

P-数据模型与数据的智能获取

徐凤生¹ 于秀清¹ 史开泉²

(德州学院数学科学学院 山东 德州 253023)¹ (山东大学数学与系统科学学院 济南 250100)²

摘 要 文中首次提出具有动态特征的 P-数据(Packet data)模型。P-数据模型通过改进 P-集合(Packet sets)模型而得到,其由内 P-数据(internal Packet data)与外 P-数据(outer Packet data)共同构成。给出了 P-数据模型与模型结构,得到了 P-数据模型与它的属性合取范式扩展-萎缩生成定理,进而得到了 P-数据推理(Packet data reasoning)模型与数据的智能获取定理;最后利用这些概念和结果,给出内 P-数据智能获取与风险数据估计的应用。

关键词 P-集合,P-数据模型,属性合取,P-数据推理,数据智能获得,应用

中图法分类号 O144 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.05.029

P-data Model and Intelligent Acquisition of P-data

XU Feng-sheng¹ YU Xiu-qing¹ SHI Kai-quan²

(College of Mathematical Sciences, Dezhou University, Dezhou, Shandong 253023, China)¹

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

Abstract P-data(Packet data) model possessing dynamic characteristics was proposed for the first time by improving P-sets(Packet sets) model in the paper. It is composed of internal P-set(internal Packet data) and outer P-set(outer Packet data), and it is obtained by improving packet sets. P-data model and model structure were given in this paper. The attribute conjunctive normal form of P-data model was got. And then, the model of P-data reasoning and the intelligent acquisition theory of data were obtained. Finally, these concepts and results were applied to internal packet data intelligent acquisition and risk estimations.

Keywords P-sets, P-data model, Attribute conjunction, P-data reasoning, Data intelligent acquisition, Applications

1 引言

2008 年,文献[1-2]基于一个事实提出了 P-集合(packet sets)的概念,给出了 P-集合的理论模型,P-集合在动态数据(动态信息)研究中得到了应用[1-19]。P-集合是由内 P-集合 $X^{\bar{F}}$ (internal packet set $X^{\bar{F}}$) 与外 P-集合 X^F (outer packet set X^F) 构成的元素集合对,或者称 $(X^{\bar{F}}, X^F)$ 是 P-集合。P-集合具有如下动态特征:给定有限普通元素集合 $X = \{x_1, x_2, \dots, x_q\} \subset U, \alpha = \{\alpha_1, \alpha_2, \dots, \alpha_k\} \subset V$ 是 X 的属性集合;1)若在 α 内补充一些属性后, α 变成 $\alpha^F, \alpha \subseteq \alpha^F$, 则在 X 内删除一些元素后, X 变成内 P-集合 $X^{\bar{F}}, X^{\bar{F}} \subseteq X$ 。2)若在 X 的属性集合 α 内删除一些属性后, α 变成 $\alpha^{\bar{F}}, \alpha^{\bar{F}} \subseteq \alpha$, 则在 X 内补充一些元素后, X 变成外 P-集合 $X^F, X \subseteq X^F$ 。3)若在 X 的属性集合 α 内补充一些属性,同时删除另一些属性后, α 变成 α^F , 同时 α 变成 $\alpha^{\bar{F}}, \alpha^{\bar{F}} \subseteq \alpha \subseteq \alpha^F$, 则 X 变成内 P-集合 $X^{\bar{F}}$, 同时 X 变成外 P-集合 $X^F, X^{\bar{F}} \subseteq X \subseteq X^F$;如果在 α 内删除属性、补充属性的过程连续发生,则 X 变成多个元素集合对 $(X_i^{\bar{F}}, X_i^F)$,

$(X_2^{\bar{F}}, X_2^F), \dots, (X_n^{\bar{F}}, X_n^F), (X^{\bar{F}}, X^F)$, 其中任意一个集合对称作 X 生成的 P-集合;所有元素集合对构成 P-集合族 $\{(X_i^{\bar{F}}, X_i^F) | i \in I, j \in J\}$ 。在一定条件下, P-集合 $(X^{\bar{F}}, X^F)$ 被还原成有限普通集合 X , P-集合中的元素与它的属性满足属性合取范式。P-集合的动态特征与数据 Y 之内的冗余数据元 $y_i \in Y$ 从 Y 内删除,数据 Y 之外的数据元 $y_j \in Y$ 入侵到 Y 内后具有相同的动态特征。数据 Y 之内的冗余数据元 $y_i \in Y$ 被删除,通过数据采集得到的数据 Y 与 Y 内优选数据元构成了新的数据 $Y^{\bar{F}}$ (未被优选的数据元从 Y 内删除)。数据 Y 之外的数据元 $y_j \in Y$ 入侵到 Y 内构成新数据 Y^F 。本文采用数据集合 Y 代替 P-集合中的元素集合 X , 来改进 P-集合理论模型,提出 P-数据模型。P-数据模型中的数据元 y_i 与它的属性满足:属性合取范式;或者, P-数据模型中的数据元 y_i 与它的属性之间的逻辑关系和 P-集合中的元素 x_i 与它的属性之间的逻辑关系相同。P-数据模型由内 P-数据模型与外 P-数据模型共同构成。本文给出了内 P-数据模型与它的属性合取范式的扩展生成;给出了外 P-数据模型与它的属性合取范式的萎缩生成;给出了内 P-数据推理结构、外 P-数据推理结构与 P-

到稿日期:2017-04-29 返修日期:2017-07-13 本文受山东省自然科学基金(ZR2010AL019),山东省统计局重点项目(KT16253)资助。

徐凤生(1965—),男,硕士,教授,主要研究方向为信息系统理论与应用,E-mail: sxxfs@126.com;于秀清(1968—),女,硕士,教授,主要研究方向为粗系统理论与应用,E-mail: sddzyxq@163.com;史开泉(1945—),男,教授,博士生导师,主要研究方向为粗系统理论与应用,E-mail: shikq@sdu.edu.cn (通信作者)。

数据推理结构。利用这些研究,还给出了内 P-数据智能获取与它在风险数据估计中的应用。P-数据模型是 P-集合理论模型应用的新研究、新结果。

本文第 2 节介绍 P-集合的结构与特征,以及 P-集合存在的事实与例子,给出 P-集合中元素 x_i 与它的属性之间满足属性合取范式的表示;第 2 节中的所有概念是非常重要的。

2 P-集合与它的属性合取范式

文献[1-2]给定有限普通元素集合 $X = \{x_1, x_2, \dots, x_q\} \subset U$, $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_k\} \subset V$ 是 X 的属性集合,称 $X^{\bar{F}}$ 是 X 生成的内 P-集合,简称 $X^{\bar{F}}$ 是内 P-集合,而且:

$$X^{\bar{F}} = X - X^- \quad (1)$$

X^- 称作 X 的 $\bar{F}$ -元素删除集合,而且:

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

如果 $X^{\bar{F}}$ 的属性集合 $\alpha^{\bar{F}}$ 满足:

$$\alpha^{\bar{F}} = \alpha \cup \{\alpha_i' | f(\beta_i) = \alpha_i' \in \alpha, f \in F\} \quad (3)$$

其中, $\beta_i \in V, \beta_i \in \alpha; f \in F$ 把 β_i 变成 $f(\beta_i) = \alpha_i' \in \alpha, X^{\bar{F}} \neq \emptyset$ 。式(1)中, $X^{\bar{F}} = \{x_1, x_2, \dots, x_p\}, p \leq q; p, q \in \mathbb{N}^+$ 。

给定有限普通元素集合 $X = \{x_1, x_2, \dots, x_q\} \subset U, \alpha = \{\alpha_1, \alpha_2, \dots, \alpha_k\} \subset V$ 是 X 的属性集合,称 X^F 是 X 生成的外 P-集合,简称 X^F 是外 P-集合,而且:

$$X^F = X \cup X^+ \quad (4)$$

X^+ 称作 X 的 F -元素补充集合,而且:

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

如果 X^F 的属性集合 α^F 满足:

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

其中, $\alpha_i \in \alpha, \bar{f} \in \bar{F}$ 把 α_i 变成 $\bar{f}(\alpha_i) = \beta_i \in \alpha; \alpha^F \neq \emptyset$; 式(4)中, $X^F = \{x_1, x_2, \dots, x_r\}, q \leq r; q, r \in \mathbb{N}^+$ 。

由 $X^{\bar{F}}$ 与 X^F 构成的元素集对称作元素集合 X 生成的 P-集合(packet sets):

$$(X^{\bar{F}}, X^F) \quad (7)$$

X 称作 P-集合 $(X^{\bar{F}}, X^F)$ 的基集合,即基础集合(ground set)。

式(3)给出:

$$\alpha_1^{\bar{F}} \subseteq \alpha_2^{\bar{F}} \subseteq \dots \subseteq \alpha_{n-1}^{\bar{F}} \subseteq \alpha_n^{\bar{F}} \quad (8)$$

得到满足式(8)的内 P-集合:

$$X_n^{\bar{F}} \subseteq X_{n-1}^{\bar{F}} \subseteq \dots \subseteq X_2^{\bar{F}} \subseteq X_1^{\bar{F}} \quad (9)$$

式(6)给出:

$$\alpha_n^{\bar{F}} \subseteq \alpha_{n-1}^{\bar{F}} \subseteq \dots \subseteq \alpha_2^{\bar{F}} \subseteq \alpha_1^{\bar{F}} \quad (10)$$

得到满足式(10)的外 P-集合:

$$X_1^F \subseteq X_2^F \subseteq \dots \subseteq X_{n-1}^F \subseteq X_n^F \quad (11)$$

利用式(9)、式(11)得到:

$$\{(X_i^{\bar{F}}, X_j^F) | i \in I, j \in J\} \quad (12)$$

式(12)称作 X 生成的 P-集合族,即 P-集合的一般表达形式。

定理 1(P-集合第一还原定理) 若 $F = \bar{F} = \emptyset$, 则 P-集合被还原成有限普通元素集合 X , 或者:

$$(X^{\bar{F}}, X^F)_{\bar{F}=F=\emptyset} = X \quad (13)$$

定理 2(P-集合第二还原定理) 若 $F = \bar{F} = \emptyset$, 则 P-集合

族被还原成有限普通集合 X , 或者:

$$\{(X_i^{\bar{F}}, X_j^F) | i \in I, j \in J\}_{\bar{F}=F=\emptyset} = X \quad (14)$$

特别说明:1)在式(1)一式(7)与式(12)中, U 是有限元素论域, X 是 U 上的有限普通元素集合(有限普通经典集合); V 是有限属性论域。2) $F = \{f_1, f_2, \dots, f_n\}, \bar{F} = \{\bar{f}_1, \bar{f}_2, \dots, \bar{f}_m\}$ 是元素(属性)迁移族, $f \in F, \bar{f} \in \bar{F}$ 是元素(属性)迁移。 $f \in F$ 的特征是:对于元素 $x_i \in U, x_i \in X, f \in F$ 把 x_i 变成 $f(x_i) = x_i' \in X$; 对于属性 $\beta \in V, \beta \in \alpha, f \in F$ 把 β 变成 $f(\beta) = \alpha' \in \alpha; \bar{f} \in \bar{F}$ 的特征是:对于元素 $x_i \in X, \bar{f} \in \bar{F}$ 把 x_i 变成 $\bar{f}(x_i) = u_i' \in X$, 对于属性 $\alpha_i \in \alpha, \bar{f} \in \bar{F}$ 把 α_i 变成 $\bar{f}(\alpha_i) = \beta_i \in \alpha$ 。元素(属性)迁移是变换(函数)的概念。

P-集合的属性合取特征如下。

给定有限普通元素集合 $X = \{x_1, x_2, \dots, x_q\} \subset U, \alpha = \{\alpha_1, \alpha_2, \dots, \alpha_k\} \subset V$ 是 X 的属性集合, $\forall x_i \in X$ 的属性 α_i 满足属性合取范式:

$$\alpha_i = \alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_k \quad (15)$$

给定内 P-集合 $X^{\bar{F}} = \{x_1, x_2, \dots, x_p\} \subset U, \alpha^{\bar{F}} = \{\alpha_1, \alpha_2, \dots, \alpha_k, \alpha_{k+1}, \dots, \alpha_m\} \subset V$ 是 $X^{\bar{F}}$ 的属性集合, $\forall x_i \in X^{\bar{F}}$ 的属性 α_i 满足属性合取范式扩展:

$$\alpha_i = (\alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_k) \wedge \alpha_{k+1} \wedge \alpha_{k+2} \wedge \dots \wedge \alpha_m \quad (16)$$

给定外 P-集合 $X^F = \{x_1, x_2, \dots, x_r\} \subset U, \alpha^F = \{\alpha_1, \alpha_2, \dots, \alpha_k\} \subset V$ 是 X^F 的属性集合, $\forall x_j \in X^F$ 的属性 α_j 满足属性合取范式萎缩:

$$\begin{aligned} \alpha_j &= (\alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_n \wedge \alpha_{n+1} \wedge \dots \wedge \alpha_k) - (\alpha_{n+1} \wedge \\ &\quad \alpha_{n+2} \wedge \dots \wedge \alpha_k) \\ &= \alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_n \end{aligned} \quad (17)$$

给定 P-集合 $(X^{\bar{F}}, X^F), (\alpha^{\bar{F}}, \alpha^F)$ 是 $(X^{\bar{F}}, X^F)$ 的属性集合, $\forall x_i \in X^{\bar{F}}$ 的属性 $\alpha_i, \forall x_j \in X^F$ 的属性 α_j 满足:

$$(\alpha_i, \alpha_j) = ((\bigwedge_{t=1}^k \alpha_t) \bigwedge_{t=k+1}^m \alpha_t, \bigwedge_{t=1}^k \alpha_t - \bigwedge_{t=\lambda+1}^k \alpha_t) \quad (18)$$

其中, $\alpha_i = (\bigwedge_{t=1}^k \alpha_t) \bigwedge_{t=k+1}^m \alpha_t, \alpha_j = \bigwedge_{t=1}^k \alpha_t - \bigwedge_{t=\lambda+1}^k \alpha_t$ 。

约定:在第 3-4 节中,用 $Y, Y^{\bar{F}}, Y^F$ 分别代替第 2 节中的元素集合 X 、内 P-集合 $X^{\bar{F}}$ 、外 P-集合 X^F , 或者, $Y = X, Y^{\bar{F}} = X^{\bar{F}}, Y^F = X^F$; 用 $y_i, y_j^{\bar{f}}, y_\lambda^f$ 分别代替第 2 节中的元素 $x_i \in X, x_j \in X^{\bar{F}}, x_\lambda \in X^F$, 或者, $y_i = x_i, y_j^{\bar{f}} = x_j, y_\lambda^f = x_\lambda$ 。 $Y, Y^{\bar{F}}, Y^F$ 分别具有属性集合 $\alpha, \alpha^{\bar{F}}, \alpha^F$ 。把第 2 节中的 P-集合理论模型(1)一式(7)、式(12)、式(15)一式(18), 在第 3 节中给出 P-数据模型与它的属性合取范式扩展-萎缩生成。

3 P-数据模型与它的属性合取范式扩展-萎缩生成

定义 1 称 Y 是元素集合 $X = \{x_1, x_2, \dots, x_q\}$ 生成的数据集合, $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_k\}$ 是 Y 的属性集合。

$$Y = \{y_1, y_2, \dots, y_n\}$$

$$= \{ \sum_{i=1}^q y_{i,1}, \sum_{i=1}^q y_{i,2}, \dots, \sum_{i=1}^q y_{i,n} \} \quad (19)$$

其中, 针对 Y 内的每一个 $y_i, y_i = \{y_{i,1}, y_{i,2}, \dots, y_{i,n}\}, y_i$ 是元

素 $x_i \in X$ 具有的 n 维数值, $i=1, 2, \dots, q$; $\forall y_{\lambda, t} \in y_i, y_{\lambda, t} \in \mathbf{R}^+$ 。

定义 2 称 $Y^{\bar{F}}$ 是内 P-集合 $X^{\bar{F}} = \{x_1, x_2, \dots, x_p\}$ 生成的内 P-数据, $\alpha^{\bar{F}} = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$ 是 $Y^{\bar{F}}$ 的属性集合。

$$Y^{\bar{F}} = \{y_1^{\bar{F}}, y_2^{\bar{F}}, \dots, y_n^{\bar{F}}\} \\ = \{\sum_{i=1}^p y_{i,1}, \sum_{i=1}^p y_{i,2}, \dots, \sum_{i=1}^p y_{i,n}\} \quad (20)$$

式(19)、式(20)中, $p \leq q, p, q \in \mathbf{N}^+$ 。

定义 3 称 Y^F 是外 P-集合 $X^F = \{x_1, x_2, \dots, x_r\}$ 生成的外 P-数据, $\alpha^{\bar{F}} = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 是 Y^F 的属性集合。

$$Y^F = \{y_1^f, y_2^f, \dots, y_n^f\} = \{\sum_{i=1}^r y_{i,1}, \sum_{i=1}^r y_{i,2}, \dots, \sum_{i=1}^r y_{i,n}\} \quad (21)$$

式(19)和式(21)中, $q \leq r, r, q \in \mathbf{N}^+$ 。

式(19)一式(21)表明:根据第 2 节中的式(1)一式(3), $Y^{\bar{F}}$ 是在 Y 内删除了数据 $y_{p+1}, y_{p+2}, \dots, y_q$ 后得到的。根据第 2 节中的式(4)一式(6), Y^F 是在 Y 内补充了数据 $y_{q+1}, y_{q+2}, \dots, y_r$ 后得到的。

定义 4 由 $Y^{\bar{F}}$ 和 Y^F 构成的数据集合对称作 P-集合 $(X^{\bar{F}}, X^F)$ 生成的 P-数据:

$$(Y^{\bar{F}}, Y^F) \quad (22)$$

式(19)一式(22)称作 P-集合生成的 P-数据模型。

定理 3(内 P-数据生成与属性合取范式扩展定理) 内 P-数据 $Y^{\bar{F}}$ 被数据 Y 生成的充分必要条件是 $y_i^{\bar{F}} \in Y^{\bar{F}}$ 的属性 $\alpha_i \in \alpha^{\bar{F}}$ 是 $y_{\lambda} \in Y$ 的属性合取范式扩展, 或者:

$$\alpha_i = (\bigwedge_{t=1}^k \alpha_t) \bigwedge_{t=k+1}^m \alpha_t \quad (23)$$

其中, $\alpha_k = \bigwedge_{t=1}^k \alpha_t$ 是 $y_{\lambda} \in Y$ 的属性合取范式, $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_k\}$ 是 Y 的属性集合。

证明:因为 Y 代替 X , Y 与 X 具有相同的属性集合 α ; 类似地, $Y^{\bar{F}}$ 与 $X^{\bar{F}}$ 具有相同的属性集合 $\alpha^{\bar{F}}$ 。1) 给定数据集合 Y , $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_k\}$ 是 Y 的属性集合; 给定数据集合 $Y^{\bar{F}}$, $\alpha^{\bar{F}} = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$ 是 $Y^{\bar{F}}$ 的属性集合。由式(15)得: $\forall y_j \in Y, y_j$ 的属性集合 $\alpha_j = \alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_k = \bigwedge_{t=1}^k \alpha_t$; 由式(16)得: $\forall y_i^{\bar{F}} \in Y^{\bar{F}}$ 的属性集合 $\alpha_i = (\alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_k) \wedge \alpha_{k+1} \wedge \dots \wedge \alpha_m = (\bigwedge_{t=1}^k \alpha_t) \bigwedge_{t=k+1}^m \alpha_t$, 从而得到式(23)。2) $\forall y_i^{\bar{F}} \in Y^{\bar{F}}$ 的属性集合 α_i 满足式(23), 或者 $\alpha_i = (\bigwedge_{t=1}^k \alpha_t) \bigwedge_{t=k+1}^m \alpha_t$, 或者 $\alpha_i = \alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_k \wedge \alpha_{k+1} \wedge \dots \wedge \alpha_m = \{\alpha_1, \alpha_2, \dots, \alpha_k, \alpha_{k+1}, \dots, \alpha_m\}$; 由定义 2 可知 $Y^{\bar{F}}$ 是内 P-集合 $X^{\bar{F}}$ 生成的内 P-数据, 或者 $Y^{\bar{F}}$ 是 Y 生成的内 P-数据。由 1)、2) 得到定理 3 成立。

定理 4(外 P-数据生成与属性合取范式萎缩定理) 外 P-数据 Y^F 被数据 Y 生成的充分必要条件是 $y_j^f \in Y^F$ 的属性 $\alpha_j \in \alpha^{\bar{F}}$ 是 $y_{\lambda} \in Y$ 的属性合取范式萎缩, 或者:

$$\alpha_j = \bigwedge_{t=1}^k \alpha_t - \bigwedge_{t=n+1}^k \alpha_t \quad (24)$$

证明:与定理 3 类似, 略。

推论 1 P-数据 $(Y^{\bar{F}}, Y^F)$ 被数据 Y 生成的充分必要条件

是 $y_i^{\bar{F}} \in Y^{\bar{F}}$ 的属性 $\alpha_i \in \alpha^{\bar{F}}$ 是 $y_{\lambda} \in Y$ 的属性合取范式扩展, 同时 $y_j^f \in Y^F$ 的属性 $\alpha_j \in \alpha^{\bar{F}}$ 是 $y_{\lambda} \in Y$ 的属性合取范式萎缩, 或者:

$$(\alpha_i, \alpha_j) = ((\bigwedge_{t=1}^k \alpha_t) \bigwedge_{t=k+1}^m \alpha_t, (\bigwedge_{t=1}^k \alpha_t - \bigwedge_{t=n+1}^k \alpha_t)) \quad (25)$$

式(25)中 $\alpha_i = (\bigwedge_{t=1}^k \alpha_t) \bigwedge_{t=k+1}^m \alpha_t, \alpha_j = \bigwedge_{t=1}^k \alpha_t - \bigwedge_{t=n+1}^k \alpha_t$ 。

推论 2 数据 Y 生成的所有 P-数据 $(Y_i^{\bar{F}}, Y_j^F), (Y_i^{\bar{F}}, Y_j^F)$ 构成 P-数据族, 或者:

$$\{(Y_i^{\bar{F}}, Y_j^F) | i \in I, j \in J\} \quad (26)$$

利用第 2 节中的 P-集合结构与第 3 节中的 P-数据模型, 在第 4 节中给出 P-数据的动态智能获取与应用。

4 P-数据的动态智能获取与应用

4.1 P-数据推理与推理结构

定义 5 称:

$$\text{if } \alpha_k^{\bar{F}} \Rightarrow \alpha_{k+1}^{\bar{F}}, \text{ then } Y_{k+1}^{\bar{F}} \Rightarrow Y_k^{\bar{F}} \quad (27)$$

是内 P-数据生成的内 P-数据推理。 $\alpha_k^{\bar{F}} \Rightarrow \alpha_{k+1}^{\bar{F}}$ 称作内 P-数据推理条件, $Y_{k+1}^{\bar{F}} \Rightarrow Y_k^{\bar{F}}$ 为内 P-数据推理结论。

式(27)中, $\alpha_k^{\bar{F}}$ 和 $\alpha_{k+1}^{\bar{F}}$ 分别是内 P-数据 $Y_k^{\bar{F}}$ 和 $Y_{k+1}^{\bar{F}}$ 的属性集合; $\alpha_k^{\bar{F}} \Rightarrow \alpha_{k+1}^{\bar{F}}$ 与 $\alpha_k^{\bar{F}} \subseteq \alpha_{k+1}^{\bar{F}}$ 等价; $Y_{k+1}^{\bar{F}} \Rightarrow Y_k^{\bar{F}}$ 与 $Y_{k+1}^{\bar{F}} \subseteq Y_k^{\bar{F}}$ 等价。

定义 6 称:

$$\text{if } \alpha_{k+1}^{\bar{F}} \Rightarrow \alpha_k^{\bar{F}}, \text{ then } Y_k^F \Rightarrow Y_{k+1}^F \quad (28)$$

是外 P-数据生成的外 P-数据推理。 $\alpha_{k+1}^{\bar{F}} \Rightarrow \alpha_k^{\bar{F}}$ 称作外 P-数据推理条件, $Y_k^F \Rightarrow Y_{k+1}^F$ 为外 P-数据推理结论。

式(28)中, $\alpha_k^{\bar{F}}$ 和 $\alpha_{k+1}^{\bar{F}}$ 分别是外 P-数据 Y_k^F 和 Y_{k+1}^F 的属性集合; $\alpha_{k+1}^{\bar{F}} \Rightarrow \alpha_k^{\bar{F}}$ 与 $\alpha_{k+1}^{\bar{F}} \subseteq \alpha_k^{\bar{F}}$ 等价; $Y_k^F \Rightarrow Y_{k+1}^F$ 与 $Y_k^F \subseteq Y_{k+1}^F$ 等价。

定义 7 称:

$$\text{if } (\alpha_k^{\bar{F}}, \alpha_{k+1}^{\bar{F}}) \Rightarrow (\alpha_{k+1}^{\bar{F}}, \alpha_k^{\bar{F}}), \text{ then } (Y_{k+1}^{\bar{F}}, Y_k^F) \Rightarrow (Y_k^{\bar{F}}, Y_{k+1}^F) \quad (29)$$

是 P-数据生成的 P-数据推理。 $(\alpha_k^{\bar{F}}, \alpha_{k+1}^{\bar{F}}) \Rightarrow (\alpha_{k+1}^{\bar{F}}, \alpha_k^{\bar{F}})$ 称作外 P-数据推理条件, $(Y_{k+1}^{\bar{F}}, Y_k^F) \Rightarrow (Y_k^{\bar{F}}, Y_{k+1}^F)$ 称作外 P-数据推理结论。式(29)中, $(\alpha_k^{\bar{F}}, \alpha_{k+1}^{\bar{F}})$ 和 $(\alpha_{k+1}^{\bar{F}}, \alpha_k^{\bar{F}})$ 分别是 P-数据 $(Y_k^{\bar{F}}, Y_{k+1}^F)$ 和 $(Y_{k+1}^{\bar{F}}, Y_k^F)$ 的属性集合; $(\alpha_k^{\bar{F}}, \alpha_{k+1}^{\bar{F}}) \Rightarrow (\alpha_{k+1}^{\bar{F}}, \alpha_k^{\bar{F}})$ 与 $\alpha_k^{\bar{F}} \subseteq \alpha_{k+1}^{\bar{F}}, \alpha_{k+1}^{\bar{F}} \subseteq \alpha_k^{\bar{F}}$ 等价; $(Y_{k+1}^{\bar{F}}, Y_k^F) \Rightarrow (Y_k^{\bar{F}}, Y_{k+1}^F)$ 与 $Y_k^{\bar{F}} \subseteq Y_{k+1}^F, Y_{k+1}^{\bar{F}} \subseteq Y_k^F$ 等价。

由定义 5—定义 7 得到以下 3 个结论, 这些结论在数据智能获取中得到应用。

结论 1 若给定属性集合 $\alpha_k^{\bar{F}}, \alpha_{k+1}^{\bar{F}}$ 与内 P-数据 $Y_k^{\bar{F}}$, 则满足 $\alpha_k^{\bar{F}} \Rightarrow \alpha_{k+1}^{\bar{F}}$ 的未知内 P-数据 $Y_{k+1}^{\bar{F}}$ 在内 P-数据 $Y_k^{\bar{F}}$ 内被智能获取, $Y_{k+1}^{\bar{F}} \subseteq Y_k^{\bar{F}}$ 。

结论 2 若给定属性集合 $\alpha_k^{\bar{F}}, \alpha_{k+1}^{\bar{F}}$ 与外 P-数据 Y_k^F , 则满足 $\alpha_{k+1}^{\bar{F}} \Rightarrow \alpha_k^{\bar{F}}$ 的未知外 P-数据 Y_{k+1}^F 在外 P-数据 Y_k^F 外被智能获取, $Y_k^F \subseteq Y_{k+1}^F$ 。

结论 3 若给定属性集合 $(\alpha_k^{\bar{F}}, \alpha_{k+1}^{\bar{F}}), (\alpha_{k+1}^{\bar{F}}, \alpha_k^{\bar{F}})$ 与 P-数据 $(Y_k^{\bar{F}}, Y_k^F)$, 则满足 $(\alpha_k^{\bar{F}}, \alpha_{k+1}^{\bar{F}}) \Rightarrow (\alpha_{k+1}^{\bar{F}}, \alpha_k^{\bar{F}})$ 的未知 P-数据

(Y_{k+1}^F, Y_{k+1}^F) 在 P-数据推理中被智能获取, $Y_{k+1}^F \subseteq Y_k^F, Y_k^F \subseteq Y_{k+1}^F$ 。

内 P-数据推理、外 P-数据推理与 P-数据推理是分别通过改进内 P-推理^[5]、外 P-推理^[5]与 P-推理^[5]得到的。

4.2 P-数据推理模型与数据智能获取-风险识别的应用

约定:为了简单而又不失应用的一般性,本节只给出内 P-数据模型与数据智能获取-风险识别的应用;外 P-数据模型与数据智能获取-风险识别的应用、P-数据模型与数据智能获取-风险识别的应用省略。 Q 是生产石油-化学产品的集团公司, $Q = \{Q_1, Q_2, Q_3, Q_4, Q_5\}$, $Q_i \in Q$ 是 Q 的子公司, $i = 1, 2, 3, 4, 5$; $\alpha = \{\alpha_1, \alpha_2, \alpha_3\}$ 是 Q 的属性集合(Q 的产品市场特征集合)。鉴于商业秘密,集团公司和子公司分别用 Q 和 Q_i 表示; Q 和 Q_i 的名称从略。属性(市场特征)分别用 $\alpha_1, \alpha_2, \alpha_3$ 表示; $\alpha_1, \alpha_2, \alpha_3$ 的名称从略。 Y 和 Y_i 分别是 Q 和 Q_i 在 2010 年 1 月—6 月的利润构成的利润离散数据分布数据集: $Y = \{Y_1, Y_2, Y_3, Y_4, Y_5\}$, $Y_i = \{y_{i,1}, y_{i,2}, y_{i,3}, y_{i,4}, y_{i,5}, y_{i,6}\}$, $i = 1, 2, 3, 4, 5$ 。 Y 和 Y_i 中的数值是真实利润值经过技术方法处理后得到的,但不影响例子的介绍与分析。

表 1 给出了 Q 和 Q_i 的利润离散分布数据 Y 和 Y_i 。

表 1 集团公司 Q 、子公司 Q_i 在 2010 年 1 月—6 月的利润 Y 和 Y_i 的离散分布

Table 1 Discrete distributions of profits Y, Y_i of group company Q and its subsidiary Q_i from January to June in 2010

k	1	2	3	4	5	6
Y_1	0.41	0.30	0.62	0.18	0.54	0.44
Y_2	0.21	0.28	0.45	0.36	0.66	0.53
Y_3	0.45	0.26	0.70	0.40	0.18	0.24
Y_4	0.19	0.35	0.46	0.68	0.77	0.16
Y_5	0.28	0.49	0.33	0.59	0.80	0.42
Y	1.54	1.68	2.56	2.21	2.95	1.79

表 1 中, Q 的利润离散分布数据集 $Y = \{y_1, y_2, y_3, y_4, y_5, y_6\} = \{\sum_{i=1}^5 y_{i,1}, \sum_{i=1}^5 y_{i,2}, \dots, \sum_{i=1}^5 y_{i,6}\}$, 表 1 生成的 Q 的利润离散分布矩阵为:

$$A = \begin{bmatrix} 0.41 & 0.21 & 0.45 & 0.19 & 0.28 \\ 0.30 & 0.28 & 0.26 & 0.35 & 0.49 \\ 0.62 & 0.45 & 0.70 & 0.46 & 0.33 \\ 0.18 & 0.36 & 0.40 & 0.68 & 0.59 \\ 0.54 & 0.66 & 0.18 & 0.77 & 0.80 \\ 0.44 & 0.53 & 0.24 & 0.16 & 0.42 \end{bmatrix} \quad (30)$$

2011 年 4 月 24 日由于子公司 Q_3 的石油管道原油泄漏而引起火灾, Q_3 的生产厂房变成一片废墟, 生产设备损坏。2011 年 5 月 3 日子公司 Q_5 的煤气(瓦斯)容器因压力超过极限值而发生爆炸, 导致煤气泄漏, 62 名工人中毒死亡。风险属性 α_4 与风险属性 α_5 入侵属性集合 α 内, α 变成 α^F :

$$\alpha^F = \alpha \cup \{\alpha_4, \alpha_5\} = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5\} \quad (31)$$

其中, α_4 和 α_5 分别表示 2011 年 4 月 24 日子公司 Q_3 的火灾与 2011 年 5 月 3 日子公司 Q_5 的爆炸。

在属性集合 α 变成 α^F 的条件下, 子公司 Q_3 与 Q_5 停产; Q 的利润离散分布表由表 1 变成表 2; Q 的利润离散分布数据集:

$$Y = \{Y_1, Y_2, Y_3, Y_4, Y_5\} \quad (32)$$

变成内 P-数据集:

$$Y^F = \{Y_1, Y_2, Y_4\} \quad (33)$$

表 2 属性集合 α_4 和 α_5 入侵属性集合 α 内后集团公司 Q 和子公司 Q_i 在 2011 年 7 月—12 月的利润 Y^F 和 Y_i^F

Table 2 Discrete distributions of profits Y^F, Y_i^F of group company Q and its subsidiary Q_i from July to December in

2011 with attribute sets α_4, α_5 transformed into α

k	7	8	9	10	11	12
Y_1	0.41	0.30	0.62	0.18	0.54	0.44
Y_2	0.21	0.28	0.45	0.36	0.66	0.53
Y_4	0.19	0.35	0.46	0.68	0.77	0.16
Y	0.81	0.93	1.53	1.22	1.97	1.13

在 α_4 和 α_5 存在的条件下, Q 的利润离散分布矩阵为:

$$A^F = \begin{bmatrix} 0.41 & 0.21 & 0.19 \\ 0.30 & 0.28 & 0.35 \\ 0.62 & 0.45 & 0.46 \\ 0.18 & 0.36 & 0.68 \\ 0.54 & 0.66 & 0.77 \\ 0.44 & 0.53 & 0.16 \end{bmatrix} \quad (34)$$

数据集 Y^F 的智能获取分析如下:

在属性集合 α 内补充属性 α_4 和 α_5 后 α 变成 α^F 的条件下, 利用第 2 节中的式(1)一式(3)、内 P-数据推理式(27)、属性集合 α 和 α^F , 由数据集 Y 得到 Y^F ; 或者:

$$\text{if } \alpha \rightarrow \alpha^F, \text{ then } Y^F \Rightarrow Y \quad (35)$$

数据 Y^F 在满足内 P-数据推理时被智能获取。如果属性 α_4 和 α_5 未出现, 则数据 Y^F 不被人们知道。

从这个简单的例子中可以看出:

(1) 在 α 与 α^F 满足关系 $\alpha \subseteq \alpha^F$ 的条件下, 数据集 Y^F 被内 P-数据推理从数据集 Y 内智能分离-挖掘, $Y^F \subseteq Y$ 。在 Y^F 存在的条件下, Q_3 和 Q_5 被智能地从 Q 内分离到 Q 外。

(2) 在 α 与 α^F 满足关系 $\alpha \subseteq \alpha^F$ 的条件下, 利润分布矩阵 A^F 被利润离散分布矩阵 A 智能生成; A 是 A^F 的增广矩阵(augmented matrix)。

(3) 把 α_4 和 α_5 定义为“风险属性”; α_4 和 α_5 入侵到属性集合 α 内, 式(33)中的数据 Y^F 被式(32)中的数据 Y 智能生成, 或者, Y^F 是满足内 P-数据模型即式(27)的 Y 生成。

(4) 在“风险属性” α_4 和 α_5 入侵到 α 内后, 集团公司 Q 的利润离散数据分布集合 Y 变成 Y^F , 使得 Q 的利润产生损失。这个结论在 2011 年 12 月 Q 公布的年度财政报告中得到证明。

5 讨论

P-集合是通过把动态特征引入到有限普通元素集合 X (静态元素集合 X) 内, 改进有限普通元素集合 X 而提出的。文献[18-19]通过改进 P-集合提出了函数 P-集合, 或者, 称函数 P-集合是把动态特征引入有限普通函数集合 S (静态函数集合 S) 内, 改进有限普通函数集合 S 而得到的。P-集合在动态数据(信息)挖掘、动态数据(信息)识别方面获得了应用;

(下转第 195 页)