



OPEN

## Data block decomposition and intelligent secure acquisition of microdata

Xiuquan Zhang<sup>1✉</sup>, Lin Shen<sup>1</sup> & Kaiquan Shi<sup>2</sup>

**P-sets (P stands for Packet) is a set model with dynamic characteristics, which is obtained by introducing dynamic characteristics into Cantor set and improving Cantor set. According to the fact that the characteristics of class I big data are completely consistent with the basic characteristics of P-sets, this paper gives research on theory and application on class I big data from the view of mathematics. Here we introduce Class I big data which need some new definitions of data block, microdata and data link. Based on these concepts, decomposition theorem of data block and microdata relation theorem are given, and then attribute reasoning theorem and microdata intelligent discovery and the intelligent secure acquisition algorithm of microdata are also proposed. By using these theoretical results, the applications of secure acquisition of microdata are presented. In summary, P-sets mathematical model provides a new theory and method for studying class I big data.**

Big data is a new and independent field formed by the mutual penetration and intersection of mathematics, statistics, probability theory, computer technology and humanities. As applications, big data and its attributes are interdependent, and the dynamic changes of big data keep pace with those of its attribute set, which is the dynamic nature of big data. Big data and its sub-data have dynamic characteristics. The attribute  $\alpha_i$  of data element  $x_i$  satisfies the "conjunctive normal form" or "disjunctive normal form" in mathematical logic. The attribute  $\alpha_i$  of class I big data element  $x_i$  meets the characteristic of attribute "conjunction"; meanwhile, the attribute  $\alpha_j$  of class II big data element  $x_j$  meets the attribute "disjunctive" characteristic. Based on these, we divide big data into two types: I and II. Dynamic and attribute characteristics are their common characteristics. Class I big data ( $x$ ) has the following characteristics:

1. ( $x$ ) has dynamic characteristics, some data elements  $x_i$  are deleted from ( $x$ ), and ( $x$ ) becomes  $(x)^{\bar{F}}$ ; whereas other data elements  $x_j$  outside ( $x$ ) are supplemented from outside ( $x$ ) to inside ( $x$ ), ( $x$ ) becomes  $(x)^F$ ;  $(x)^{\bar{F}}$  and  $(x)^F$  together constitute big data ( $x$ ), where  $(x)^{\bar{F}} \subseteq (x)$  and  $(x) \subseteq (x)^F$ ;
2. The attribute set  $\alpha$  of ( $x$ ) satisfies "attribute conjunction" characteristic, the concept of "conjunction" comes from mathematical logic;
3. Data element  $x_j$  coexists with its attribute  $\alpha_j$ ;
4. Big data ( $x$ ) has an important relationship with mathematical concepts and mathematical models.

Characteristics 1–4 are hidden in big data, which cannot be lost or avoided.

Many scholars have devoted themselves to the application research of combining mathematical models with big data, and have achieved some achievements. Based on the mechanism analysis of big data technology in the food supply chain. Li et al.<sup>1</sup> established a basic mathematical model called PEFS to evaluate and optimize the stability of the food system. A prediction model was estimated and evaluated using vector autoregressive model with time series data of long- and short-term interest rates<sup>2</sup>. Based on big data, Fan et al.<sup>3</sup> proposed a grey predictive mathematical model which used to analysis on the effect of Escherichia coli infection on patients with lupus nephritis. However, there is little research on big data features 1–4 in these references, and they focus on combining big data with other disciplines and only give application research, but ignore the theoretical research of big data, pay no attention to the mathematical concept connotation and dynamic characteristics contained in big data, and no one combines big data with mathematics to give theoretical research of big data from the perspective of mathematics.

<sup>1</sup>School of Mathematics and Statistics, Huanghuai University, Zhumadian, China. <sup>2</sup>School of Mathematics and Systems Science, Shandong University, Jinan, China. ✉email: zhangxiuquan@huanghuai.edu.cn

Shi<sup>4</sup> obtained P-sets by introducing dynamic characteristics into Cantor set  $X$  and improving  $X$ . Cantor set  $X$  is given, and  $\alpha$  is attribute set of  $X$ , the concepts and characteristics of P-sets are as follows:

- I Supplementing some attributes in  $\alpha$ , then  $\alpha$  becomes  $\alpha^F, \alpha \subseteq \alpha^F; X$  becomes internal P-sets  $X^{\bar{F}}, X^{\bar{F}} \subseteq X$ ;
- II Deleting some attributes in  $\alpha$ , then  $\alpha$  becomes  $\alpha^{\bar{F}}, \alpha^{\bar{F}} \subseteq \alpha; X$  becomes outer P-sets  $X^F, X \subseteq X^F$ ;
- III If attributes are supplemented and deleted in  $\alpha$  at the same time,  $\alpha$  becomes  $\alpha^F$  and  $\alpha^{\bar{F}}$ , where  $\alpha^{\bar{F}} \subseteq \alpha \subseteq \alpha^F; X$  becomes internal P-sets  $X^{\bar{F}}$  and outer P-sets  $X^F$ , where  $X^{\bar{F}} \subseteq X \subseteq X^F$ ; or  $X$  becomes  $(X^{\bar{F}}, X^F), (X^{\bar{F}}, X^F)$  is the P-sets generated by  $X$ .

The mathematical characteristics of the P-sets such as quantitative characteristics, algebraic characteristics, geometrical characteristics, random characteristics, and theory applications are studied by scholars. By using P-sets, L. A. Zadeh fuzzy set is improved<sup>5</sup>. An algebraic model of P-sets is proposed<sup>6</sup>. By introducing assistant set into function P-sets to expand the function P-sets<sup>7</sup>. More dynamic characteristics and applications of P-sets are discussed<sup>8-10</sup>. P-information fusion and application based on P-sets are obtained<sup>11-14</sup>. The inverse P-sets is proposed<sup>15-21</sup>, which is the dual form of P-sets. Function P-sets is given<sup>22,23</sup>, which is the functional form of P-sets; and the function inverse P-sets is also proposed<sup>24-26</sup>, which is the functional form of inverse P-sets. However, some of these achievements generalize or expand P-sets, while others discuss their applications in dynamic information systems using P-sets as tools, such as intelligent data mining, risk tracking and recognition, and intelligent image recognition. no one has applied the P-sets mathematical model to big data theory and application research in the available literature.

P-sets and class I big data have the same dynamic characteristics and attribute "conjunction" characteristics. In view of I-III, it is easy to get that P-sets is a new mathematical model and method for studying big data. It provides theoretical support for studying the structure, characteristics and basic theory of class I big data.

The purpose of this paper is to give research on theory and application of Class I big data by using the mathematical model with dynamic characteristics.

### Dynamic model with attribute conjunction

Cantor set  $X = \{x_1, x_2, \dots, x_q\} \subset U$  is given, and  $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_k\} \subset V$  is the attribute set of  $X, X^{\bar{F}}$  is referred to as internal P-sets generated by  $X$ , referred to as that  $X^{\bar{F}}$  is internal P-sets for short,

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

$X^-$  is referred to as  $\bar{F}$ -element deleted set of  $X$ ,

$$X^- = \{x_i | x_i \in X, \bar{f}(x_i) = u_i \bar{\in} X, \bar{f} \in \bar{F}\}. \tag{2}$$

If attribute set  $\alpha^F$  of  $X^{\bar{F}}$  meets

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

where in (3),  $\beta_i \in V, \beta_i \bar{\in} \alpha, f \in F$  changes  $\beta_i$  into  $f(\beta_i) = \alpha'_i \in \alpha$ ; in (1),  $X^{\bar{F}} \neq \emptyset, X^{\bar{F}} = \{x_1, x_2, \dots, x_p\}, p < q; p, q \in N^+$ .

Cantor set  $X = \{x_1, x_2, \dots, x_q\} \subset U$  is given, and  $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_k\} \subset V$  is the attribute set of  $X, X^F$  is referred to as outer P-sets generated by  $X$ , referred to as that  $X^F$  is outer P-sets for short,

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

$X^+$  is referred to as  $F$ -element supplemented set of  $X$ ,

$$X^+ = \{u_i | u_i \in U, u_i \bar{\in} X, f(u_i) = x'_i \in X, f \in F\}. \tag{5}$$

If attribute set  $\alpha^{\bar{F}}$  of  $X^F$  meets

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

where in (6),  $\alpha_i \in \alpha, \bar{f} \in \bar{F}$  changes  $\alpha_i$  into  $\bar{f}(\alpha_i) = \beta_i \bar{\in} \alpha, \alpha^{\bar{F}} \neq \emptyset$ ; in (4),  $X^F = \{x_1, x_2, \dots, x_r\}, q < r; q, r \in N^+$ .

The set pair which is composed of internal packet set  $X^{\bar{F}}$  and outer packet set  $X^F$  is referred to as P-sets generated by  $X$ , referred to as P-sets for short and written as

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

From (9) and (11), we get:

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

(8) is referred to as the family of P-sets generated by  $X$ , which is the general expression of P-sets, where both  $I$  and  $J$  are indicator sets.

From (1) to (8), we get:

**Proposition 1** Under the condition of  $F = \bar{F} = \emptyset$ , P-sets  $(X^{\bar{F}}, X^F)$  and Cantor set  $X$  meet:

$$(X^{\bar{F}}, X^F)_{F=\bar{F}=\emptyset} = X. \tag{9}$$

**Proof** 1. if  $F = \bar{F} = \emptyset$ , from (3), we get:  $\alpha^F = \alpha \cup \{\alpha'_i | f(\beta_i) = \alpha'_i \in \alpha, f \in F\}$ .

$= \alpha \cup \emptyset = \alpha, \{\alpha'_i | f(\beta_i) = \alpha'_i \in \alpha, f \in F\} = \emptyset$ ; and  $X^- = \emptyset$  in the Formula (2),  $X^{\bar{F}} = X - X^- = X - \emptyset = X$  in the Formula (1).

2. if  $F = \bar{F} = \emptyset$ , from (6), we get:  $\alpha^{\bar{F}} = \alpha - \{\beta_i | \bar{f}(\alpha_i) = \beta_i \in \alpha, \bar{f} \in \bar{F}\}$

$\alpha - \emptyset = \alpha; \{\beta_i | \bar{f}(\alpha_i) = \beta_i \in \alpha, \bar{f} \in \bar{F}\} = \emptyset$ , and  $X^+ = \emptyset$  in the Formula (5),  $X^F = X \cup X^+ = X \cup \emptyset = X$  in the Formula (4).

Based on 1 and 2, we can complete this Proposition.

**Proposition 2** Under the condition of  $F = \bar{F} = \emptyset$ , the family of P-sets  $\{(X_i^{\bar{F}}, X_i^F) | i \in I, j \in J\}$  and Cantor set  $X$  meet:

$$\{(X_i^{\bar{F}}, X_i^F) | i \in I, j \in J\}_{F=\bar{F}=\emptyset} = X \tag{10}$$

The proof is similar to Proposition 1, and it is omitted.

From Formula (1)–(8) and Proposition 1 and 2, we can easily get the dynamic characteristics of P-sets as follows:

Under the condition we continuously add attributes to  $\alpha$ ,  $X$  generates an internal P-sets  $X_i^{\bar{F}}$ ; similarly we delete attributes continuously in  $\alpha$ ,  $X$  dynamically generates an outer P-sets  $X_i^F$ ; if the attributes are supplemented and deleted at the same time in  $\alpha$ ,  $X$  dynamically generates P-sets  $(X_i^{\bar{F}}, X_i^F), i, j = 1, 2, \dots, n$ .

**Remark** 1.  $U$  is a finite element domain and  $V$  is a finite attribute domain;

2.  $f \in F$  and  $\bar{f} \in \bar{F}$  are element (attribute) transfer;  $F = \{f_1, f_2 \dots f_n\}$  and  $\bar{F} = \{\bar{f}_1, \bar{f}_2, \dots, \bar{f}_n\}$  are the family of element (attribute) transfer, element (attribute) transfer is a concept of function or transformation;

3. The characteristics of  $f \in F$  are that: for element  $u_i \in U, u_i \in X, f \in F$  changes  $u_i$  into  $f(u_i) = x'_i \in X$ ; for attribute  $\beta_i \in V, \beta_i \in \alpha, f \in F$  changes  $\beta_i$  into  $f(\beta_i) = \alpha'_i \in \alpha$ ;

4. The characteristics of  $\bar{f} \in \bar{F}$  are that: for element  $x_i \in X, \bar{f} \in \bar{F}$  changes  $x_i$  into  $\bar{f}(x_i) = u_i \in X$ ; for attribute  $\alpha_i \in \alpha, \bar{f} \in \bar{F}$  changes  $\alpha_i$  into  $\bar{f}(\alpha_i) = \beta_i \in \alpha$ ;

5. The dynamic characteristics of Formula (1) are the same as the dynamic characteristics of down-counter  $T = T - 1$ ;

6. The dynamic characteristics of Formula (4) are the same as the dynamic characteristics of accumulator  $T = T + 1$ . For example, for the Formula (4)  $X_1^F = X \cup X_1^+$ , let  $X = X_1^F, X_2^F = X \cup X_2^+ = (X_1^F \cup X_1^+) \cup X_2^+, \dots$ , so on.

**Fact and evidence of existence of P-sets.**  $X = \{x_1, x_2, x_3, x_4, x_5\}$  is a finite commodity element set of five apples, and  $\alpha = \{\alpha_1, \alpha_2, \alpha_3\}$  is the attribute set confined in  $X$ , where  $\alpha_1$  denotes red color,  $\alpha_2$  denotes sweet taste,  $\alpha_3$  denotes produced by Henan province of China. Obviously,  $x_i$  has attributes  $\alpha_1, \alpha_2$  and  $\alpha_3$ ; the attribute  $\alpha_i$  of  $x_i$  meets "conjunctive normal form",  $\forall x_i \in X, i = 1, 2, \dots, 5$ , moreover

$$\alpha_i = \alpha_1 \wedge \alpha_2 \wedge \alpha_3$$

Let  $\alpha_4$  denotes weight is 150 g, supplementing attribute  $\alpha_4$  in  $\alpha$ ,  $\alpha$  is changed into  $\alpha^F = \{\alpha_1, \alpha_2, \alpha_3\} \cup \{\alpha_4\}$ , and  $X$  is changed into internal P-sets  $X^{\bar{F}} = X - \{x_4, x_5\} = \{x_1, x_2, x_3\}$ . Obviously,  $x_i$  has attributes  $\alpha_1, \alpha_2, \alpha_3$  and  $\alpha_4, \forall x_i \in X^{\bar{F}}, i = 1, 2, 5$ , moreover

$$\alpha_i = (\alpha_1 \wedge \alpha_2 \wedge \alpha_3) \wedge \alpha_4 = \alpha_1 \wedge \alpha_2 \wedge \alpha_3 \wedge \alpha_4.$$

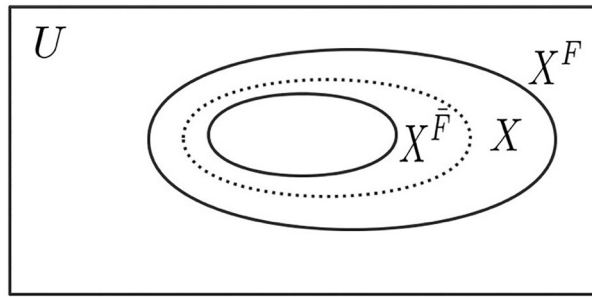
If deleting attribute  $\alpha_3$  in  $\alpha$ ,  $\alpha$  is changed into  $\alpha^{\bar{F}} = \{\alpha_1, \alpha_2, \alpha_3\} - \{\alpha_3\} = \{\alpha_1, \alpha_2\}$ , and  $X$  is changed into outer P-sets  $X^F = X \cup \{x_6, x_7\} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$ . Obviously,  $x_i$  has attributes  $\alpha_1, \alpha_2, \forall x_i \in X^F, i = 1, 2, \dots, 7$ , moreover

$$\alpha_i = (\alpha_1 \wedge \alpha_2 \wedge \alpha_3) - \wedge \alpha_3 = \alpha_1 \wedge \alpha_2$$

This simple fact and logical feature can be accepted by ordinary people. The relationship among  $X^{\bar{F}}, X^F$  and finite ordinary element set  $X$  are shown in Fig. 1.

Agreement: we call that  $(x) = X, (x)^{\bar{F}} = X^{\bar{F}}, (x)^F = X^F, ((x)^{\bar{F}}, (x)^F) = (X^{\bar{F}}, X^F)$ ; in "Dynamic model with attribute conjunction" section;  $(x), (x)^{\bar{F}}, (x)^F$  and  $((x)^{\bar{F}}, (x)^F)$  are data, element  $x_i$  in  $X$  is data element of  $X$ , which will be used directly in "Data block decomposition and microdata generation" section—"Microdata intelligent-security acquisition algorithm" section.

By using characteristics and concept of P-sets in "Dynamic model with attribute conjunction" section, we get "Data block decomposition and microdata generation" section.



**Figure 1.**  $X^{\bar{F}}$  and  $X^F$  are the internal P-sets and outer P-sets generated by  $X$  respectively,  $X^{\bar{F}}$  and  $X^F$  constitute P-Sets  $(X^{\bar{F}}, X^F)$ ;  $X^{\bar{F}}$  and  $X^F$  are represented by solid lines, and finite ordinary element set  $X$  is represented by dotted lines,  $U$  is a finite element domain.

### Data block decomposition and microdata generation

**Definition 1** If  $(x)^{\bar{F}}$  and  $(x)^F$  satisfy,

$$(x)^{\bar{F}} = \{(x)^{\bar{F}}_i | (x)^{\bar{F}}_{i+1} \subseteq (x)^{\bar{F}}_i, \alpha_i^{\bar{F}} \subseteq \alpha_{i+1}^{\bar{F}}; i = 1, 2, \dots, n\} \tag{11}$$

$$(x)^F = \{(x)^F_j | (x)^F_j \subseteq (x)^F_{j+1}, \alpha_j^{\bar{F}} \subseteq \alpha_{j+1}^{\bar{F}}; j = 1, 2, \dots, n - 1\} \tag{12}$$

then  $(x)^{\bar{F}}$  and  $(x)^F$  are called  $\alpha^F$ -data block and  $\alpha^{\bar{F}}$ -data block of  $(x)$  respectively. Otherwise,  $\alpha_i^{\bar{F}}$  is the attribute set of  $(x)^{\bar{F}}_i$ ;  $(x)^{\bar{F}}_i$  is the  $\alpha^{\bar{F}}$ -block of  $(x)^{\bar{F}}$ ;  $\alpha_j^{\bar{F}}$  is the attribute set of  $(x)^F_j$ ;  $(x)^F_j$  is the  $\alpha^{\bar{F}}$ -block of  $(x)^F$ ; where  $(x)^{\bar{F}}_i \in (x)^{\bar{F}}$ ,  $(x)^F_j \in (x)^F$ .

**Definition 2** We call  $(x)$  is big data composed of  $\alpha^F$ -data block  $(x)^{\bar{F}}$  and  $\alpha^{\bar{F}}$ -data block  $(x)^F$  respectively, and,

$$\begin{aligned} (x) &= (x)^{\bar{F}} \cup (x)^F \\ &= \left( \bigcup_{i=1}^n (x)^{\bar{F}}_i \right) \cup \left( \bigcup_{j=1}^n (x)^F_j \right) \end{aligned} \tag{13}$$

**Definition 3** Under the condition of supplementing attributes in  $\alpha_k^{\bar{F}}$ , if  $\nabla(x)^{\bar{F}}_{k+1}$ ,  $(x)^{\bar{F}}_k$  and  $(x)^-$  satisfy,

$$\nabla(x)^{\bar{F}}_{k+1} = (x)^{\bar{F}}_k - (x)^- \tag{14}$$

then  $\nabla(x)^{\bar{F}}_{k+1}$  is called  $\alpha_k^{\bar{F}}$ -microdata generated by  $(x)^{\bar{F}}_k$ , where  $(x)^-$  is composed of deleted data elements  $x_i$  in  $(x)^{\bar{F}}_k$ ;  $\nabla(x)^{\bar{F}}_{k+1} \subseteq (x)^{\bar{F}}_k$ .

**Definition 4** Under the condition of deleted attributes in  $\alpha_k^{\bar{F}}$ , if  $\Delta(x)^F_{k+1}$ ,  $(x)^F_k$  and  $(x)^+$  satisfy,

$$\Delta(x)^F_{k+1} = (x)^F_k \cup (x)^+ \tag{15}$$

then  $\Delta(x)^F_{k+1}$  is called  $\alpha_k^{\bar{F}}$ -microdata generated by  $(x)^F_k$ , where  $(x)^+$  is composed of data element  $x_j$  supplemented into  $(x)^F_k$ ; where  $\Delta(x)^F_{k+1} \supseteq (x)^F_k$ .

From (11) to (15), we get:

**Theorem 1** (Decomposition Theorem of internal  $\alpha^F$ -data block).

The sufficient and necessary conditions for  $(x)^{\bar{F}}_k$  is the  $\alpha^F$ -data block decomposed from  $(x)$  are: there is  $\Delta\alpha \neq \emptyset$ , so that the attribute set  $\alpha_k^{\bar{F}}$  of  $(x)^{\bar{F}}_k$  and attribute sets  $\alpha$  of  $(x)$  satisfy

$$\alpha_k^{\bar{F}} - (\alpha \cup \Delta\alpha) = \emptyset \tag{16}$$

where  $\Delta\alpha \cap \alpha = \emptyset$ .

**Proof** From (1)–(3) in "Dynamic model with attribute conjunction" section, we get:

1. if attribute set  $\Delta\alpha$  is supplemented by attribute set  $\alpha_k^F$  of  $(x)_k^F$  and attribute set  $\alpha$  of  $(x)$ , and it satisfies  $\alpha_k^F = \alpha \cup \Delta\alpha$ , or  $\alpha_k^F - (\alpha \cup \Delta\alpha) = \emptyset$ , then  $(x)_k^F$  is the internal decomposition of  $(x)$ ,  $(x)_k^F \subseteq (x)$ .

2. If  $(x)_k^F$  is the internal decomposition of  $(x)$ , then  $\alpha_k^F = \alpha \cup \Delta\alpha$ , or  $\alpha_k^F - (\alpha \cup \Delta\alpha) = \emptyset$ . Among them,  $\Delta\alpha$  is the attribute set added to  $\alpha$ ,  $\Delta\alpha \neq \emptyset$ . Theorem 1 is proved from 1 and 2.  $\square$

**Theorem 2** (Decomposition Theorem outer  $\alpha^F$ -data block).

The sufficient and necessary conditions for  $(x)_k^F$  is the  $\alpha^F$ -data block decomposed from  $(x)$  are: there is  $\nabla\alpha \neq \emptyset$ , so that the attribute set  $\alpha_k^F$  of  $(x)_k^F$  and attribute set  $\alpha$  of  $(x)$  satisfy

$$\alpha_k^F - (\alpha - \nabla\alpha) = \emptyset \tag{17}$$

where  $\nabla\alpha \cap \alpha \neq \emptyset$ .

Its proof is similar to Theorem 1, and it is omitted.

**Theorem 3** The necessary and sufficient conditions for the existence of  $\alpha_{k+1}^F$ -microdata  $\nabla(x)_{k+1}^F$  are: the attribute set  $\alpha_k^F$  of  $(x)_k^F$  and the attribute set  $\alpha_{k+1}^F$  of  $\nabla(x)_{k+1}^F$  satisfy.

$$\text{card}(\alpha_{k+1}^F) - \text{card}(\alpha_k^F) > 0 \tag{18}$$

**Proof** From Formula (1) to (3) and Formula (14) in "Dynamic model with attribute conjunction" section, we get:  $\nabla(x)_{k+1}^F = (x)_k^F - (x)^- = (x)_{k+1}^F$  if  $\alpha_k^F$  and  $\alpha_{k+1}^F$  are attribute sets of  $(x)_k^F$  and  $\nabla(x)_{k+1}^F$  respectively, then  $\alpha_k^F \subseteq \alpha_{k+1}^F$ , or  $\text{card}(\alpha_{k+1}^F) - \text{card}(\alpha_k^F) > 0$ .  $\square$

**Theorem 4** The necessary and sufficient conditions for the existence of  $\alpha_k^F$ -microdata  $\Delta(x)_{k+1}^F$  are: the attribute set  $\alpha_k^F$  of  $(x)_k^F$  and the attribute set  $\alpha_{k+1}^F$  of  $\Delta(x)_{k+1}^F$  satisfy.

$$\text{card}(\alpha_{k+1}^F) - \text{card}(\alpha_k^F) < 0 \tag{19}$$

Its proof is similar to Theorem 3, and it is omitted, where in (18) and (19),  $\text{card} = \text{cardinal number}$ .

**Theorem 5** ( $\alpha^F$ -microdata relation Theorem) If there is  $k$ , then  $\alpha_k^F$ -microdata  $\nabla(x)_k^F$  meets.

$$\bigcup_{i=1}^{k-1} \nabla(x)_i^F \subseteq \nabla(x)_k^F \subseteq \bigcup_{i=k+1}^n \nabla(x)_i^F \tag{20}$$

**Theorem 6** ( $\alpha^F$ -microdata relation Theorem) If there is  $\lambda$ , then  $\alpha_\lambda^F$ -microdata  $\Delta(x)_\lambda^F$  meets.

$$\bigcup_{j=1}^{\lambda-1} \Delta(x)_j^F \subseteq \Delta(x)_\lambda^F \subseteq \bigcup_{j=\lambda+1}^n \Delta(x)_j^F \tag{21}$$

Propositions 3 and 4 are obtained from (11) to (21):

**Propositions 3**  $\alpha^F$ -data block  $(x)_i^F$  constitutes  $\alpha^F$ -data block chain of  $(x)$  as follows:

$$\{(x)_n^F, \alpha_n^F\} \leftarrow \{(x)_{n-1}^F, \alpha_{n-1}^F\} \leftarrow \dots \leftarrow \{(x)_2^F, \alpha_2^F\} \leftarrow \{(x)_1^F, \alpha_1^F\} \tag{22}$$

**Propositions 4**  $\alpha^F$ -data block  $(x)_j^F$  constitutes  $\alpha^F$ -data block chain of  $(x)$  as follows:

$$\{(x)_1^F, \alpha_1^F\} \leftarrow \{(x)_2^F, \alpha_2^F\} \leftarrow \dots \leftarrow \{(x)_{n-1}^F, \alpha_{n-1}^F\} \leftarrow \{(x)_n^F, \alpha_n^F\} \tag{23}$$

By utilizing the concepts and the model in "Dynamic model with attribute conjunction" section, and the theoretical results in "Data block decomposition and microdata generation" section, we will give "Attribute reasoning and intelligent discovery of microdata" section.

### Attribute reasoning and intelligent discovery of microdata

Suppose  $\nabla(x)_{k+1}^F$  is the  $\alpha_k^F$ -microdata generated by  $(x)_k^F$ , and  $\alpha_{k+1}^F$  and  $\alpha_k^F$  are their attribute sets respectively, if they meet

$$\text{if } \alpha_k^F \Rightarrow \alpha_{k+1}^F, \text{ then } \nabla(x)_{k+1}^F \Rightarrow (x)_k^F \tag{24}$$

(24) is referred to as attribute reasoning generated by  $\alpha^F$ -microdata  $\nabla(x)_{k+1}^F$ ,  $\alpha_k^F \Rightarrow \alpha_{k+1}^F$  is referred to as reasoning condition,  $\nabla(x)_{k+1}^F \Rightarrow (x)_k^F$  is referred to as inference conclusion.

Suppose  $\Delta(x)_{k+1}^F$  is the  $\alpha_k^{\bar{F}}$ -microdata generated by  $(x)_k^F$ , and  $\alpha_{k+1}^{\bar{F}}$  and  $\alpha_k^{\bar{F}}$  are their attribute sets respectively, if they meet

$$\text{if } \alpha_{k+1}^{\bar{F}} \Rightarrow \alpha_k^{\bar{F}}, \text{ then } (x)_k^F \Rightarrow \Delta(x)_{k+1}^F \tag{25}$$

(25) is referred to as attribute reasoning generated by  $\alpha^{\bar{F}}$ -microdata  $\Delta(x)_{k+1}^F, \alpha_{k+1}^{\bar{F}} \Rightarrow \alpha_k^{\bar{F}}$  is referred to as reasoning condition,  $(x)_k^F \Rightarrow \Delta(x)_{k+1}^F$  is referred to as inference conclusion. where in (24) and (25), " $\Rightarrow$ " is equal to " $\subseteq$ ".

If  $(\nabla(x)_{k+1}^F, (x)_k^F), ((x)_k^{\bar{F}}, \Delta(x)_{k+1}^F)$  and  $(\alpha_k^F, \alpha_{k+1}^{\bar{F}}), (\alpha_{k+1}^F, \alpha_k^{\bar{F}})$  meet

$$\text{if } (\alpha_k^F, \alpha_{k+1}^{\bar{F}}) \Rightarrow (\alpha_{k+1}^F, \alpha_k^{\bar{F}}), \text{ then } (\nabla(x)_{k+1}^F, (x)_k^F) \Rightarrow ((x)_k^{\bar{F}}, \Delta(x)_{k+1}^F) \tag{26}$$

(26) is referred to as attribute reasoning generated by  $(\alpha^F, \alpha^{\bar{F}})$ -microdata  $(\nabla(x)_{k+1}^F, \Delta(x)_{k+1}^F), (\alpha_k^F, \alpha_{k+1}^{\bar{F}}) \Rightarrow (\alpha_{k+1}^F, \alpha_k^{\bar{F}})$  is referred to as reasoning condition,  $(\nabla(x)_{k+1}^F, (x)_k^F) \Rightarrow ((x)_k^{\bar{F}}, \Delta(x)_{k+1}^F)$  is referred to as inference conclusion. Where (26) represents if  $\alpha_k^F \Rightarrow \alpha_{k+1}^F$ , then  $\nabla(x)_{k+1}^F \Rightarrow (x)_k^{\bar{F}}$ ; if  $\alpha_{k+1}^{\bar{F}} \Rightarrow \alpha_k^{\bar{F}}$ , then  $(x)_k^F \Rightarrow \Delta(x)_{k+1}^F$ .

From (24) to (26), we get:

**Theorem 7** ( $\alpha^F$ -microdata  $\nabla(x)_{k+1}^{\bar{F}}$  intelligent decomposition-discovery Theorem) Under the condition of reasoning (24),  $\alpha^F$ -microdata  $\nabla(x)_{k+1}^{\bar{F}}$  is intelligently decomposed—discovered from  $\alpha^{\bar{F}}$ -data block  $(x)_k^{\bar{F}}$ , or,

1.  $\nabla(x)_{k+1}^{\bar{F}}$  and  $(x)_k^{\bar{F}}$  satisfy

$$\nabla(x)_{k+1}^{\bar{F}} \cap (x)_k^{\bar{F}} \neq \phi \tag{27}$$

2. The separation coefficient  $\eta_{k+1}^{\bar{F}}$  of  $\nabla(x)_{k+1}^{\bar{F}}$  separated from  $(x)_k^{\bar{F}}$  satisfies

$$\eta_{k+1}^{\bar{F}} - 1 < 0 \tag{28}$$

where  $\eta_{k+1}^{\bar{F}} = \text{card}(\nabla(x)_{k+1}^{\bar{F}} / \text{card}((x)_k^{\bar{F}})), 1 = \text{card}((x)_k^{\bar{F}} / \text{card}(x)_k^{\bar{F}}) = \eta = 1$  is the self separation coefficient of  $(x)_k^{\bar{F}}$ .

**Proof** 1. From Formula (1)–(3) in "Dynamic model with attribute conjunction" section and Formula (14) in "Data block decomposition and microdata generation" section, we get:  $\nabla(x)_{k+1}^{\bar{F}} = (x)_k^{\bar{F}} - (x)^- = (x)_{k+1}^{\bar{F}}, \nabla(x)_{k+1}^{\bar{F}}$  and  $(x)_k^{\bar{F}}$  meet  $\nabla(x)_{k+1}^{\bar{F}} \subseteq (x)_k^{\bar{F}}$ , or  $\nabla(x)_{k+1}^{\bar{F}} \cup (x)_k^{\bar{F}} \neq \emptyset$ , we get (27); or  $\alpha^F$ -microdata  $\nabla(x)_{k+1}^{\bar{F}}$  is intelligently decomposed-discovered from  $\alpha^{\bar{F}}$ -data block  $(x)_k^{\bar{F}}$ .

2. Because  $\text{card}(\nabla(x)_{k+1}^{\bar{F}}) < \text{card}((x)_k^{\bar{F}})$  so  $\eta_{k+1}^{\bar{F}} = \text{card}(\nabla(x)_{k+1}^{\bar{F}} / \text{card}((x)_k^{\bar{F}})) < \text{card}((x)_k^{\bar{F}} / \text{card}(x)_k^{\bar{F}}) = \eta = 1$ , or  $\eta_{k+1}^{\bar{F}} - 1 < 0$ , and we get (28). □

**Theorem 8** ( $\alpha^{\bar{F}}$ -microdata  $\Delta(x)_{k+1}^F$  intelligent decomposition-discovery Theorem) Under the condition of reasoning (25),  $\alpha^{\bar{F}}$ -microdata  $\Delta(x)_{k+1}^F$  is intelligently decomposed—discovered outside  $\alpha^F$ -data block  $(x)_k^F$ , or,

1.  $\Delta(x)_{k+1}^F$  and  $(x)_k^F$  satisfy

$$\Delta(x)_{k+1}^F \cap (x)_k^F \neq \phi \tag{29}$$

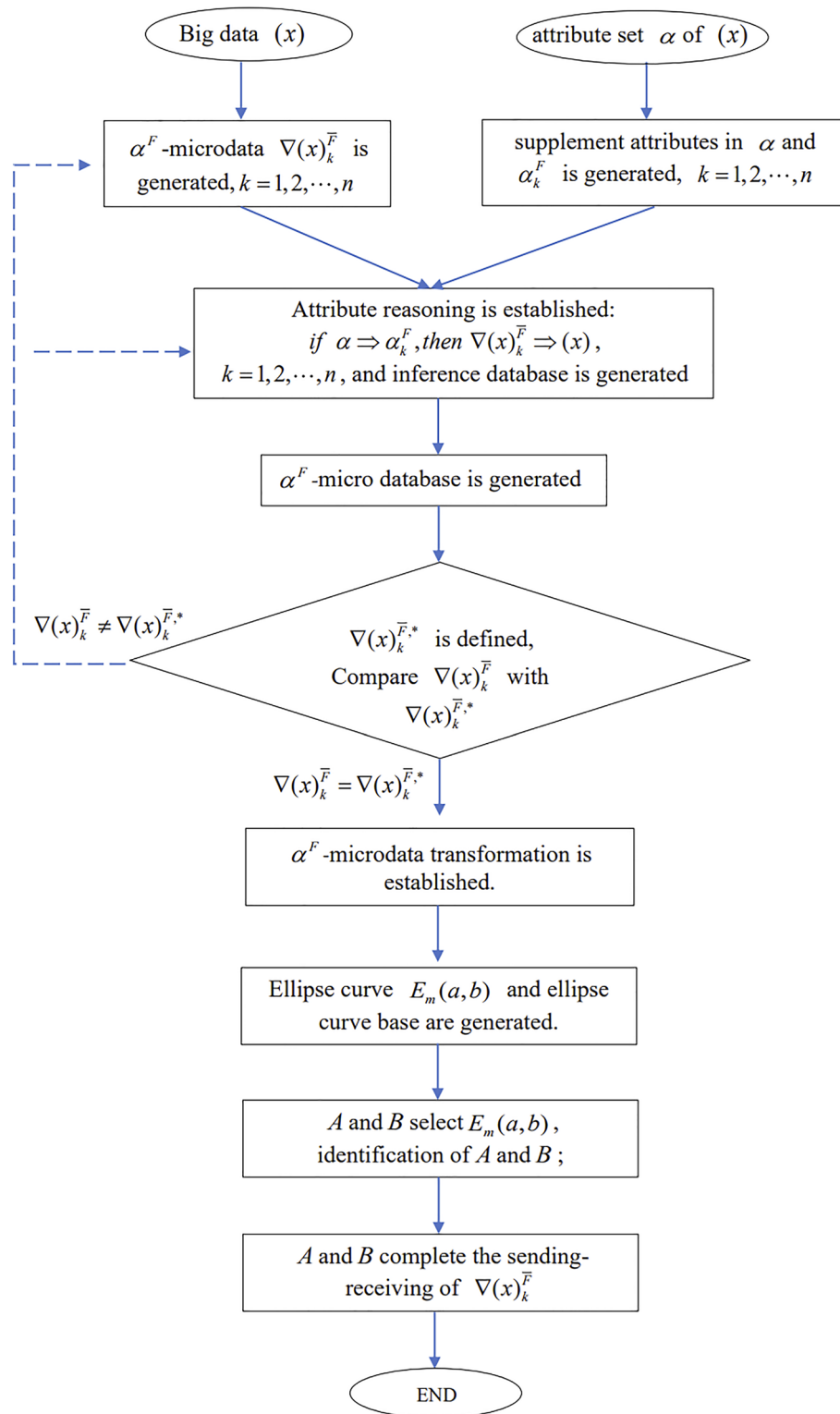
2. The separation coefficient  $\eta_{k+1}^F$  of  $\Delta(x)_{k+1}^F$  separated outside  $(x)_k^F$  satisfies

$$\eta_{k+1}^F - 1 > 0 \tag{30}$$

where  $\eta_{k+1}^F = \text{card}(\Delta(x)_{k+1}^F) / \text{card}((x)_k^F), 1 = \text{card}(x)_k^F / \text{card}((x)_k^F) = \eta = 1$  is the Self separation coefficient of  $(x)_k^F$ .

Its proof is similar to Theorem 7, and will be omitted. The Propositions 5 is a direct conclusion of Theorems 7 and 8.

**Propositions 5** Under the condition of reasoning (26),  $\alpha^F$ -microdata  $\nabla(x)_{k+1}^{\bar{F}}$  and  $\alpha^{\bar{F}}$ -microdata  $\Delta(x)_{k+1}^F$  are intelligently decomposed—discovered within  $(x)_k^{\bar{F}}$  and outside  $(x)_k^F$  respectively.



**Figure 2.**  $\alpha^F$ -microdata intelligent-security acquisition algorithm.

### Microdata intelligent-security acquisition algorithm

In this section, only the intelligent decomposition and secure acquisition algorithms of  $\alpha^F$ -microdata are given, it is a part of  $(\alpha^F, \alpha^{\bar{F}})$ -microdata intelligent decomposition and secure acquisition algorithm; the complete  $(\alpha^F, \alpha^{\bar{F}})$ -microdata intelligent decomposition and secure acquisition algorithm is omitted. The  $\alpha^F$ -microdata intelligent decomposition—security acquisition algorithm is shown in Fig. 2.

The detailed process of the algorithm is as follows:

- (1) Big data  $(x)$  and its attribute set  $\alpha$  are given, which are the initial values of the algorithm;
- (2)  $\alpha^F$ -microdata  $\nabla(x)_k^{\bar{F}}$  is generated; supplement attributes in  $\alpha$  and  $\alpha_k^F$  is generated,  $k = 1, 2, \dots, n$ ;
- (3) Attribute reasoning is established: if  $\alpha \Rightarrow \alpha_k^F$ , then  $\nabla(x)_k^{\bar{F}} \Rightarrow (x)$ ,  $k = 1, 2, \dots, n$ , and inference database is generated;
- (4)  $\alpha^F$ -micro database is generated;
- (5) Given the standard microdata  $\nabla(x)_k^{\bar{F},*}$ , Compare  $\nabla(x)_k^{\bar{F}}$  with  $\nabla(x)_k^{\bar{F},*}$ , if  $\nabla(x)_k^{\bar{F}} \neq \nabla(x)_k^{\bar{F},*}$ , then return to (2) and (3) and start cycle; if  $\nabla(x)_k^{\bar{F}} = \nabla(x)_k^{\bar{F},*}$ , then go to the next step;
- (6)  $\alpha^F$ -microdata transformation is established;
- (7) Ellipse curve  $E_m(a, b)$  and ellipse curve base are generated;
- (8)  $A$  and  $B$  select  $E_m(a, b)$ , identification of  $A$  and  $B$ ;
- (9)  $A$  and  $B$  complete the sending-receiving of  $\nabla(x)_k^{\bar{F}}$ ;
- (10) The algorithm ends.

By using the preliminary concept in "Dynamic model with attribute conjunction" section, the theoretical results are deduced from "Data block decomposition and microdata generation" section and "Attribute reasoning and intelligent discovery of microdata" section and the intelligent security acquisition algorithm in "Microdata intelligent-security acquisition algorithm" section, and we get "Microdata encryption–decryption and its secure acquisition-application" section.

### Microdata encryption–decryption and its secure acquisition-application

The example in this section comes from  $\alpha^F$ -data block  $(x)_t^{\bar{F}}$  of "commodity big data  $(x)$ ",  $t = 1, 2, \dots, n$ ; the name of  $(x)_t^{\bar{F}}$  is "commodity profit estimation";  $x_i \in (x)_t^{\bar{F}}$  is the commodity,  $\alpha_t^F$  is the attribute set of  $(x)_t^{\bar{F}}$ , and  $\alpha_i \in \alpha_t^F$  is the attribute of  $x_i$ ; for convenience, the names of  $x_i$  and  $\alpha_i$  are omitted.

$$(x)_t^{\bar{F}} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\} \tag{31}$$

$$\alpha_t^F = \{\alpha_1, \alpha_2, \alpha_3\} \tag{32}$$

$\alpha_i = \alpha_1 \wedge \alpha_2 \wedge \alpha_3$  is the attribute of  $x_i$ ,  $y_1 \sim y_8$  is the sales market of  $x_1 \sim x_8$ , and  $y_1 \sim y_8$  is distributed in 8 areas of the city.

**Profit Estimation of Commodity  $x_i$ .** Make  $\lambda$  surveys on the profit value of  $x_i$ , and use the statistical methods to get Table 1.

The "—" in Table 1 indicates that the profit valuation  $\xi_i$  is between 2 and 3;  $\xi_1 = 7, \xi_4 = 9, \xi_5 = 10$  and  $\xi_8 = 9$  are 7%, 9%, 10% and 9% respectively. The commodity profit investigator obtains two new attributes  $\alpha_4$  and  $\alpha_5$  of  $x_1, x_4, x_5$  and  $x_8$  while obtaining  $\xi_i$ , where  $\alpha_4$  stands for that  $x_1, x_4, x_5$  and  $x_8$  installments are allowed;  $\alpha_5$  stands for that  $x_1, x_4, x_5$  and  $x_8$  allowing return or exchange after purchase. When both  $\alpha_4$  and  $\alpha_5$  exist,  $\alpha_t^F$  in (32) generates  $\alpha_{t+1}^F$ .

$$\alpha_{t+1}^F = \alpha_t^F \cup \{\alpha_4, \alpha_5\} = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5\} \tag{33}$$

**Intelligent Generation of  $\alpha_{t+1}^F$ -microdata  $\nabla(x)_{t+1}^{\bar{F}}$ .** By using (24) in "Attribute reasoning and intelligent discovery of microdata" section, we get:

$$\text{if } \alpha_t^F \Rightarrow \alpha_{t+1}^F, \text{ then } \nabla(x)_{t+1}^{\bar{F}} \Rightarrow (x)_t^{\bar{F}} \tag{34}$$

(34) gives

$$\begin{aligned} \nabla(x)_{t+1}^{\bar{F}} &= (x)_t^{\bar{F}} - \{x_1, x_3, x_6, x_7\} \\ &= \{x_1, x_4, x_5, x_8\} \end{aligned}$$

The attribute of  $\forall x_i \in (x)_{t+1}^{\bar{F}}$  is  $\alpha_i = (\alpha_1 \wedge \alpha_2 \wedge \alpha_3) \wedge \alpha_4 \wedge \alpha_5, i = 1, 4, 5, 8$ .

By using Table 1, the profit values composition points  $(\xi_1, \xi_4) = (7, 9)$  and  $(\xi_5, \xi_8) = (10, 9)$  of  $x_1, x_4, x_5$  and  $x_8$  are obtained.

**Secure acquisition application of  $\alpha_{k+1}^F$ -microdata  $\nabla(x)_{k+1}^{\bar{F}}$ .** Ellipse curve model and its encryption decryption algorithm

$x_i$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$
$\xi_i$	7	—	—	9	10	—	—	9

**Table 1.**  $\xi_i$  is the pre estimate of  $x_i$  profit,  $i = 1, 2, \dots, 8$ .



Call  $E(k) = \{(x, y) | x, y \in k, p(x, y) = 0\} \cup \{\mathcal{O}\}$  is an ellipse curve, which is established by the following equation proposed by Koblitz<sup>27</sup>:

$$y^2 = x^3 + ax + b \tag{35}$$

where discriminant  $D = (4a^3 + 27b^2) \bmod m \neq 0, k$  is a number field,  $a, b \in k; \mathcal{O}$  is a point at infinity,  $m$  is a prime number. Equation (35) is a simplified form of weistrass equations  $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ .

Taken  $P(x_1, y_1), Q(x_1, y_1) \in E(k)$  arbitrarily,  $R(x_3, y_3)$  are points generated by  $P(x_1, y_1) \oplus Q(x_1, y_1)$  in  $E(k)$  arbitrarily,  $\oplus$  is the dot add operation in  $E(k), R(x_3, y_3)$  is obtained as follows:

1. If  $P(x_1, y_1) \neq Q(x_2, y_2)$ , then there are

$$\begin{aligned} x_3 &= (\lambda^2 - x_1 - x_2) \bmod m, \\ y_3 &= (\lambda(x_1 - x_3) - y_1) \bmod m, \\ \lambda &= \frac{y_2 - y_1}{x_2 - x_1}. \end{aligned} \tag{36}$$

2. If  $P(x_1, y_1) = Q(x_2, y_2)$ , then there are

$$\begin{aligned} x_3 &= (\lambda^2 - x_1 - x_2) \bmod m \\ y_3 &= (\lambda(x_1 - x_3) - y_1) \bmod m, \\ \lambda &= \frac{3x_1^2 + a}{2y_1}. \end{aligned} \tag{37}$$

Let  $P_m$  be the plaintext of  $\alpha_{k+1}^F$ -microdata  $\nabla(x)_{k+1}^F$ , and  $C_m$  be the ciphertext of  $P_m$ .  $A$  is the encryptor of plaintext  $P_m$  and the sender of ciphertext  $C_m$ ;  $B$  is the decryptor of ciphertext  $C_m$  and the accepter of plaintext  $P_m, n_A$  is the private key to  $A, P_A = n_A G$  is the public key of  $A; n_B$  is the private key to  $B, P_B = n_B G$  is the public key of  $B; P_A$  is given to  $B$  publicly and  $P_B$  is given to  $A$  publicly.  $A$  and  $B$  choose the same basic point  $G \in E(k), n_A, n_B \in N^+$ .

I Information encryption process:  $A$  chooses  $P_m = \nabla(x)_{k+1}^F, k \in N^+$ , the public key of  $A$  and  $B$  is  $P_B$ , then  $A$  gives the ciphertext  $C_m$  of  $P_m$  as follows:

$$C_m = \{kG, P_m + kP_B\} \tag{38}$$

Then  $A$  sends  $C_m$  to  $B$ .

II Information decryption process:  $B$  accepts  $C_m$ , if  $n_B(kG) = C_m^1 P_m + kP_B = C_m^2$ , then  $B$  obtains plaintext  $P_m$  from ciphertext  $C_m$

$$\begin{aligned} C_m^2 - C_m^1 &= P_m + kP_B - n_B(kG) \\ &= P_m + kP_B - k(n_B G) \\ &= P_m + kP_B - kP_B = P_m \end{aligned} \tag{39}$$

where  $P_m$  is the point composed of profit values  $\xi_1, \xi_4, \xi_5$  and  $\xi_8$  of  $x_1, x_4, x_5$ , and  $x_8: (\xi_1, \xi_4) = (7, 9), (\xi_5, \xi_8) = (10, 9)$ .

Using  $(\xi_1, \xi_4), (\xi_5, \xi_8)$  to determine ellipse curve  $y^2 = x^3 + ax + b \bmod m = x^3 + x + 6 \bmod 11$ , Get point set:  $E_m(a, b) = \{(2, 4), (2, 7), (3, 5), (3, 6), (5, 2), (5, 9), (7, 2), (7, 9), (8, 3), (8, 8), (10, 2), (10, 9)\}$ ; obviously:  $(\xi_1, \xi_4) = (7, 9)$  and  $(\xi_5, \xi_8) = (10, 9) \in E_m(a, b)$ .

$A$  chooses  $k = 3$ , the basic point is  $G = (2, 7), n_B = 7, P_B = n_B G$  is the private key and public key of  $B$  respectively.

1.  $A$  chooses  $P_m = (7, 9) = 8G$ , assume  $k = 3$  and the public key of  $G$  and  $B$  is  $P_B, A$  gives the ciphertext  $C_m$  of  $P_m$ :

$$\begin{aligned} C_m &= \{kG, P_m + kP_B\} = \{6G, 8G + 3(7 \times 2G)\} \\ &= \{6G, 8G + 42G\} = \{6G, 50G\} \\ &= \{6G, 2G\} = \{(5, 9), (2, 7)\} \end{aligned} \tag{40}$$

2. After  $B$  accepts  $C_m$ , use  $n_B$  to get  $P_m$

$$\begin{aligned} C_m^2 - C_m^1 &= P_m + kP_B - n_B(kG) \\ &= 2G - 7(3G) = 2G - 42G \\ &= -40G = -4G = 8G = (7, 9) = P_m \end{aligned} \tag{41}$$

The encryption–decryption of  $P_m = (10, 9) = 12G$  is similar to (40) and (41), and which will be omitted.

### Important description of application examples

1. The application example of the paper is taken from the "commodity profit estimation" block in "commodity big data", which is the compression and simplification of the original case. Because of the special nature of the example, some information can not be made public. The profit value  $\xi_i$  of commodity  $x_i$  is an important secret of company C; the profit value  $\xi_i$  should be safe, should not be disclosed, and should not be stolen by others. Company C shall safely receive the  $\xi_i$  which sent by market investigators and shall make the decision on commodity  $x_i$  access to the market.

2. In the competition of commodity market, if the profit value  $\xi_i$  of  $x_i$  is stolen by others, then company C\*, which produces similar commodity  $x_i$ , will quickly occupies the market, thus company C will lose the market and the profit of the commodity. Therefore, the security of data transmission is particularly important, the encryption and decryption process in the information security algorithm ensures the security transmission of information.

3. The security of ellipse curve cryptography comes from the difficulty of solving discrete logarithms, point set  $E_m(k)$  and dot add operation  $\oplus$  defined on  $E_m(k)$  form Abel Group  $\langle E_m(k), \oplus \rangle$ . In the example, we only use  $n_B$  and  $P_B$  of B, and the  $n_A$  and  $P_A$  of A are omitted.

### Confirmation of application examples

The applications and the methods of the examples have been applied and confirmed in 8 business districts in Zhengzhou, China. The profits  $\xi_2, \xi_3, \xi_6$  and  $\xi_7$  of  $x_2, x_3, x_6$  and  $x_7$  in Table 1 are between 2 and 3;  $y_2, y_3, y_6$  and  $y_7$  were closed by company C in July 2020.

## Discussion

Big data is a branch of theory and application research with mathematical concept connotation, which has infiltrated into many application researches and aroused people's interest. What is the structure and characteristics of big data? Can it be classified? What is the basis for classification? In the existing literature, these problems have not been discussed. In this paper, the theory and application of Class I big data are presented by using the p-sets mathematical model with dynamic characteristics and attribute "Conjunction" characteristics. Restore the mathematical characteristics of big data with mathematical methods and models, and understand the mathematical structure and logical characteristics of big data. These characteristics are hidden in big data and do not attract people's attention. Finally, the paper selects the "financial security" data block in the "financial big data" as an example, uses the big data block theory given in this paper, and combines the elliptic curve to give the security algorithm of financial information encryption and decryption.

In the research of big data, most scholars combine mathematical models with big data to carry out application research, rarely pay attention to the theoretical research of big data, who do not recognize the dynamic and logical characteristics of big data, which lack strict reasoning and proof. In this paper, the author not only gives the theoretical research of big data by using mathematical models, but also pays attention to the combination of theoretical research with application research. Perhaps the research and results presented in this paper provide a research idea and mathematical method for researchers who are engaged in the fundamental theory and application of big data. The infiltration of mathematical concepts into the study of big data may make people acquire new theoretical understanding of big data, and it becomes an inevitable process from the simple study application to the theoretical study of big data.

In the inverse P-sets, attribute  $\alpha_j$  of element  $x_j$  conforms to the attribute "disjunctive" feature. Inverse P-sets model shares the same characteristics of as class II big data, which is a mathematical model and method to study class II big data. Applying inverse P-sets to study the characteristics and application of class II big data is the aim which we will meet in the future. Applying for inverse P-sets, The application and characteristics of class II big data is the aim which we will put forward in the future.

Function P-sets and function inverse P-sets are mathematical models and methods to study class I big data in function form and class II big data in function form respectively. All of P-sets, inverse P-sets, function P-sets, function inverse P-sets are the preparation for studying the application of big data, understanding the characteristics of big data, and giving mathematical models as well as methods.

### Data availability

The datasets used and analysed during the current study available from the corresponding author on reasonable request.

Received: 19 April 2022; Accepted: 26 March 2023

Published online: 04 April 2023

## References

- Li, Z. S., Huang, Y. & Li, Z. J. Stability evaluation and optimization of food system using PEFS mathematical model and big data technology. *J. Phys. Conf. Ser.* **1952**(4), 042113–042123 (2021).
- Kim, J. H. A study on estimation and prediction of vector time series model using financial big data (interest rates). *TURCOMAT* **12**(5), 309–316 (2021).
- Fan, M. X. *et al.* Big data-based grey forecast mathematical model to evaluate the effect of *Escherichia coli* infection on patients with *Lupus nephritis*. *J. Results Phys.* **26**(1), 104339–104354 (2021).
- Shi, K. Q. P-sets and its applications. *J. Adv. Syst. Sci. Appl.-Ions.* **9**, 209–219 (2009).
- Shi, K. Q. & Li, S. W. Separated fuzzy set  $(A^F, A^F)$  and the intelligent fusion of fuzzy information. *J. Shandong Univ. Nat. Sci.* **57**(07), 1–13 (2022) (in Chinese)
- Li, X. C. An algebraic model of P-sets. *J. Shangqiu Norm. Univ.* **36**, 1–5 (2020) (in Chinese).
- Yu, X. Q., Xu, F. S. & J. N. Function  $p(\sigma, \tau)$ -set and its characteristics. *J. Jilin Univ. Nat. Sci.* **56**, 53–59 (2018) (in Chinese).
- Fan, C. X. & Lin, K. K. P-sets and the reasoning-identification of disaster information. *Int. J. Conver. Inf. Technol.* **7**, 337–345 (2012).

9. Lin, H. K. & Fan, C. X. The dual form of P-reasoning and identification of unknown attribute. *Int. J. Digit. Content Technol. Appl.* **6**, 121–131 (2012).
10. Zhang, X. Q. P-augmented matrix and its application in dynamic tracking recognition. *J. Anhui Univ. Nat. Sci.* **46**(3), 53–58 (2022) (in Chinese).
11. Tang, J. H., Zhang, L. & Shi, K. Q. Outer P-information law reasoning and its application in intelligent fusion and separating of information law. *J. Microsyst. Technol.* **24**(10), 4389–4398 (2018).
12. Zhang, X. Q., Zhang, J. Y. & Shi, K. Q. Dynamic boundary characteristics of P-sets and information dynamic fusion generation. *J. Xinyang Norm. Univ.* **35**(3), 364–368 (2022) (in Chinese).
13. Hao, X. M. & Li, N. N. Quantitative characteristics and applications of P-information hidden mining. *J. Shandong Univ. Nat. Sci.* **54**, 9–14 (2019) (in Chinese).
14. Liu, J. Q. & Zhang, H. Y. Information P-dependence and P-dependence mining-sieving. *J. Comput. Sci.* **45**, 202–206 (2018) (in Chinese).
15. Shi, K. Q. Inverse P-sets. *J. Shandong Univ. Nat. Sci.* **47**(1), 98–109 (2012) (in Chinese).
16. Zhang, X. Q., Shen, L. & Shi, K. Q. ( $\alpha^F$ ,  $\alpha^F$ )-information fusion generated by information segmentation and its intelligent retrieval. *J. Math.* **10**(5), 713 (2022).
17. Fan, C. X. & Huang, S. L. Inverse P-reasoning discovery identification of Inverse P-information. *Int. J. Digit. Content Technol. Appl.* **6**(20), 735–744 (2012).
18. Lin, K. K. & Fan, C. X. Embedding camouflage of inverse P-information and application. *Int. J. Conver. Inf. Technol.* **7**, 471–480 (2012).
19. Yu, X. Q. & Xu, F. S. Random inverse packet information and its acquisition. *J. Appl. Math. Nonlinear Sci.* **5**, 357–366 (2020).
20. Li, S. W. & Shi, K. Q. Inverse separated fuzzy set on of fuzzy information and secure acquisition. *J. Shandong Univ. Nat. Sci.* **57**(09), 1–14 (2022) (in Chinese).
21. Zhang, L. & Ren, X. F. The relationship between abnormal information system and inverse P-augmented matrices. *J. Shandong Univ. Nat. Sci.* **54**(9), 15–21 (2019) (in Chinese).
22. Shi, K. Q. Function P-Sets. *Int. J. Mach. Learn. Cybern.* **2**, 281–288 (2011).
23. Chen, B. H. & Zhang, L. Attribute relations of data compound-decomposition and data intelligent acquisition. *J. Fuzzy Syst. Math.* **35**(3), 167–174 (2021) (in Chinese).
24. Shi, K. Q. Function inverse P-sets and information law fusion. *J. Shandong Univ. Nat. Sci.* **47**, 73–80 (2012) (in Chinese).
25. Zhang, L., Ren, X. F. & Shi, K. Q. Inverse p-matrix reasoning model-based the intelligent dynamic separation and acquisition of educational information. *Microsyst. Technol.* **24**, 4415–4421 (2018).
26. Tang, J. H., Chen, B. H., Zhang, L. & Bai, X. R. Function inverse P-sets and the dynamic separation of inverse P-information laws. *J. Shandong Univ. Nat. Sci.* **48**, 104–110 (2013) (in Chinese).
27. Koblitz, N. Elliptic curve crypto systems. *J. Math Comput.* **48**(17), 203–209 (1987).

### Author contributions

Conceptualization, X.Z.; methodology, X.Z.; writing—original draft preparation, X.Z.; writing—review and editing, L.S.; software, L.S.; investigation and supervision, K.S.; validation and funding acquisition, K.S. All authors have read and agreed to the published version of the manuscript.

### Funding

This work was sponsored by the National Natural Science Foundation of China (NSFC) under Grant 12171193, and the Fund of Young Backbone Teacher in Henan Province under Grant 2021GGJS158, Key project of Henan Education Department (No. 23B110012), Research and Development Fund of Huanghuai University (XKPY-2022013).

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to X.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023