计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.191000072

大数据分解-融合及其智能获取

刘纪芹[1]　　史开泉[2]
1 山东财经大学数学与数量经济学院　济南 250014
2 山东大学数学学院　济南 250100
(sdfiljq@126.com)

摘　要　文中给出了通过大数据分解、融合生成的大数据分解-融合以及大数据距离；利用这些概念，给出了大数据并-交分解定理以及大数据交-并分解定理与它们的属性合取关系、大数据融合的智能生成定理与大数据融合的距离关系、大数据分解-融合的识别准则与大数据分解-融合获取的智能算法与算法的过程，以及这些理论结果在大数据分解-融合智能获取的应用。文中给出了∧型大数据新的特征，∧型大数据是利用P-集合模型得到的。

关键词：大数据；分解-融合；识别准则；智能算法

中图法分类号　O144,TP391

Big Data Decomposition-Fusion and Its Intelligent Acquisition

LIU Ji-qin[1] and SHI Kai-quan[2]
1 School of Mathematics and Quantitative Economics, Shandong University of Finance and Economics, Jinan 250014, China
2 School of Mathematics, Shandong University, Jinan 250100, China

Abstract　The concepts of big data decomposition-fusion and big data distance generated by big data decomposition and fusion are given. By using these concepts, an union-intersection decomposition theorem of big data, an intersection-union decomposition theorem of big data and their attribute conjunction relation are given. Intelligent generation theorems and the distance relationship of big data fusion are given. A recognition criterion of big data decomposition-fusion, an intelligent algorithm and algorithm process of big data decomposition-fusion acquisition are given. The application of these theoretical results in big data decomposition-fusion intelligent acquisition is presented. In this paper, the new characteristics of ∧-type big data are given, ∧-type big data is obtained by using P-sets model.

Keywords　Big data, Decomposition-Fusion, Recognition criterion, Intelligent algorithm

1　引言

本文是文献[1]的研究的继续。针对文献[1]中∧型大数据的特征，给出了新的讨论，本文简称∧型大数据为大数据，以便不引起混乱与误解。

文献[2-20]给出了多个关于大数据的理论与应用研究。无论来自哪个应用领域的大数据都满足如下特征：(1)大数据与它的属性相伴存在；(2)不同的时刻 $t\in T$，大数据的状态是不相同的，或者大数据的状态跟随 $t\in T$ 的变化有了动态特征；(3)大数据的数据元 x_i 的属性 α_i 满足一定的逻辑关系。特征(1)－特征(3)与P-集合(packet sets)的基本特征完全吻合。P-集合[1,21-67]成为研究大数据特征-应用的数学工具与数学方法。

对文献[1]中的结果再认识，得到：(1)给定大数据的基数据(x)与它的属性集合 α。1)若在 k 时刻 $k\in T$，α 内补充属性，α 生成 α_k^F，$\alpha\subseteq\alpha_k^F$，则$(x)$生成$(x)_k^{\overline{F}}$，$(x)_k^{\overline{F}}\subseteq(x)$。或者连续在 α 内补充属性，α 生成 $\alpha_1^F,\alpha_2^F,\cdots,\alpha_{n-1}^F,\alpha_n^F$；$\alpha\subseteq\alpha_1^F\subseteq\alpha_2^F\subseteq\cdots\subseteq\alpha_{n-1}^F\subseteq\alpha_n^F$；$(x)$连续生成$(x)_1^{\overline{F}},(x)_2^{\overline{F}},\cdots,(x)_{n-1}^{\overline{F}},(x)_n^{\overline{F}}$；$(x)_n^{\overline{F}}\subseteq(x)_{n-1}^{\overline{F}}\subseteq\cdots\subseteq(x)_2^{\overline{F}}\subseteq(x)_1^{\overline{F}}\subseteq(x)$。换言之，在确定的 $k\in T$ 的条件下，$(x)_k^{\overline{F}}$ 是(x)的一个分解，或者称$(x)_k^{\overline{F}}$ 是(x)的 $\overline{F}$-分解。2)若在 $k\in T$ 时刻，α 内删除属性，α 生成 $\alpha_k^{\overline{F}}$，$\alpha_k^{\overline{F}}\subseteq\alpha$，则$(x)$生成$(x)_k^F$，$(x)\subseteq(x)_k^F$；或者连续在 α 内删除属性，α 生成 $\alpha_1^{\overline{F}},\alpha_2^{\overline{F}},\cdots,\alpha_{n-1}^{\overline{F}},\alpha_n^{\overline{F}}$；$\alpha_n^{\overline{F}}\subseteq\alpha_{n-1}^{\overline{F}}\subseteq\cdots\subseteq\alpha_2^{\overline{F}}\subseteq\alpha_1^{\overline{F}}\subseteq\alpha$；$(x)$连续生成$(x)_1^F,(x)_2^F,\cdots,(x)_{n-1}^F,(x)_n^F$；$(x)\subseteq(x)_1^F\subseteq(x)_2^F\subseteq\cdots\subseteq(x)_{n-1}^F\subseteq(x)_n^F$。换言之，在确定的 $k\in T$ 的条件下，$(x)_k^F$ 是(x)的一个分解，或者称$(x)_k^F$ 是(x)的 F-分解。事实上，1)与2)存在于大数据的动态分析与应用中，由(1)中给出的认识得

到稿日期：2019-10-12　返修日期：2019-12-12　　本文已加入开放科学计划(OSID)，请扫描上方二维码获取补充信息。
基金项目：国家社会科学基金项目(71663010)；山东省自然科学基金项目(zr2013aq019)
This work was supported by the National Social Science Foundation of China (71663010) and Natural Science Foundation of Shandong Province, China (zr2013aq019).
通信作者：史开泉(shikq@sdu.edu.cn)

到：(2)中(x)生成的$\overline{F}$-分解$(x)_k^{\overline{F}}$的本质是(x)内的一些冗余数据元x_i构成$\nabla(x)$，$\nabla(x)$是从(x)内被融合到(x)外得到的；或者$\overline{F}$-分解生成数据外-融合$(x)_k^{\overline{F}}=(x)-\nabla(x)$，或者外-融合$(x)_k^{\overline{F}}$是$(x)$内删除$\nabla(x)$的剩余。$(x)$生成的$F$-分解$(x)_k^F$的本质是$(x)$内缺失的一些数据元$x_j$构成$\Delta(x)$，$\Delta(x)$是从$(x)$外被融合到$(x)$内得到的；或者$F$-分解生成数据内-融合$(x)_k^F=(x)\cup\Delta(x)$，或者内-融合$(x)_k^F$是$(x)$内补充$(x)$内的缺失$\Delta(x)$的生成。显然，(1)中的$\overline{F}$-分解、$F$-分解各自生成交叉-复合概念：$\overline{F}$-分解与外-融合，$F$-分解与内-融合。交叉-复合概念为认识-挖掘大数据的特征与这些特征的应用给出概念的支持。如果引入推理概念：if $A_k^{\overline{F}}\Rightarrow A$，then $(x)_k^{\overline{F}}\Rightarrow(x)$；if $A\Rightarrow A_k^F$，then $(x)\Rightarrow(x)_k^F$，则得到智能获取$\overline{F}$-分解与外-融合$(x)_k^{\overline{F}}$的方法，得到智能获取F-分解与内-融合$(x)_k^F$的方法，智能获取算法被生成。

本文的主要结果如下：给出了大数据分解、分解与融合交叉-合成概念，即分解-融合；给出了大数据距离；利用这些概念给出了大数据并-交分解定理、大数据交-并分解定理与它们的属性合取关系；给出了大数据融合的智能生成定理与大数据融合的距离关系；给出了大数据分解-融合的识别准则与大数据分解-融合获取的智能算法与算法过程；给出了这些理论结果在大数据智能分析-识别与获取的应用。

2 相关概念

2.1 P-集合及其特征

给定有限普通元素集合$X=\{x_1,x_2,\cdots,x_q\}\subset U$，$\alpha=\{\alpha_1,\alpha_2,\cdots,\alpha_k\}\subset V$是$X$的属性集合，称$X^{\overline{F}}$是$X$生成的内P-集合(internal packet set)，简称$X^{\overline{F}}$是内P-集合。

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

X^{-}称作X的$\overline{F}$-元素删除集合。

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

如果α^F是$X^{\overline{F}}$的属性集合，则：

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

式(3)中，$\beta_i\in V$，$\beta_i\overline{\in}\alpha$，$f\in F$把$\beta_i$生成$f(\beta_i)=\alpha'_i\in\alpha$；式(1)中，$X^{\overline{F}}\neq\emptyset$，$X^{\overline{F}}=\{x_1,x_2,\cdots,x_p\}$，$p<q$；$p,q\in\mathbf{N}^+$。

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

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

X^{+}称作X的F-元素补充集合。

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

如果$\alpha^{\overline{F}}$是X^F的属性集合，则：

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

式(6)中，$\alpha_i\in\alpha$，$\overline{f}\in\overline{F}$把$\alpha_i$生成$\overline{f}(\alpha_i)=\beta_i\overline{\in}\alpha$；$\alpha^{\overline{F}}\neq\emptyset$；式(4)中，$X^F=\{x_1,x_2,\cdots,x_r\}$，$q<r$，$q,r\in\mathbf{N}^+$。

由内P-集合00$X^{\overline{F}}$与外P-集合X^F构成的有限普通元素集合对，称作X生成的P-集合(P=Packet)，简称P-集合：

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

有限普通元素集合X称作P-集合$(X^{\overline{F}},X^F)$的基集合(基础集合)。

α内不断补充属性，由式(3)给出：

$$\alpha_1^F\subseteq\alpha_2^F\subseteq\cdots\subseteq\alpha_{n-1}^F\subseteq\alpha_n^F\tag{8}$$

得到满足式(8)的内P-集合：

$$X_n^{\overline{F}}\subseteq X_{n-1}^{\overline{F}}\subseteq\cdots\subseteq X_2^{\overline{F}}\subseteq X_1^{\overline{F}}\tag{9}$$

α内不断删除属性，由式(6)给出：

$$\alpha_n^{\overline{F}}\subseteq\alpha_{n-1}^{\overline{F}}\subseteq\cdots\subseteq\alpha_2^{\overline{F}}\subseteq\alpha_1^{\overline{F}}\tag{10}$$

得到满足式(10)的外P-集合：

$$X_1^F\subseteq X_2^F\subseteq\cdots\subseteq X_{n-1}^F\subseteq X_n^F\tag{11}$$

在α内不断补充属性的同时删除属性，由式(9)、式(11)得到：

$$\{(X_i^{\overline{F}},X_j^F)\mid i\in I,j\in J\}\tag{12}$$

式(12)称作X生成的P-集合族，式(12)是P-集合的一般形式。

P-集合$(X^{\overline{F}},X^F)$是把动态特征引入到有限普通元素集合X内，改进有限普通元素集合X得到的具有动态特征的新集合模型。在什么条件下，P-集合$(X^{\overline{F}},X^F)$能被还原成X？命题1、命题2给出了问题的结论。

命题1 在$F=\overline{F}=\emptyset$的条件下，P-集合$(X^{\overline{F}},X^F)$与X满足：

$$(X^{\overline{F}},X^F)_{F=\overline{F}=\emptyset}=X\tag{13}$$

命题2 在$F=\overline{F}=\emptyset$的条件下，P-集合族$\{(X_i^{\overline{F}},X_j^F)\mid i\in I,j\in J\}$与$X$满足：

$$\{(X_i^{\overline{F}},X_j^F)\mid i\in I,j\in J\}_{F=\overline{F}=\emptyset}=X\tag{14}$$

命题1、命题2的证明是简单的，证明略。

2.2 P-增广矩阵结构与生成

给定有限普通元素集合$X=\{x_1,x_2,\cdots,x_q\}$，$\forall x_i\in X$具有n个元素值$y_{1,i},y_{2,i},\cdots,y_{n,i}$；$\boldsymbol{y}_i=(y_{1,i},y_{2,i},\cdots,y_{n,i})^{\mathrm{T}}$是$y_{1,i},y_{2,i},\cdots,y_{n,i}$生成的向量，$i=1,2,\cdots q$；以$\boldsymbol{y}_i$为列得到矩阵$\boldsymbol{A}$，$\boldsymbol{A}$称作$X$生成的元素值矩阵。

$$\boldsymbol{A}=\begin{pmatrix}y_{1,1}&y_{1,2}&\cdots&y_{1,q}\\y_{2,1}&y_{2,2}&\cdots&y_{2,q}\\\vdots&\vdots&\ddots&\vdots\\y_{n,1}&y_{n,2}&\cdots&y_{n,q}\end{pmatrix}\tag{15}$$

给定内P-集合$X^{\overline{F}}=\{x_1,x_2,\cdots,x_p\}$，$\boldsymbol{A}^{\overline{F}}$称作$X^{\overline{F}}$生成的$\boldsymbol{A}$的内P-增广矩阵。

$$\boldsymbol{A}^{\overline{F}}=\begin{pmatrix}y_{1,1}&y_{1,2}&\cdots&y_{1,p}\\y_{2,1}&y_{2,2}&\cdots&y_{2,p}\\\vdots&\vdots&\ddots&\vdots\\y_{n,1}&y_{n,2}&\cdots&y_{n,p}\end{pmatrix}\tag{16}$$

给定外P-集合$X^F=\{x_1,x_2,\cdots,x_r\}$，$\boldsymbol{A}^F$称作$X^F$生成的$\boldsymbol{A}$的外P-增广矩阵。

$$\boldsymbol{A}^F=\begin{pmatrix}y_{1,1}&y_{1,2}&\cdots&y_{1,r}\\y_{2,1}&y_{2,2}&\cdots&y_{2,r}\\\vdots&\vdots&\ddots&\vdots\\y_{n,1}&y_{n,2}&\cdots&y_{n,r}\end{pmatrix}\tag{17}$$

由内P-增广矩阵$\boldsymbol{A}^{\overline{F}}$与外P-增广矩阵$\boldsymbol{A}^F$构成的矩阵对，称作P-集合$(X^{\overline{F}},X^F)$生成的$\boldsymbol{A}$的P-增广矩阵。

$$(\boldsymbol{A}^{\overline{F}},\boldsymbol{A}^F)\tag{18}$$

称

$$\{(\boldsymbol{A}_i^{\overline{F}},\boldsymbol{A}_j^F)\,|\,i\in I,j\in J\} \tag{19}$$

是 P-集合族$\{(X_i^{\overline{F}},X_j^F)\,|\,i\in I,j\in J\}$生成的 $\boldsymbol{A}$ 的 P-增广矩阵族，式(19)是 P-增广矩阵的一般形式。

式(15)－式(17)中，$p<q<r,p,q,r\in\mathbf{N}^+$。

特别说明：(1)外 P-增广矩阵 $\boldsymbol{A}^F$ 与 $\boldsymbol{A}$ 的普通增广矩阵 $\boldsymbol{A}^*$ 是同一个概念，或者 $\boldsymbol{A}^F=\boldsymbol{A}^*$。(2)在普通数学中，$\boldsymbol{A}$ 是 $m\times n$ 阶矩阵，如果在 n 列中补充 λ 列得到 $\boldsymbol{A}_{m\times(n+\lambda)}$，则 $\boldsymbol{A}_{m\times(n+\lambda)}$ 被定义成 $\boldsymbol{A}$ 的增广矩阵。如果在 n 列中删去 t 列，$t<n$，得到 $\boldsymbol{A}_{m\times(n-t)}$，$\boldsymbol{A}_{m\times(n-t)}$ 是否是 $\boldsymbol{A}$ 的增广矩阵？在普通数学中 $\boldsymbol{A}_{m\times(n-t)}$ 没有定义，但 $\boldsymbol{A}_{m\times(n-t)}$ 在信息系统应用研究中常常遇到。基于这个事实，利用 P-集合$(X^{\overline{F}},X^F)$的结构与动态特性，改进普通增广矩阵的定义得到内 P-增广矩阵 $\boldsymbol{A}^{\overline{F}}$。与此类似，得到 P-增广矩阵$(\boldsymbol{A}^{\overline{F}},\boldsymbol{A}^F)$。(3)内 P-增广矩阵 $\boldsymbol{A}^{\overline{F}}$、外 P-增广矩阵 $\boldsymbol{A}^F$、P-增广矩阵$(\boldsymbol{A}^{\overline{F}},\boldsymbol{A}^F)$是利用 P-集合的结构与动态特征，改进普通增广矩阵 $\boldsymbol{A}^*$ 的概念、结构得到增广矩阵新概念、新结构。在一定条件下，$\boldsymbol{A}^{\overline{F}}$，$\boldsymbol{A}^F$ 与$(\boldsymbol{A}^{\overline{F}},\boldsymbol{A}^F)$被还原成普通增广矩阵 $\boldsymbol{A}^*$。

应当指出：(1)P-集合的存在事实、逻辑特征与 P-集合的更多特征与应用，见文献[1,21-67]。(2) P-增广矩阵的更多讨论、特征与应用，见文献[49]。

2.3 ∧型大数据结构与特征

给定(x)，α 是(x)的属性集合，∧是属性的合取运算；$\rho^{\overline{F}}\in(0,1)$是 X 内元素被迁移到(x)外的可能性(概率值)，$\rho^F\in(0,1)$是 X 外的元素被迁移到(x)内的可能性(概率值)，称

$$\{((x)_i^{\overline{F}},(x)_j^F)\,|\,i\in I,j\in J:\alpha,\wedge,(\rho^{\overline{F}},\rho^F)\} \tag{20}$$

是(x)生成的具有合取特征的大数据，简称∧型大数据；称$(x)_i^{\overline{F}}$，$(x)_j^F$ 分别是∧型大数据的合取前项、合取后项。

式(20)中，$(x)=X$，$(x)_i^{\overline{F}}=X_i^{\overline{F}}$，$(x)_j^F=X_j^F$，$((x)_i^{\overline{F}},(x)_j^F)=(X_i^{\overline{F}},X_j^F)$。∧型大数据的更多特征与应用见文献[1]。

利用第 2 节中的预备概念，针对文献[1]给出的结果，第 3 节—第 5 节给出了新的研究。

约定：在第 3 节－第 5 节中，文献[1]中的∧型大数据简称大数据；∧型大数据前项$(x)^{\overline{F}}$、后项$(x)^F$ 分别称为$\overline{F}$-数据$(x)^{\overline{F}}$、F-数据$(x)^F$。内 P-增广矩阵 $\boldsymbol{A}^{\overline{F}}$、外 P-增广矩阵 $\boldsymbol{A}^F$、P-增广矩阵$(\boldsymbol{A}^{\overline{F}},\boldsymbol{A}^F)$分别称作内 P-矩阵 $\boldsymbol{A}^{\overline{F}}$、外 P-矩阵 $\boldsymbol{A}^F$、P-矩阵$(\boldsymbol{A}^{\overline{F}},\boldsymbol{A}^F)$。

在基数据(x)的属性集合 α 变化的条件下，$\alpha\subseteq\alpha_1^F\subseteq\alpha_2^F\subseteq\cdots\subseteq\alpha_{n-1}^F\subseteq\alpha_n^F$；$(x)$生成$\overline{F}$-数据：$(x)_n^{\overline{F}}\subseteq(x)_{n-1}^{\overline{F}}\subseteq\cdots\subseteq(x)_2^{\overline{F}}\subseteq(x)_1^{\overline{F}}\subseteq(x)$；显然，$\forall(x)_i^{\overline{F}}$ 是(x)的一个分解。在基数据(x)的属性集合 α 变化的条件下，$\alpha_n^{\overline{F}}\subseteq\alpha_{n-1}^{\overline{F}}\subseteq\cdots\subseteq\alpha_2^{\overline{F}}\subseteq\alpha_1^{\overline{F}}\subseteq\alpha$，$(x)$生成 F-数据：$(x)\subseteq(x)_1^F\subseteq(x)_2^F\subseteq\cdots\subseteq(x)_{n-1}^F\subseteq(x)_n^F$；显然，$\forall(x)_j^F$ 是(x)的一个分解。引入时间参数 $t\in T$，$(x)_t^{\overline{F}}$ 是(x)在 $t\in T$ 时刻生成的$\overline{F}$-数据分解；$(x)_t^F$ 是(x)在 $t\in T$ 时刻生成的 F-数据分解。

3 大数据分解定理与大数据分解的属性关系

定理 1(大数据并-交分解定理) 若 α，α_i^F 分别是基数据(x)、$\overline{F}$-数据$(x)_i^{\overline{F}}$ 的属性集合，则：

$$(x)=\bigcup_{\substack{\alpha\subseteq\alpha_i^F\\ i\in I}}(x)_i^{\overline{F}} \tag{21}$$

$$\alpha=\bigcap_{\substack{\alpha\subseteq\alpha_i^F\\ i\in I}}\alpha_i^F \tag{22}$$

证明：由式(9)得到：$(x)_n^{\overline{F}},(x)_{n-1}^{\overline{F}},\cdots,(x)_2^{\overline{F}},(x)_1^{\overline{F}}$ 是(x)在式(8)条件下的生成，满足$(x)_n^{\overline{F}}\subseteq(x)_{n-1}^{\overline{F}}\subseteq\cdots\subseteq(x)_2^{\overline{F}}\subseteq(x)_1^{\overline{F}}\subseteq(x)$；$\forall k$，$(x)_k^{\overline{F}}$ 是满足 $\alpha\subseteq\alpha_k^F$ 条件下的分解；显然$(x)=(x)_n^{\overline{F}}\cup(x)_{n-1}^{\overline{F}}\cup\cdots\cup(x)_2^{\overline{F}}\cup(x)_1^{\overline{F}}\cup(x)$，得到式(21)。由式(8)知：$\alpha,\alpha_1^F,\alpha_2^F,\cdots,\alpha_{n-1}^F,\alpha_n^F$ 满足 $\alpha\subseteq\alpha_1^F\subseteq\alpha_2^F\subseteq\cdots\subseteq\alpha_{n-1}^F\subseteq\alpha_n^F$，则有 $\alpha=\alpha\cap\alpha_1^F\cap\alpha_2^F\cap\cdots\cap\alpha_{n-1}^F\cap\alpha_n^F$，得到式(22)。

推论 1 $\forall t,x_i\in(x)_p^{\overline{F}}$ 的属性 α_i 满足属性合取扩展。

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

其中，$\alpha=\{\alpha_1,\alpha_2,\cdots,\alpha_k\}$是$(x)$的属性集合，$\alpha_p^F=\{\alpha_1,\alpha_2,\cdots,\alpha_k,\alpha_{k+1},\cdots,\alpha_\lambda\}$是$(x)_p^{\overline{F}}$ 的属性集合。

推论 2 存在$\nabla(x)\neq\varnothing$，(x)的并分解冗余$\nabla(x)$与$(x)_p^{\overline{F}}$，(x)满足：

$$(x)_p^{\overline{F}}=(x)-\nabla(x) \tag{24}$$

定理 2(大数据交-并分解定理) 若 α，$\alpha_j^{\overline{F}}$ 分别是基数据(x)、F-数据$(x)_j^F$ 的属性集合，则：

$$(x)=\bigcap_{\substack{\alpha_j^{\overline{F}}\subseteq\alpha\\ j\in J}}(x)_j^F \tag{25}$$

$$\alpha=\bigcup_{\substack{\alpha_j^{\overline{F}}\subseteq\alpha\\ j\in J}}\alpha_j^{\overline{F}} \tag{26}$$

定理 2 可由式(11)、式(10)得到，证明略。

推论 3 $\forall t,x_j\in(x)_q^F$ 的属性 α_j 满足属性合取收缩。

$$\alpha_j=(\bigwedge_{t=1}^{k}\alpha_t)-\bigwedge_{t=\eta+1}^{k}\alpha_t \tag{27}$$

其中，$\alpha=\{\alpha_1,\alpha_2,\cdots,\alpha_\eta,\alpha_{\eta+1},\cdots,\alpha_k\}$是$(x)$的属性集合，$\alpha_q^{\overline{F}}=\{\alpha_1,\alpha_2,\cdots,\alpha_\eta\}$是$(x)_q^F$ 的属性集合。

推论 4 存在 $\Delta(x)\neq\varnothing$，(x)的交分解的补充 $\Delta(x)$与$(x)_q^F$，(x)满足：

$$(x)_q^F=(x)\cup\Delta(x) \tag{28}$$

定理 3 $\overline{F}$-数据分解$(x)_k^{\overline{F}}$ 与 F-数据分解$(x)_k^F$ 构成的分解对$((x)_k^{\overline{F}},(x)_k^F)$满足：

$$((x)_k^{\overline{F}},(x)_k^F)=(\bigcap_{\substack{\alpha\subseteq\alpha_k^F\\ k\in I}}(x)_k^{\overline{F}},\bigcup_{\substack{\alpha_k^{\overline{F}}\subseteq\alpha\\ k\in J}}(x)_k^F) \tag{29}$$

$\overline{F}$-数据分解$(x)_k^{\overline{F}}$ 的属性集合 α_k^F 与 F-数据分解$(x)_k^F$ 的属性集合 $\alpha_k^{\overline{F}}$ 构成的属性集合对$(\alpha_k^F,\alpha_k^{\overline{F}})$满足：

$$(\alpha_k^F,\alpha_k^{\overline{F}})=(\bigcup_{\substack{\alpha\subseteq\alpha^F\\ k\in I}}\alpha_k^F,\bigcap_{\substack{\alpha_k^{\overline{F}}\subseteq\alpha\\ k\in J}}\alpha_k^{\overline{F}}) \tag{30}$$

式(29)中，$(x)_k^{\overline{F}}=\bigcap_{\substack{\alpha\subseteq\alpha_k^F\\ k\in I}}(x)_k^{\overline{F}}$，$(x)_k^F=\bigcup_{\substack{\alpha_j^{\overline{F}}\subseteq\alpha\\ k\in J}}(x)_k^F$；式(30)中，$\alpha_k^F=\bigcup_{\substack{\alpha\subseteq\alpha^F\\ k\in I}}\alpha_k^F$，$\alpha_k^{\overline{F}}=\bigcap_{\substack{\alpha^{\overline{F}}\subseteq\alpha\\ k\in J}}\alpha_k^{\overline{F}}$。

推论 5 $\forall x_i\in(x)_p^{\overline{F}}$ 的属性 α_i，$\forall x_j\in(x)_q^F$ 的属性 α_j 满足属性合取扩展-收缩。

$$(\alpha_i,\alpha_j)=((\bigwedge_{t=1}^{k}\alpha_t)\bigwedge_{t=k+1}^{\lambda}\alpha_t,(\bigwedge_{t=1}^{k}\alpha_t)-\bigwedge_{t=\eta+1}^{k}\alpha_t) \tag{31}$$

推论 6 存在$\nabla(x)_k\neq\varnothing$，$\Delta(x)_k\neq\varnothing$，$((x)_{k+1}^{\overline{F}},(x)_{k+1}^F)$与$((x)_k^{\overline{F}}-\nabla(x)_k,(x)_k^F\cup\Delta(x)_k)$满足：

$$((x)_{k+1}^{\overline{F}},(x)_{k+1}^F)=((x)_k^{\overline{F}}-\nabla(x)_k,(x)_k^F\cup\Delta(x)_k) \tag{32}$$

式(31)中，$\alpha_i=(\bigwedge_{t=1}^{k}\alpha_t)\bigwedge_{t=k+1}^{\lambda}\alpha_t$；$\alpha_j=(\bigwedge_{t=1}^{k}\alpha_t)-\bigwedge_{t=\eta+1}^{k}\alpha_t$；式(32)中，$(x)_{k+1}^{\bar F}=(x)_k^{\bar F}-\nabla(x)_k$，$(x)_{k+1}^{F}=(x)_k^{F}\cup\Delta(x)_k$。

第3节中的概念与结果给出启迪：(x)的分解$(x)_k^{\bar F}$，$(x)_k^{\bar F}\subset(x)$是(x)内的$\nabla(x)\subset(x)$从(x)内被融合到(x)外得到的，$\nabla(x)$是(x)生成$(x)_k^{\bar F}$的冗余。(x)的分解$(x)_k^F$，$(x)\subset(x)_k^F$是(x)外的$\Delta(x)$，$\Delta(x)\cap(x)=\varnothing$从$(x)$外被融合到$(x)$内得到的，$\Delta(x)$是$(x)$生成$(x)_k^F$的补充。这是在大数据动态分析-识别中常见的事实。启迪指出：基数据(x)的并分解式(21)、基数据(x)的交分解式(25)生成一个新的复合概念：分解-融合。分解-融合是对具有动态特征的大数据的数学分析。显然，分解-融合是潜藏在大数据动态分析与应用的一个重要概念。如果定义$(x)_k^{\bar F}$是(x)生成的数据融合，则(x)与$(x)_k^{\bar F}$之间存在距离$\sigma_k^{\bar F}$；定义$(x)_k^F$是(x)生成的数据融合，则(x)与$(x)_k^F$之间存在距离σ_k^F。显然，$(x)_k^{\bar F}$的存在决定于$\nabla(x)$与$\sigma_k^{\bar F}$；$(x)_k^F$的存在决定于$\Delta(x)$与σ_k^F。如果$(x)_k^{\bar F}$与(x)满足推理条件，$(x)_k^{\bar F}$从(x)内被获取；$(x)_k^F$与(x)满足推理条件，$(x)_k^F$从(x)外被获取。因此，$(x)_k^{\bar F}$与(x)之间的距离$\sigma_k^{\bar F}$成为识别$(x)_k^{\bar F}$的度量尺度，$(x)_k^F$与(x)之间的距离σ_k^F成为识别$(x)_k^F$的度量尺度。根据这个启迪，得到第4节的内容。

4 大数据智能融合与融合智能识别尺度

给定基数据(x)，$\exists x_i\in(x)$，$\bar f\in\bar F$把x_i生成$\bar f(x_i)=x_i'\overline{\in}(x)$，$x_i'$构成$\nabla(x)_k$。若$(x)$，$\nabla(x)_k$与$\bar F$-数据分解$(x)_k^{\bar F}$满足：

$$(x)_k^{\bar F}=(x)-\nabla(x)_k \tag{33}$$

则称$(x)_k^{\bar F}$是(x)生成的数据外-融合。

给定基数据(x)，$\exists x_j\overline{\in}(x)$，$f\in F$把$x_j$生成$f(x_j)=x_j'\in(x)$，$x_j'$构成$\Delta(x)_k$。若$(x)$，$\Delta(x)_k$与$F$-数据分解$(x)_k^F$满足：

$$(x)_k^F=(x)\cup\Delta(x)_k \tag{34}$$

则称$(x)_k^F$是(x)生成的数据内-融合。

由$(x)_k^{\bar F}$与$(x)_k^F$构成的数据融合对

$$((x)_k^{\bar F},(x)_k^F) \tag{35}$$

称为由(x)生成的数据外-内融合。

式(33)、式(34)中，$x_i\in(x)$是(x)内的数据元，$x_j\overline{\in}(x)$是(x)外的数据元。

给定A，$A_k^{\bar F}$与(x)，$(x)_k^{\bar F}$满足：

$$\text{if } A_k^{\bar F}\Rightarrow A,\text{then } (x)_k^{\bar F}\Rightarrow(x) \tag{36}$$

式(36)称作数据外-融合$(x)_k^{\bar F}$被(x)的智能生成，如果$(x)_k^{\bar F}$与(x)满足推理条件$A_k^{\bar F}\Rightarrow A$。

给定A，A_k^F与(x)，$(x)_k^F$满足：

$$\text{if } A\Rightarrow A_k^F,\text{then } (x)\Rightarrow(x)_k^F \tag{37}$$

式(37)称作数据内-融合$(x)_k^F$被(x)的智能生成，如果$(x)_k^F$与(x)满足推理条件$A\Rightarrow A_k^F$。

式(36)、式(37)中，$A_k^{\bar F}\Rightarrow A$与$A_k^{\bar F}\subseteq A$等价，$A\Rightarrow A_k^F$与$A\subseteq A_k^F$等价。

给定$(A_k^{\bar F},A)$，(A,A_k^F)与$((x)_k^{\bar F},(x))$，$((x),(x)_k^F)$满足：

$$\text{if } (A_k^{\bar F},A)\Rightarrow(A,A_k^F),\text{then } ((x)_k^{\bar F},(x))\Rightarrow((x),(x)_k^F) \tag{38}$$

式(38)称作数据外-内融合被(x)的智能生成，如果$((x)_k^{\bar F},(x))$与$((x),(x)_k^F)$满足推理条件$(A_k^{\bar F},A)\Rightarrow(A,A_k^F)$。

式(38)中，$(A_k^{\bar F},A)\Rightarrow(A,A_k^F)$表示$A_k^{\bar F}\Rightarrow A$，$A\Rightarrow A_k^F$。

给定数据外-融合$(x)_k^{\bar F}$，y_i是数据元$x_i\in(x)_k^{\bar F}$的值，给定基数据(x)，y_j是数据元$x_j\in(x)$的值；称$\sigma_k^{\bar F}$是$(x)_k^{\bar F}$关于(x)的距离。

$$\sigma_k^{\bar F}=\frac{(\sum_{i=1}^{p}y_i^2)^{\frac{1}{2}}}{(\sum_{j=1}^{q}y_j^2)^{\frac{1}{2}}} \tag{39}$$

给定数据内-融合$(x)_k^F$、基数据(x)，称σ_k^F是$(x)_k^F$关于(x)的距离。

$$\sigma_k^{F}=\frac{(\sum_{i=1}^{r}y_i^2)^{\frac{1}{2}}}{(\sum_{j=1}^{q}y_j^2)^{\frac{1}{2}}} \tag{40}$$

式(39)、式(40)中，$(x)=\{x_1,x_2,\cdots,x_q\}$，$(x)_k^{\bar F}=\{x_1,x_2,\cdots,x_p\}$，$(x)_k^F=\{x_1,x_2,\cdots,x_r\}$，$p<q<r$，$p,q,r\in\mathrm{N}^+$。

由$\sigma_k^{\bar F}$与σ_k^F构成的距离对，称作数据外-内融合$((x)_k^{\bar F},(x)_k^F)$关于(x)的距离。

$$(\sigma_k^{\bar F},\sigma_k^F) \tag{41}$$

由式(33)－式(41)得到以下定理。

定理4 数据外-融合$(x)_k^{\bar F}$被(x)智能生成的充分必要条件是存在$\nabla(x)_k\neq\varnothing$，$((x)-\nabla(x)_k)$，$(x)$与$A_k^{\bar F}$，$A$满足：

$$\text{if } A_k^{\bar F}\Rightarrow A,\text{then } ((x)-\nabla(x)_k)\Rightarrow(x) \tag{42}$$

存在$\Delta\alpha_k\neq\varnothing$，$(x)_k^{\bar F}$的属性集合$\alpha_k^F$，$(x)$的属性集合$\alpha$与$\Delta\alpha_k$满足：

$$\text{if } A_k^{\bar F}\Rightarrow A,\text{then } \alpha\Rightarrow(\alpha\cup\Delta\alpha_k) \tag{43}$$

证明：(1) 1)由式(1)、式(2)与式(15)、式(16)得到：若$A_k^{\bar F}$，A分别是$(x)_k^{\bar F}=(x)-\nabla(x)_k$，$(x)$的数值生成，$A_k^{\bar F}$与$A$满足$A_k^{\bar F}\subseteq A$，或者$A_k^{\bar F}\Rightarrow A$；$(x)_k^{\bar F}=((x)-\nabla(x)_k)\subseteq(x)$，或者$(x)_k^{\bar F}\Rightarrow(x)$，则在$A_k^{\bar F}\Rightarrow A$的条件下，得到$(x)_k^{\bar F}=((x)-\nabla(x)_k)\subseteq(x)$；或者$((x)-\nabla(x)_k)\Rightarrow(x)$。2)若$(x)_k^{\bar F}=((x)-\nabla(x)_k)\subseteq(x)$或者$(x)_k^{\bar F}=((x)-\nabla(x)_k)\Rightarrow(x)$，$A_k^{\bar F}$，$A$分别是$(x)_k^{\bar F}=(x)-\nabla(x)_k$，$(x)$的数值生成，则在$(x)_k^{\bar F}=((x)-\nabla(x)_k)\subseteq(x)$，或者$(x)_k^{\bar F}=((x)-\nabla(x)_k)\Rightarrow(x)$条件下得到$A_k^{\bar F}$与$A$满足$A_k^{\bar F}\Rightarrow A$。由1)与2)得到式(42)。

(2) 因为$A_k^{\bar F}$，A分别是$(x)_k^{\bar F}=((x)-\nabla(x)_k)$，$(x)$的数值生成，$A_k^{\bar F}\subseteq A$；$\alpha_k^F$，$\alpha$分别是$(x)_k^{\bar F}$，$(x)$的属性集合，$\alpha\subseteq\alpha_k^F$；$A_k^{\bar F}$，$A$分别具有与$(x)_k^{\bar F}$，$(x)$相同的属性集合$\alpha_k^F$，$\alpha$；又因为$\alpha_k^F$与$\alpha$满足$\alpha\subseteq(\alpha\cup\Delta\alpha)=\alpha_k^F$，$\Delta\alpha\neq\varnothing$；显然，若$A_k^{\bar F}\Rightarrow A$，则有$\alpha\Rightarrow(\alpha\cup\Delta\alpha)=\alpha_k^F$，得到式(43)。

推论7 数据外-融合$(x)_k^{\bar F}$关于(x)的距离$\sigma_k^{\bar F}$是单位离散区间(0,1]的内点，或者

$$\sigma_k^{\bar F}\in(0,1] \tag{44}$$

其中，(0,1]是数值0与1生成的单位离散区间，$1=\frac{(\sum_{i=1}^{q}y_i^2)^{\frac{1}{2}}}{(\sum_{j=1}^{q}y_j^2)^{\frac{1}{2}}}$是$(x)$的自身距离。

证明：由式(39)与(x)的自身距离直接得到式(44)。

定理 5 数据内-融合$(x)_k^F$被(x)智能生成的充分必要条件是存在$\Delta(x)_k\neq\varnothing$，$(x)\cup\Delta(x)_k$，$(x)$与$A_k^F$，$A$满足：

$$\text{if } A\Rightarrow A_k^F,\text{then } (x)\Rightarrow((x)\cup\Delta(x)_k) \tag{45}$$

存在$\nabla\alpha_k\neq\varnothing$，$(x)_k^F$的属性集合$\alpha_k^{\bar{F}}$，$(x)$的属性集合$\alpha$与$\nabla\alpha_k$满足：

$$\text{if } A\Rightarrow A_k^F,\text{then } (\alpha-\nabla\alpha_k)\Rightarrow\alpha \tag{46}$$

定理 5 的证明与定理 4 类似，证明略。

推论 8 数据内-融合$(x)_k^F$关于(x)的距离σ_k^F是单位离散区间(0,1]的外点，或者

$$\sigma_k^F\overline{\in}(0,1] \tag{47}$$

证明：由式(40)与(x)的自身距离直接得到式(47)。

由定理 4、定理 5、推论 7、推论 8 直接得到定理 6、推论 9。

定理 6 数据外-内融合$((x)_k^{\bar{F}},(x)_k^F)$被$(x)$智能生成的充分必要条件是存在$\nabla(x)_k\neq\varnothing$，$\Delta(x)_k\neq\varnothing$；$(((x)-\nabla(x)_k),(x))$，$((x),(x)\cup\Delta(x)_k)$与$(A_k^{\bar{F}},A)$，$(A,A_k^F)$满足：

$$\text{if } (A_k^{\bar{F}},A)\Rightarrow(A,A_k^F),\text{then } (((x)-\nabla(x)_k),(x))\Rightarrow((x),(x)\cup\Delta(x)_k) \tag{48}$$

存在$\Delta\alpha_k\neq\varnothing$，$\nabla\alpha_k\neq\varnothing$；$(A_k^{\bar{F}},A)$，$(A,A_k^F)$与$(\alpha,(\alpha-\nabla\alpha_k))$，$((\alpha\cup\Delta\alpha_k),\alpha)$满足：

$$\text{if } (A_k^{\bar{F}},A)\Rightarrow(A,A_k^F),\text{then } (\alpha,(\alpha-\nabla\alpha_k))\Rightarrow((\alpha\cup\Delta\alpha_k),\alpha) \tag{49}$$

推论 9 数据外-内融合$((x)_k^{\bar{F}},(x)_k^F)$关于$(x)$的距离$(\sigma_k^{\bar{F}},\sigma_k^F)$生成的离散区间$[\sigma_k^{\bar{F}},\sigma_k^F]$与单位离散区间(0,1]满足：

$$[\sigma_k^{\bar{F}},\sigma_k^F]\cap(0,1]\neq\varnothing \tag{50}$$

利用式(33)－式(50)得到以下命题。

命题 3 数据外-融合$(x)_{k+1}^{\bar{F}}$与$(x)_k^{\bar{F}}$满足$(x)_{k+1}^{\bar{F}}\cap(x)_k^{\bar{F}}\neq\varnothing$。

命题 4 数据内-融合$(x)_{k+1}^F$与$(x)_k^F$满足$(x)_{k+1}^F\cap(x)_k^F\neq\varnothing$。

命题 5 数据外-内融合$((x)_{k+1}^{\bar{F}},(x)_{k+1}^F)$与$((x)_k^{\bar{F}},(x)_k^F)$满足$((x)_{k+1}^{\bar{F}},(x)_{k+1}^F)\cap((x)_k^{\bar{F}},(x)_k^F)\neq\varnothing$。

其中，$((x)_{k+1}^{\bar{F}},(x)_{k+1}^F)\cap((x)_k^{\bar{F}},(x)_k^F)\neq\varnothing$表示$(x)_{k+1}^{\bar{F}}\cap(x)_k^{\bar{F}}\neq\varnothing$，$(x)_{k+1}^F\cap(x)_k^F\neq\varnothing$。

将第 3 节与第 4 节给出的讨论与结果交叉-整合，得到大数据分解-融合识别准则与大数据分解-融合获取智能算法。

大数据分解-融合识别准则如下：

数据$\bar{F}$-分解与外-融合$(x)_k^{\bar{F}}$与基数据(x)辨识$IDE((x)_k^{\bar{F}},(x))$，满足：

$$\sigma_k^{\bar{F}}-1<0 \tag{51}$$

数据F-分解与内-融合$(x)_k^F$与基数据(x)辨识$IDE((x)_k^F,(x))$，满足：

$$\sigma_k^F-1>0 \tag{52}$$

式(51)－式(52)中，IDE＝identification。

大数据分解-融合获取智能算法如图 1 所示，其过程如下。

图 1 中，模块 I－I* 给定基数据(x)与它的属性集合α，该步骤是算法的准备。模块 II－II* 在时间$k\in T$，α内补充属性，α生成α_k^F；(x)生成$\bar{F}$-分解$(x)_k^{\bar{F}}$。模块 III(x)生成复合概念：$\bar{F}$-分解与外-融合$(x)_k^{\bar{F}}$，$(x)_k^{\bar{F}}$是分解-融合交叉生成。模块 III 与模块 IV 合作，生成推理模块 V：if $A_k^{\bar{F}}\Rightarrow A$，then $(x)_k^{\bar{F}}\Rightarrow(x)$，$k=1,2,\cdots,n$。模块 VI 给定标准$(x)_k^{\bar{F},*}$，$\sigma_k^{\bar{F},*}$；$(x)_k^{\bar{F},*}$与$\sigma_k^{\bar{F},*}$构成复合比较-识别模块。模块 VII 生成双标准$\langle(x)_k^{\bar{F},*},\sigma_k^{\bar{F},*}\rangle$比较模块结构；实现：1)若$(x)_k^{\bar{F}}\neq(x)_k^{\bar{F},*}$，$\sigma_k^{\bar{F}}\neq\sigma_k^{\bar{F},*}$，则返回到 IV 内，获取新的$(x)_k^{\bar{F}}$，$\sigma_k^{\bar{F}}$，…；2)若$(x)_k^{\bar{F}}=(x)_k^{\bar{F},*}$，$\sigma_k^{\bar{F}}=\sigma_k^{\bar{F},*}$，则返回到 V，VI 内，生成$(x)_{k+1}^{\bar{F}}$，$(x)_{k+2}^{\bar{F}}$，…，$(x)_n^{\bar{F}}$。模块 VIII 生成满足标准的$(x)_1^{\bar{F}}$，$(x)_2^{\bar{F}}$，…，$(x)_n^{\bar{F}}$。

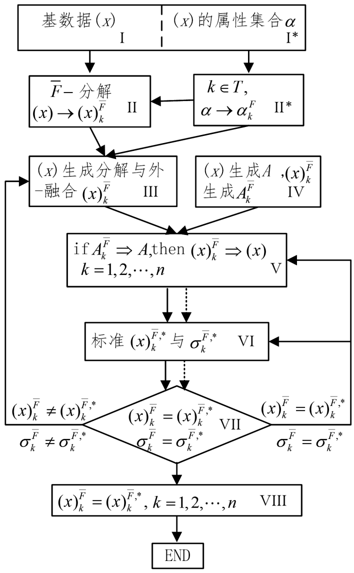

图 1 大数据分解-融合获取智能算法框图

Fig. 1 Block diagram of intelligent algorithm for big data decomposition-fusion acquisition

特别说明：图 1 是$\bar{F}$-数据分解-融合获取智能算法框图与算法的过程；F-数据分解-融合获取智能算法框图与算法的过程与图 1 类似，略。

5 大数据分解-融合的智能获取的应用

例子取自金融风险大数据-风险评估系统中的$\bar{F}$-数据在风险属性-入侵条件下的大数据状态智能发现。在风险属性入侵之前，基数据(x)生成$\bar{F}$-数据的动荡状态不被人们事先知道。

给定基数据(x)与(x)的属性集合α。

$$(x)=\{x_1,x_2,x_3,x_4,x_5,x_6\} \tag{53}$$

$$\alpha=\{\alpha_1,\alpha_2,\alpha_3\} \tag{54}$$

式(53)中，$\forall x_j\in(x)$是(x)的数据元，x_j的属性$\alpha_j=\alpha_1\wedge\alpha_2\wedge\alpha_3=\bigwedge_{i=1}^{3}\alpha_i$，$j=1,2,\cdots,6$，“$\wedge$”是“合取算子”。

$\boldsymbol{A}$是(x)生成的矩阵：

$$\boldsymbol{A}=\begin{pmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} & r_{16}\\ r_{21} & r_{22} & r_{23} & r_{24} & r_{25} & r_{26}\\ r_{31} & r_{32} & r_{33} & r_{34} & r_{35} & r_{36}\\ r_{41} & r_{42} & r_{43} & r_{44} & r_{45} & r_{46}\end{pmatrix}$$

$$=\begin{pmatrix} 2.80 & 1.04 & 2.71 & 1.21 & 2.33 & 1.77\\ 1.98 & 2.78 & 0.98 & 2.36 & 3.10 & 2.06\\ 1.63 & 1.66 & 1.74 & 1.40 & 2.17 & 1.15\\ 2.36 & 1.38 & 1.96 & 1.62 & 1.06 & 2.70\end{pmatrix} \tag{55}$$

(1) $\forall k\in T$,风险属性 α_4 入侵到 α 内,α 生成 α_k^F;由式(1)—式(3),在 α_4 入侵的条件下,(x)生成$\overline{F}$-数据$(x)_k^{\overline{F}}$,$(x)_k^{\overline{F}}$ 具有属性集合 α_k^F。

$$(x)_k^{\overline{F}}=\{x_1,x_2,x_4,x_5,x_6\} \tag{56}$$

$$\alpha_k^F=\alpha\cup\{\alpha_4\}=\{\alpha_1,\alpha_2,\alpha_3,\alpha_4\} \tag{57}$$

由式(53)—式(57)得到:

$$\text{if } \alpha\Rightarrow\alpha_k^F,\text{then } (x)_k^{\overline{F}}\Rightarrow(x) \tag{58}$$

由式(58)得到:$(x)_k^{\overline{F}}$ 满足推理条件 $\alpha\Rightarrow\alpha_k^F$,$(x)$被智能分解成$\overline{F}$-数据$(x)_k^{\overline{F}}\subseteq(x)$,$(x)_k^{\overline{F}}$ 在(x)内被智能获取;$x_3\in(x)$从(x)内被融合到(x)外,(x)生成$(x)_k^{\overline{F}}=\{x_1,x_2,x_4,x_5,x_6\}$。

显然,$(x)_k^{\overline{F}}$ 生成 $\boldsymbol{A}$ 的内 P-矩阵 $\boldsymbol{A}_k^{\overline{F}}$。

$$\boldsymbol{A}_k^{\overline{F}}=\begin{pmatrix} r_{11} & r_{12} & r_{14} & r_{15} & r_{16}\\ r_{21} & r_{22} & r_{24} & r_{25} & r_{26}\\ r_{31} & r_{32} & r_{34} & r_{35} & r_{36}\\ r_{41} & r_{42} & r_{44} & r_{45} & r_{46}\end{pmatrix} =\begin{pmatrix} 2.80 & 1.04 & 1.21 & 2.33 & 1.77\\ 1.98 & 2.78 & 2.36 & 3.10 & 2.06\\ 1.63 & 1.66 & 1.40 & 2.17 & 1.15\\ 2.36 & 1.38 & 1.62 & 1.06 & 2.70\end{pmatrix} \tag{59}$$

式(59)是式(55)的子矩阵,$\boldsymbol{A}_k^{\overline{F}}\subseteq\boldsymbol{A}$;$\boldsymbol{A}_k^{\overline{F}}$ 与 $\boldsymbol{A}$ 满足 $\boldsymbol{A}_k^{\overline{F}}\Rightarrow\boldsymbol{A}$;$(x)_k^{\overline{F}}$ 与(x)满足识别准则:$\sigma_k^{\overline{F}}<1$。

(2) $\forall k+\lambda\in T$,风险属性 $\alpha_5,\alpha_6,\alpha_7$ 同时入侵到 α_k^F 内,α_k^F 生成 $\alpha_{k+\lambda}^F$;$\overline{F}$-数据$(x)_k^{\overline{F}}$ 生成$(x)_{k+\lambda}^{\overline{F}}$,$(x)_{k+\lambda}^{\overline{F}}$ 具有属性集合 $\alpha_{k+\lambda}^F$。

$$(x)_{k+\lambda}^{\overline{F}}=\{x_1,x_4,x_5\} \tag{60}$$

$$\alpha_{k+\lambda}^F=\alpha_k^F\cup\{\alpha_5,\alpha_6,\alpha_7\}=\{\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5,\alpha_6,\alpha_7\} \tag{61}$$

由式(56)、式(57)、式(60)、式(61)得到:

$$\text{if } \alpha_k^F\Rightarrow\alpha_{k+\lambda}^F,\text{then } (x)_{k+\lambda}^{\overline{F}}\Rightarrow(x)_k^{\overline{F}} \tag{62}$$

由式(62)得到:$(x)_{k+\lambda}^{\overline{F}}$满足推理条件 $\alpha_k^F\Rightarrow\alpha_{k+\lambda}^F$,$(x)_k^{\overline{F}}$ 被智能分解成$\overline{F}$-数据$(x)_{k+\lambda}^{\overline{F}}$,$(x)_{k+\lambda}^{\overline{F}}\subseteq(x)_k^{\overline{F}}$,$(x)_{k+\lambda}^{\overline{F}}$在$(x)_k^{\overline{F}}$ 内被智能获取。$x_2,x_6\in(x)_k^{\overline{F}}$ 从$(x)_k^{\overline{F}}$ 内被融合到$(x)_k^{\overline{F}}$ 外,$(x)_k^{\overline{F}}$ 生成$(x)_{k+\lambda}^{\overline{F}}=\{x_1,x_4,x_5\}$。

$(x)_{k+\lambda}^{\overline{F}}$生成 $\boldsymbol{A}$ 的内 P-矩阵 $\boldsymbol{A}_{k+\lambda}^{\overline{F}}$。

$$\boldsymbol{A}_{k+\lambda}^{\overline{F}}=\begin{pmatrix} r_{11} & r_{14} & r_{15}\\ r_{21} & r_{24} & r_{25}\\ r_{31} & r_{34} & r_{35}\\ r_{41} & r_{44} & r_{45}\end{pmatrix}=\begin{pmatrix} 2.80 & 1.21 & 2.33\\ 1.98 & 2.36 & 3.10\\ 1.63 & 1.40 & 2.17\\ 2.36 & 1.62 & 1.06\end{pmatrix} \tag{63}$$

$\boldsymbol{A}_{k+\lambda}^{\overline{F}}$,$\boldsymbol{A}_k^{\overline{F}}$ 与 $\boldsymbol{A}$ 满足 $\boldsymbol{A}_{k+\lambda}^{\overline{F}}\Rightarrow\boldsymbol{A}_k^{\overline{F}}\Rightarrow\boldsymbol{A}$;或者 $\boldsymbol{A}_{k+\lambda}^{\overline{F}}\subseteq\boldsymbol{A}_k^{\overline{F}}\subseteq\boldsymbol{A}$;$(x)_{k+\lambda}^{\overline{F}}$与$(x)$满足识别准则:$\sigma_{k+\lambda}^{\overline{F}}<1$。

(3)例子的结果式(56)、式(60)的认证:式(56)、式(60)的结果与某金融机构在 2017 年 1—4 季度实际利润年报相符,平均误差在 1.037%,式(56)、式(60)的结果被认证。

(4)$\overline{F}$-数据融合$(x)^{\overline{F}}$状态存在的估计与$\overline{F}$-数据融合的数学特征。

1)从式(1)—式(3)看到:$\overline{F}$-数据融合$(x)^{\overline{F}}$(内 P-集合 $X^{\overline{F}}$)是风险属性 α_i^* 从基数据(x)的属性集合 α 外被融合到 α 内,(x)生成$(x)^{\overline{F}}$。换一个说法,α_i^* 被融合到 α 的可能性 $\rho^F=1$(α_i^* 融合到 α 内的概率值 $\rho^F=1$),这是$(x)^{\overline{F}}$存在的极限情形。在实际中,金融机构(股份制银行集团)投资企业要对风险属性 α_i^* 的存在与 α_i^* 入侵到 α 内的风险进行评估;或者金融机构利用集值统计迭代的方法对 ρ^F 给出估计。在一般情况下,得到的结果 $\rho^F\approx1$ 或者“风险可控”。如果 $\rho^F<0.5$,金融机构对投资风险保持警觉,这个事实是金融系统中的一个常识。式(20)给出了$(x)^{\overline{F}}$存在的界定 $0<\rho^F\leqslant1$。

2)因为受到一些原因的限制,例子中的属性 α_i 的名称被忽略。矩阵 $\boldsymbol{A}_k^{\overline{F}}$,$\boldsymbol{A}_{k+\lambda}^{\overline{F}}$中的数值是真实数值经过技术方法后得到的,利用技术方法后的数值不影响例子的分析。$A_k^{\overline{F}}$,$A_{k+\lambda}^{\overline{F}}$中的每一列是 x_i 在一年中 4 个季度的盈利值,例子中的 x_1~x_6 是金融机构的 6 个分支机构(股份制银行集团的 6 个分支)。例子中,k,$k+\lambda\in T$ 是风险属性入侵的数据参数。

3)因为风险属性的存在以及入侵到 α 内,(x)的 x_i 被融合到(x)外,(x)生成$(x)_k^{\overline{F}}$,$(x)_{k+\lambda}^{\overline{F}}$;或者$(x)$被分解成$(x)_k^{\overline{F}}$,$(x)_{k+\lambda}^{\overline{F}}$是一个普通常识。用数学的概念认识分解-融合$(x)_k^{\overline{F}}$,$(x)_{k+\lambda}^{\overline{F}}$得到:$(x)_k^{\overline{F}}$ 是关于属性集合 α_k^F 的等价类,$(x)_k^{\overline{F}}=[(x)_k^{\overline{F}}]_{\alpha_k^F}$;$(x)_{k+\lambda}^{\overline{F}}$是关于属性集合 $\alpha_{k+\lambda}^F$ 的等价类,$(x)_{k+\lambda}^{\overline{F}}=[(x)_{k+\lambda}^{\overline{F}}]_{\alpha_{k+\lambda}^F}$。因为 $\forall x_i\in(x)_k^{\overline{F}}$ 关于 α_k^F 满足自反性、对称性与传递性。显然,数学概念对大数据概念给出理论提升。

结束语 本文是文献[1]对大数据研究的继续,给出了 ∧ 型大数据的一些新的结果。∧ 型大数据是从具有动态特征的 P-集合中得到的,利用 P-集合的结构分析了 ∧ 型大数据的动态特征。应该说,在应用中遇到的大数据与它的状态不是一成不变的。在 P-集合中,内 P-集合的生成是由有限普通元素集合 X 的属性集合 α 内被补充属性,(x)生成$(x)^{\overline{F}}$;外 P-集合的生成是有限普通元素集合 X 的属性集合 α 内删除属性,(x)生成$(x)^F$。这两个简单特征用数据融合概念(信息[68-71]融合概念)认识得到:内 P-集合 $X^{\overline{F}}$是$\overline{F}$-数据融合$(x)^{\overline{F}}$,外 P-集合 X^F 是 F-数据融合$(x)^F$。显然,把数据融合概念引入到大数据基础理论研究中,理由充分,方法可行。(x)生成$(x)^{\overline{F}}$用数据分解(信息分解)概念认识得到:$(x)^{\overline{F}}$是(x)的$\overline{F}$-分解,$(x)^F$ 是(x)的 F-分解。

$\overline{F}$-数据融合$(x)^{\overline{F}}$是(x)内的一些数据元 x_i 从(x)内被融合到(x)外,F-数据融合$(x)^F$ 是(x)外的一些数据元 x_j 从(x)外被融合到(x)内。$\overline{F}$-分解$(x)^{\overline{F}}$是(x)的属性集合 α 内补充属性,(x)被分解成$(x)^{\overline{F}}$,$(x)^{\overline{F}}\subset(x)$;$F$-分解$(x)^F$ 是(x)的属性集合 α 内删除属性,(x)被分解成$(x)^F$,$(x)\subset(x)^F$。本文把分解、融合两个概念交叉、组合得到一个新的合成概念:分解-融合。把分解-融合概念应用到大数据的基本理论研究中,以分解-融合为出发点,获取大数据的一些新特征。

在大数据基础理论与应用研究中,人们已越来越多地认识到多个数学概念、方法的融合与人工智能方法、模型的交叉,逐步认识大数据概念的内涵,获取大数据新的应用。事实上,大数据是应用数学、智能系统、认知科学相结合生成的一个崭新的应用领域,这个认识是文献[1]与本文的主题。

参 考 文 献

[1] SHI K Q. Big data structure-logic characteristics and big data

law [J]. Journal of Shandong University (Natural Science), 2019, 54(2): 1-29.

[2] WANG F Y, CARLEY K M, ZENG D, et al. Social computing: from social informatics to social intelligence [J]. IEEE Intelligent systems, 2007, 22(2): 79-83.

[3] ZHANG N, WANG F Y, ZHU F, et al. DynaCAS: Computational experiments and decision support for ITS [J]. IEEE Intelligent Systems, 2008, 23(6): 19-23.

[4] WANG F Y. Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications [J]. IEEE Transactions on Intelligent Transportation Systems, 2010, 11(3): 630-638.

[5] WANG F Y. Scanning the issue and beyond : Parallel driving with software vehicular robots for safety and smartness [J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(4): 1381-1387.

[6] LV Y, DUAN Y, KANG W, et al. Traffic flow prediction with big data: A deep learning approach [J]. IEEE Trans. Intelligent Transportation Systems, 2015, 16(2): 865-873.

[7] LV Y, ZHANG X, KANG W, et al. Managing emergency traffic evacuation with a partially random destination allocation strategy: a computational-experiment-based optimization approach[J]. IEEE Trans. Intelligent Transportation Systems, 2015, 16(4): 2182-2191.

[8] YUAN Y, WANG F Y, ZENG D. Developing a cooperative bidding framework for sponsored search markets-An evolutionary perspective [J]. Information Sciences, 2016, 369: 674-689.

[9] WANG F Y. The emergence of intelligent enterprises: From CPS to CPSS [J]. IEEE Intelligent Systems, 2010, 25(4): 85-88.

[10] GOODHOPE K, KOSHY J, KREPS J, et al. Building LinkedIn's real-time activity data pipeline [J]. IEEE Data Eng. Bull., 2012, 35(2): 33-45.

[11] MCKUSICK K, QUINLAN S. GFS: evolution on fast- forward [J]. Communications of the ACM, 2010, 53(3): 42-49.

[12] COOPER B F, RAMAKRISHNAN R, SRIVASTAVA U, et al. PNUTS: Yahoo!'s hosted data serving platform [J]. Proceedings of the VLDB Endowment, 2008, 1(2): 1277-1288.

[13] MELNIK S, GUBAREV A, LONG J J, et al. Dremel: interactive analysis of web-scale datasets [J]. Proceedings of the VLDB Endowment, 2010, 3(1/2): 330-339.

[14] BU Y, HOWE B, BALAZINSKA M, et al. HaLoop: efficient iterative data processing on large clusters [J]. Proceedings of the VLDB Endowment, 2010, 3(1/2): 285-296.

[15] LAM W, LIU L, PRASAD S T S, et al. Muppet: MapReduce-style processing of fast data [J]. Proceedings of the VLDB Endowment, 2012, 5(12): 1814-1825.

[16] CHEN S. CHEETAH: a high performance, custom data warehouse on top of MapReduce [J]. Proceedings of the VLDB Endowment, 2010, 3(1/2): 1459-1468.

[17] CANIM M, MIHAILA G A, BHATTACHARJEE B, et al. SSD bufferpool extensions for database systems [J]. Proceedings of the VLDB Endowment, 2010, 3(1/2): 1435-1446.

[18] LUO T, LEE R, MESNIER M, et al. Storage-DB: heterogeneity-aware data management to exploit the full capability of hybrid storage systems [J]. Proceedings of the VLDB Endowment, 2012, 5(10): 1076-1087.

[19] WONG P C, SHEN H W, JOHNSON C R, et al. The top 10 challenges in extreme-scale visual analytics[J]. IEEE computer graphics and applications, 2012, 32(4): 63-67.

[20] PIKE R, DORWARD S, GRIESEMER R, et al. Interpreting the data: parallel analysis with Sawzall [J]. Scientific Programming, 2005, 13(4): 277-298.

[21] SHI K Q. P-sets [J]. Journal of Shandong University (Natural Science), 2008, 43(11): 77-84.

[22] SHI K Q. P-sets and its applications [J]. Advances in Systems Science and Applications, 2009, 9(2): 209-219.

[23] SHI K Q. P-sets and its applied characteristics [J]. Computer Science, 2010, 37(8): 1-8.

[24] SHI K Q. P-reasoning and P-reasoning discovery- identification of information [J]. Computer Science, 2011, 38(7): 1-9.

[25] SHI K Q. P-sets, inverse P-sets and the intelligent fusion- filter identification of information [J]. Computer Science, 2012, 39(4): 1-13.

[26] FAN C X, LIN H K. P-sets and the reasoning- identification of disaster information [J]. International Journal of Convergence Information Technology, 2012, 7(1): 337-345.

[27] LIN H K, FAN C X. The dual form of P-reasoning and identification of unknown attribute [J]. International Journal of Digital Content Technology and its Applications, 2012, 6(1): 121-131.

[28] ZHANG L, CUI Y Q, SHI K Q. Outer P-sets and data internal recovery [J]. Systems Engineering and Electronics, 2010, 32(6): 1233-1238.

[29] WANG Y, GENG H Q, SHI K Q. P-sets and dependence- discovery of dynamic information [J]. Systems Engineering and Electronics, 2011, 33(9): 2035-2038.

[30] LI Y Y, LIN H K, SHI K Q. Characteristics of data discrete interval and data discovery-application [J]. Systems Engineering and Electronics, 2011, 33(10): 2258-2262.

[31] ZHANG L, TANG J H, SHI K Q. The fusion of internal packet information and its feature of attribute conjunction [J]. Journal of Shandong University (Natural Science), 2014, 49(2): 93-97.

[32] SHI K Q, ZHANG L. Internal P-set and data outer- recovery [J]. Journal of Shandong University (Natural Science), 2009, 44(4): 8-14.

[33] ZHANG G Y, ZHOU H Y, SHI K Q. P-sets and the recovery-identification double [J]. Systems Engineering and Electronics, 2010, 32(9): 1919-1924.

[34] ZHANG L, REN X F. P-sets and $(f, \overline{f})$-heredity [J]. Quantitative Logic and Soft Computing, 2010, 2(1): 735-743.

[35] ZHANG L, XIU M, SHI K Q. P-sets and application of internal-outer data circle [J]. Quantitative Logic and Soft Computing, 2010, 2(1): 581-591.

[36] QIU Y F, CHEN B H. f-model generated by P-sets [J]. Quantitative Logic and Soft Computing, 2010, 2(1): 613-620.

[37] LI Y Y, ZHANG L, SHI K Q. Generation and recovery of compressed data and redundant data[J]. Quantitative Logic and Soft Computing, 2010, 2(1): 661-671.

[38] XIU M, SHI K Q, ZHANG L. P-sets and $\overline{F}$-data selection-discovery [J]. Quantitative Logic and Soft Computing, 2010, 2(1): 791-799.

[39] ZHAO S L, FAN C X, SHI K Q. Outer P-information generation and its reasoning-searching discovery[J]. Journal of Shandong University (Natural Science), 2012, 47(1): 99-104.

[40] ZHAO S L, WU S L, SHI K Q. Internal P- reasoning information recovery and attribute hiding searching discovery [J]. Computer Science, 2013, 40(4): 209-213.

[41] SHI K Q. Function P-sets[J]. Journal of Shandong University (Natural Science), 2011, 46(2): 62-69.
[42] SHI K Q. Function P-sets[J]. International Journal of Machine Learning and Cybernetics, 2011, 2(4): 281-288.
[43] SHI K Q. P-information law intelligent fusion and soft information image intelligent generation[J]. Journal of Shandong University(Natural Science), 2014, 49(4): 1-17.
[44] TANG J H, ZHANG L, SHI K Q. Intelligent fusion of information law and its inner separation[J]. Computer Science, 2015, 42(2): 204-209.
[45] LIN R, FAN C X. Function P-sets and dynamic characteristics of information law[J]. Journal of Shandong University (Natural Science), 2012, 47(1): 121-126.
[46] REN X F, ZHANG L, SHI K Q. Two types of dynamic information law models and their applications in information camouflage and risk identification[J]. Computer Science, 2018, 45(9): 230-236.
[47] CHEN B H, ZHANG L, SHI K Q. Intelligent dynamic fusion of packet information and the intelligent state recognition of information law[J]. Journal of Shandong University (Natural Science), 2018, 53(2): 83-87.
[48] TANG J H, ZHANG L, SHI K Q, et al. Outer P- information law reasoning andits application in intelligent fusion and separating of information law[J]. Microsystem Technologies, 2018, 24(10): 4389-4398.
[49] SHI K Q. P-augmented matrix and dynamic intelligent discovery-identification of information[J]. Journal of Shandong University (Natural Science), 2015, 50(10): 1-12.
[50] LIN H K, FAN C X. Embedding camouflage of inverse P-information and application[J]. International Journal of Convergence Information and Technology, 2012, 7(20): 471-480.
[51] FAN C X, HUANG S L. Inverse P-reasoning discovery identification of inverse P-information[J]. International Journal of Digital Content Technology and its Applications, 2012, 6(20): 735-744.
[52] SHI K Q. Function inverse P-sets and the hiding information generated by function inverse P-information law fusion[C]// Proceedings of the 13th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, Sanya, China, 2014: 224-237.
[53] GUO H L, REN X F, ZHANG L. Relationships between dynamic data mining and P-augmented matrix[J]. Journal of Shandong University (Natural Science), 2016, 51(8): 105-110.
[54] GUO H L, CHEN B H, TANG J H. Inverse P-sets and intelligent fusion mining-discovery of information[J]. Journal of Shandong University (Natural Science), 2013, 48(8): 97-103, 110.
[55] ZHANG L, REN X F, SHI K Q. The dynamic segmentation characteristics of P-augmented matrix and the dynamic intelligent acquisition of P-information[J]. International Journal of Applied Decision Sciences, 2016, 9(4): 413-425.
[56] ZHANG L, REN X F, SHI K Q, et al. Inverse packet matrix reasoning model-based the intelligent dynamic separation and acquisition of educational information[J]. Microsystem Technologies, 2018, 24(10): 4415-4421.
[57] REN X F, ZHANG L, SHI K Q, et al. Inverse P-augmented matrix method-based the dynamic findings of unknown information [J]. Microsystem Technologies, 2018, 24(10): 4187-4192.
[58] ZHANG L, REN X F, SHI K Q. Inverse P-information law models and the reality-camouflage intelligent transformations of information image[C]// Proceedings of the 2016 International Conference on Network and Information Systems for Computers, Washington: IEEE, 2016: 337-341.
[59] ZHANG L, REN X F, SHI K Q. Inverse P-data models and data intelligent separation[C]// Proceedings of the 2016 International Conference on Electronic Information Technology and Intellectualization, DOI: 10.12783/dtcse/iceiti2016/6107, 2016.
[60] REN X F, ZHANG L, SHI K Q. Surplus- deficiency of cardinal number and inverse P-augmented matrices[J]. Journal of Shandong University (Natural Science), 2015, 50(10): 13-18, 26.
[61] ZHANG L, REN X F. Surplus-deficient theorem of cardinal number and data internal-outer mining- separation [J]. Journal of Shandong University (Natural Science), 2015, 50(8): 90-94.
[62] GUO H L, ZHANG L. Data separation and its attribute state characteristics [J]. Journal of Shandong University (Natural Science), 2017, 52(12): 89-94.
[63] TANG J H, CHEN B H. Intelligent fusion of internal inverse packet information and expansion relationship of attribute disjunction[J]. Journal of Shandong University (Natural Science), 2014, 49(2): 89-92, 97.
[64] ZHANG L, REN X F, SHI K Q. Intelligent switch- camouflage of information laws and P-law augmented matrices [J]. Journal of Shandong University (Natural Science), 2016, 51(8): 90-97.
[65] REN X F, ZHANG L, SHI K Q. Boundary Characteristics of Inverse P-sets and System Condition Monitoring[J]. Computer Science, 2016, 43(10): 211-213, 255.
[66] REN X F, ZHANG L. Perturbation theorems of inverse P-sets and perturbation-based data mining[J]. Journal of Shandong University (Natural Science), 2016, 51(12): 54-60.
[67] LIU J Q, ZHANG H Y. Information P-dependence and P-dependence mining-sieving [J]. Computer Science, 2018, 45(7): 202-206.
[68] SHI K Q, YAO B X. Function S-rough sets and law identification[J]. Science in China E: Information Science, 2008, 38(4): 553-564.
[69] SHI K Q, ZHAO J L. Function S-rough sets and security-authentication of hiding law[J]. Science in China E: Information Science, 2008, 38(8): 1234-1243.
[70] SHI K Q, YAO B X. Function S-rough sets and law identification[J]. Science in China F: Information Science, 2008, 51(5): 499-510.
[71] SHI K Q, ZHAO J L. Function S-rough sets and security-authentication of hiding law[J]. Science in China F: information Sciences, 2008, 51(7): 924-935.

LIU Ji-qin, born in 1968, Ph.D, professor. Her main research interests include big data intelligent analysis and application and rough system theory and application.

SHI Kai-quan, born in 1945, professor, Ph.D supervisor. His main research interests include big data theory and application, information intelligent system. He proposed S-rough sets, function S-rough sets, P-sets, inverse P-sets, function P-sets and function inverse P-sets.