bar{r}}^{\bar{r}} = (y_1, y_2, ..., y_r)$ and $y_{\bar{r}}^F = (y_1, y_2, ..., y_r)$ are the characteristics value set of $(x)^{\bar{r}}$ and $(x)^{\bar{r}}$, respectively.

By the definitions 5-8, following propositions are gotten directly.

Proposition 2: If the dependence measure θ of data (x) and the dependence measure θ^* of data (x)* satisfy $\theta^*-\theta \ge 0$, then (x)* is an \overline{F} -data of (x), namely (x)* = (x)^{\overline{F}} and vice versa.

Proposition 3: If the dependence measure θ of data (x) and the dependence measure θ° of data (x)° satisfy $\theta^{\circ} - \theta \le 0$, then (x)° is an F-data of (x), namely (x)° = (x)^F and vice versa. Where, $\theta = card(\alpha)/card(\alpha)$ is the dependence measure of (x), α is the attribute set of (x).

Theorem 7: The separation theorem of abnormal \overline{F} -data. If the module ρ^* of data (x)* and the module ρ of (x) satisfy:

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

Then:

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

 $(x)^*$ is separated from (x) and $(x)^*$ is an F-data of (x).

Where:

- $GRD((x)^*) = Card((x)^*)/card((x))$ is the granulation degree of (x)*
- GRD((x)) = Card ((x))/card ((x)) is the granulation degree of (x)

CRD = Granulation degree

Proof: Because ρ^* and ρ are the modules of (x)* and (x), respectively and they satisfy $\rho^*-\rho \le 0$, by the Eq. 29 and 30, $y^* = \{y_1, y_2, ..., y_p\} \subseteq \{y_1, y_2, ..., y_q\} = y$. y^* and y are the characteristic value set of (x)* and (x). As (x)* = $\{x_1, x_2, ..., x_p\} \subseteq \{x_1, x_2, ..., x_q\} = (x)$ and GRD ((x)*) = card ((x)*)/card ((x)) \le card ((x))/card ((x)) = GRD ((x)), namely GRD ((x)*)-GRS ((x)) \le 0. Thus (x)* is separated from (x). As (x)* \subseteq (x), data (x) one-directional depends on (x)*, there is that (x)* is an F-data of (x) by the theorem 2. So (x)* = (x)^F.

The theorem 7 get a fact: As (x) loses some data elements x_i , (x) becomes smaller (or card ((x)) is decreased). So (x) changes into (x)*. (x)* is an abnormal data of (x) (or $(x)^{\overline{r}}$ is an abnormal data of (x)), namely (x)* hides in (x)(or $(x)^* \subset (x)$). Thus (x)* is separated from (x).

Theorem 8: The separation theorem of abnormal F-data. If the module ρ° of data (x)° and the module ρ of data (x) satisfy:

$$\rho^{\circ}$$
- $\rho \ge 0$ (34)

Then:

$$GRD((x)^{\circ})-GRD((x)) \ge 0$$
 (35)

The $(x)^{\circ}$ is separated from (x) and $(x)^{\circ}$ is an F-data of (x). The proof is similar to the proof of Theorem 7, so it is omitted.

The theorem 8 get a fact: As some data elements x_j invade into data (x), (x) becomes larger (or, card ((x)) is increased) and (x) changes into (x)°. The (x)° is an abnormal data of (x), (or (x)^F is an abnormal data of (x)). The (x)° hides out of (x) (or (x) \subset (x)°)). Thus (x)° is separated out (x).

Based on the theorems 7 and 8, following theorem is gotten directly.

Theorem 9: The recovery theorem of abnormal \overline{F} -data. The necessary and sufficient condition of abnormal data (x)* being restored into data (x) is that the attribute set α^F of (x)* and the attribute set α of (x) satisfy following relation:

$$\alpha^{\mathrm{F}} \cdot \alpha = \emptyset \tag{36}$$

In fact, by the Eq. 1-3, $\alpha^F = \alpha \cup \{\alpha' | f(\beta = \alpha' \in \alpha, f \in F\}$. Namely, α^F and α satisfies $\alpha \subseteq \alpha^F$. Therefore, if $\{\alpha' | f(\beta = \alpha' \in \alpha, f \in F\} = \emptyset$, then $\alpha^F = \alpha$, namely $\alpha^F - \alpha = \emptyset$. So (x)* has the same attribute set with (x). By the Eq. 29 and 30, $\rho = \rho^*$, namely, (x) = (x)*. Thus (x)* is restored into (x).

Theorem 10: The recovery theorem of abnormal F-data. The necessary and sufficient condition of abnormal data $(x)^\circ$ being

restored into data (x) is that the attribute set $\alpha^{\overline{F}}$ of (x)° and the attribute set α of (x) satisfy following relation:

$$\alpha - \alpha^{\overline{F}} = \emptyset \tag{37}$$

The proof can be gotten according to the Eq. 4-6, 29 and 31. So, it is omitted.

RESULTS

Application of separating abnormal data

Assumption 3: $y = \{y_1, y_2, ..., y_q\}, y^{\bar{r}} = \{y_1, y_2, ..., y_p\}$ and $y^{\bar{r}} = \{y_1, y_2, ..., y_r\}$ are called characteristics value set of (x), $(x)^{\bar{r}}, (x)^{\bar{r}}$, respectively in section 4. For simplicity and without misunderstandings, y, $y^{\bar{r}}$, $y^{\bar{r}}$ are called data in this section.

The example of comes from an identification system for computer vision. A subsystem of the system is taken without losing generality. Figure 1 is a simplified block diagram of this subsystem.

In Fig. 1, ① is the module generated by data module $\rho^{\overline{F}}$ and ρ^{F} . ② is the comparison module of the data module $\rho^{\overline{F}}$ and ρ^{F} . ③ and ④ are the memory modules of standard module $\rho = 1$. ⑤ is the early warning module of the data module $\rho^{\overline{F}}$. ⑥ is the early warning module of the data module $\rho^{\overline{F}}$.

The working process of the subsystem in Fig. 1 as follows. If the subsystem is normal condition, the data $y = \{y_1, y_2, ..., y_q\}$, enters into the module ① from the port G and y generates the data module $\rho = ||y||/||y|| = 1$ in ①. The data y enters into the module ② and ③ outputs B = "1", A = "1", D = "0", C = "1" and E = "0". The B = "1" expresses that B outputs the standard data $y = \{y_1, y_2, ..., y_q\}$. There will be two cases as follows if the system occurs abnormal state (I): As the data element $y_1 \in y$ are lost from data the y which enters into ① from the port G, the



Fig. 1: Subsystem of a computer vision-identification system

Table 1: Standard data in the computer vision identification system

| 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 |
| - | | | | | | |

| Table 2: Data which has data loss in the computer vision identification system | | | | | | | | | | |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|--|--|--|--|
| y ₁ | y ₂ | y ₃ | y ₄ | y ₅ | y ₆ | y ₇ | | | | |
| 1.83 | 1.92 | - | 1.04 | 1.16 | 1.52 | - | | | | |

standard data $y = \{y_1, y_2, ..., y_q\}$ changes into $y^{\overline{F}} = \{y_1, y_2, ..., y_q\}$ y_p , $p \le q$. The $y^{\overline{F}}$ enters into ① and generates $\rho^{\overline{F}} =$ $\|\mathbf{y}^{\overline{\mathbf{r}}}\|/\|\mathbf{y}\| \leq \|\mathbf{y}\|/\|\mathbf{y}\| = \rho = 1$, namely $\rho^{\overline{\mathbf{r}}} < \rho = 1$. The module ③ outputs "0". The module (5) outputs D = "1", namely the system gives early warning and the system outputs A = "0", B = "0". Where, B = "0" expresses that the output B gives the abnormal data $y^{\overline{F}}$. (II) As data y which enters into ① from the port G is invade by data y'_{k} , the standard data $y = \{y_1, y_2, ..., y_q\}$ changes into $y^{F} = \{y_{1}, y_{2}, ..., y_{t}\}, q \le r$. The y^{F} enters into ① and generates $\rho^{F} = \|y^{F}\| / \|y\| \ge \|y\| / \|y\| = \rho = 1$, namely $\rho^{F} \ge \rho = 1$. The module ④ outputs "0" and the module 6 outputs E = "1". The system gives early warning and outputs C = "0" and B = "0". Where, B = "0" expresses that output B gives the abnormal data y^F. The states I and II express that the abnormal state of the system (the data y^F or y^F) is separated and identified from the normal state (the standard data y). Where, the port G is the input of the system and B is the output of the subsystem. "0" and "1" are logic values of the state B.

In order to discuss easily, the following separation-identification criterion is given.

Separation-identification criterion of abnormal data: The module ρ^* of data y^* ($\rho^* = \rho^{\overline{r}}$ or $\rho^* = \rho^{\overline{r}}$) and the module ρ of y satisfy:

IDE
$$(\rho^*, \rho)$$
 (38)

Namely, the output data of system is separated and the logic value of output is "0".

The experimental data of the computer vision identification system is taken. For briefness and without losing generality, taking $y = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$ is the sampling data of B in Fig. 1 and the Table 1 lists the values $y = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$.

The data of the Table 1 comes from the experimental data that is processed by the technical means and it does not affect the analytical results and discussion about separation-identification of abnormal data. The data meet meets the requirements of the image of the vision-identification system (distortion degree is 0.03). The vision-identification system can correctly identify a moving image which is an image monitoring. The system outputs

B = "1" and satisfies the Eq. 29. Namely $\rho = 1$. There is that A = "1", D = "0", C = "1", E = "0". The B = "1" expresses that B outputs the standard data $y = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$.

In order to use the system to detect a case accurately and assess the practicability and stability of the system, an experiment is presented for state I. The parameters of the module ① is adjusted in the Fig. 1 so that the data in the Table 1 is lost, then the Table 2 is obtained.

In the Table 2 "-" means "null data". The data make the image of the vision-identification system distortion (The distortion degree is 0.87). The output B = "0" satisfies the Eq. 32, namely $\rho^* = 3.43/4.08 \le 0.84 \le 1$. In the figure, A = "0", "0", D = "1", E = "1". The B = "0" expresses B give the abnormal data y = {y₁, y₂, y₃, y₄, y₅, y₆} = y^F.

Note: Adjust the parameters of the module ① in Fig. 1 is adding a "filter" in an actual system and adding the "filter" is equivalent to α being supplemented by attributes in the Eq. 14. Apparently, if α is the attribute set of $y = \{y_1, y_2, ..., y_q\}, \alpha^F$ is the attribute set of $y^{\overline{F}} = \{y_1, y_2, ..., y_p\}$ and $\alpha \subseteq \alpha^F$, there is $y^{\overline{F}} \subseteq y$.

Using the Table 1 and 2, $\theta^{\overline{r}} = \operatorname{card} (\alpha^{\overline{r}})/\operatorname{card}(\alpha) \ge 1$. In the Eq. 27, $y^{\overline{r}} \Rightarrow y$ by the theorem 2. Based on the theorem 7, $y^{\overline{r}}$ is an abnormal data of y and $\rho^{\overline{r}} \le \rho$. By the Eq. 38, the abnormal data $y^{\overline{r}}$ is separated and identified from the standard data y. Another experiment (for the state II) is omitted.

DISCUSSION

Abnormal data is often encountered in computer control systems and computer information systems. But there is less the theory about researching abnormal data. When data $(x) = \{x_1, x_2, ..., x_n\}$ is transferred in computer system, the system gets data $(x)^{\overline{F}} = \{x_1, x_2, ..., x_p\}$ $(p \le q)$ or $(x)^F = \{x_1, x_2, ..., x_r\}$ $(q \le r)$ sometimes. In fact, (x) changes into $(x)^{\overline{F}}$, there is $(x)^{\overline{F}} \subseteq (x)$. On the other hand, when (x) changes into $(x)^F$, $(x) \subseteq (x)^F$. The reason of result in two changes is that the attribute α of (x) is reduced or added. But the fact is often ignored. The study finds back the fact ignored and discusses the changes for the abnormal data $(x)^{\overline{F}}$ and $(x)^{\overline{F}}$ on based the new mathematical model P-sets. The study proposes the concepts of (\overline{F} , F)-data and data dependence. The (F, F)-data consists of F-data and F-data. Here, F-data, F-data and (F, F)-data are all abnormal. The generations of these abnormal data are related to the change of the attribute α of (x) and have dependence characteristics. The study gives the dependence characteristics and dependence-identification theorems of (F, F)-data. By defining the module of data, the paper provides the method of separating-recovery (F, F)-data and the criterion of separating abnormal data. The paper gives the application about separating abnormal data. The discussion and results in this study can extend to computer control systems so as to obtain some new study and can be applied in data mining, data filtration and data discovery, fault detection and ruling etc. But P-sets not only has dynamic characteristics^{1,2,6,7} but also random characteristics^{18,27,28}, so there is phenomenon III in addition the phenomena of I and II in the Introduction. That is when the phenomena I and II simultaneously appear, (x) generates data $(x)^{F}$ and $(x)^{F}$. The phenomenon is more complex than the phenomena I and II. It cannot be solved directly by using the method of this study. And it will be researched in this study.

ACKNOWLEDGMENT

This study was supported by the Natural Science Foundation of China (70871117), the Innovative Team on Science and Technology of Ningde Normal University of China (2013T08).

REFERENCES

- 1. Shi, K.Q., 2008. P-sets. J. Shandong Univ. (Nat. Sci.), 43: 77-84.
- 2. Shi, K.Q., 2009. P-sets and its applications. Int. J. Adv. Syst. Sci. Applic., 9: 209-219.
- 3. Li, Y., J. Li, Q. Ruan and Y. Wu, 2016. Separation of dynamic data and recovery of P-abnormal data. Int. J. Mach. Learning Cybernetics. 10.1007/s13042-015-0483-x
- 4. Li, Y., 2012. P-recursion information and recursion search ring-application. J. Convergence Inform. Technol., 7: 42-53.
- 5. Lin, H.K., Y.Y. Li and C.X. Fan, 2012. Information outer-camouflage and true-false information. Int. J. Digital Content Technol. Applic., 6: 19-29.
- 6. Shi, K.Q., 2010. P-sets and its applied characteristics. Comput. Sci., 37: 1-8.
- Shi, K.Q., 2012. P-sets, inverse P-sets and the intelligent fusion-filter identification of information. Comput. Sci., 39: 1-13.
- Yu, X.Q., 2012. P-law reasoning and unknown law discovery-application. J. Convergence Inform. Technol., 7: 146-154.
- 9. Shi, K.Q., 2011. P-reasoning and P-reasoning discovery-identification of information. Comput. Sci., 38: 1-9.
- 10. Fan, C.X. and H.K. Lin, 2012. P-sets and the reasoning-identification of disaster information. J. Convergence Inform. Technol., 7: 337-345.
- Lin, H.K. and C.X. Fan, 2012. The dual form of P-reasoning and identification of unknown attribute. Int. J. Digital Content Technol. Applic., 6: 121-131.

- Li, Y.Y., Q.S. Ruan, S.L. Zhang, H.L. Yang and Y.M. Mao, 2011. Identification and recovery of inward-recursion information. J. Shandong Univ. (Nat. Sci.), 46: 121-126.
- Huang, S., X. Yu and R. Song, 2010. P-sets and P-information law recovery-identification. Proceedings of the 2nd International Workshop on Intelligent Systems and Applications, May 22-23, 2010, Wuhan, pp: 1-4.
- 14. Yuying, L., Z. Li and Z. Shiliang, 2011. Geometric identification of search states on (*F, F*)-information. Proceedings of the 6th International Conference on Computer Science and Education, August 3, 2011, Milwaukee, pp: 1189-1193.
- 15. Li, Y.Y., C.X. Fan and K.Q. Shi, 2011. Mixed memory information and memory information filtering. Syst. Eng. Electr., 33: 1824-1828.
- Lin, R. and C. Fan, 2012. Packet sets and identification of inward-convergence information. J. Convergence Inform. Technol., 7: 157-164.
- 17. Shi, K.Q., 2014. P-information law intelligent fusion and soft information #br# image intelligent generation. J. Shandong Univ. (Nat. Sci.), 49: 1-17.
- Shi, K.Q., J.H. Tang and L. Zhang, 2015. Intelligent fusion of inverse packet information and recessive transmission of information's intelligent hiding. Syst. Eng. Electr., 37:599-605.
- Wang, Y., H. Geng and K. Shi, 2010. The mining of dynamic information based on P-sets and its applications. Int. J. Adv. Syst. Sci. Applic., 10: 234-240.
- 20. Yu, X.Q., 2011. Internal P-set and security transmission-identification of information. J. Comput., 6: 2249-2254.
- 21. Yu, X.Q., 2012. F-data family and restoring-recovery of deformity data sets. Adv. Inform. Sci. Serv. Sci., 4: 363-370.
- 22. Guanyu, Z., L. Enzhong and Z. Hui, 2010. Information gene and identification of its information Knock-out/Knock-in. Int. J. Adv. Syst. Sci. Applic., 10: 308-315.
- 23. Lin, H.K. and Y.Y. Li, 2012. Data circle characteristics of information and information state identification. J. Convergence Inform. Technol., 7: 157-167.
- 24. Shi, K.Q. and L. Zhang, 2009. Internal P-sets and data outer-recovery. J. Shandong Univ. (Nat. Sci.), 44: 8-14.
- 25. Lin, H. and Y. Li, 2010. P-sets and its P-separation theorems. Int. J. Adv. Syst. Sci. Applic., 10: 209-215.
- 26. Liu, J.Q., 2010. P-probabilities and its application. Int. J. Adv. Syst. Sci. Applic., 10: 200-208.
- 27. Guo, Z.L., 2011. Random characteristics of P-sets. Fuzzy Syst. Math., 25: 170-174.
- Guo, Z.L., S.L. Zhao and K.Q. Shi, 2013. System state detection-recognition based on random P-sets. Comput. Sci., 40: 178-180.