文章编号: 1001- 2486(2010) 04- 0145- 05

基于饱和信息维的信息资源聚焦的数学模型*

吴 强[1], 李建平[1], 黄宏斌[2], 胡小荣[1], 黄建华[1]
(1. 国防科技大学 理学院, 湖南 长沙 410073; 2. 国防科技大学 信息系统与管理学院, 湖南 长沙 410073)

摘 要: 运用饱和信息维的增减技术对信息获取子空间进行结构上的剖分, 将饱和信息维的"值"的获取分解为"白"与"灰"两部分, 用饱和信息维列作为因子列来构造灰色关联空间, 用关联序来描述信息之间的相关联程度, 得到信息资源聚焦的一种新的数学模型。

关键词: 饱和信息维; 灰关联空间; 关联度; 关联序; 信息聚焦
中图分类号: TP391 文献标识码: A

Information Resource Focusing Mathematical Model Besed on Saturation Information Dimension

WU Qiang[1], LI Jian-ping[1], HUANG Hong-bin[2], HU Xiao-rong[1], HUANG Jian-hua[1]
(1. College of Science, National Univ. of Defense Technology, Changsha 410073, China;
2. College of Information System and Management, National Univ. of Defense Technology, Changsha 410073, China)

Abstract: The increasing or decreasing technology of the saturated information dimension was applied to decompose and analyze the structure of information acquisition subspace one, and the acquisition of the value of saturated information dimension was divided into two parts—White and Grey. The saturation information dimension was used as the correlative order, applied to describe the grey correlative space, and the correlative oder was used to describe the correlative degree between information, thus obtaining a new mathematical model focusing on information resource.

Key words: saturation information dimension; grey correlation space; correlate degree; correlative order; information focusing

信息技术的高速发展, 已影响到科学技术的所有领域。信息及信息的获取已成为发展现代科学技术所应考虑的首要问题之一。信息获取科学与技术[3], 是信息化时代传统的传感技术和检测技术上的学科升华, 它在科学的深度和技术的广度上突破了原来的界限, 系统地解决信息获取所面临的科学和技术问题。信息的获取依赖于信息与信息系统的广泛应用, 依赖于对一切所需信息准确、适时、有效的掌握, 这需要精确化的信息服务支持。文献[1]提出了"信息资源聚焦服务"的概念, 对信息资源进行组织、索引, 形成以用户关注信息点为核心的满足用户需求的信息资源集合, 为用户提供有效的信息服务, 以解决"信息缺乏"和"信息泛滥"的矛盾。本文在信息获取科学的基础上, 运用饱和信息维的增减技术对信息获取子空间进行结构上的剖分, 将饱和信息维的"值"的获取分解为"白"与"灰"两部分, 用饱和信息维列作为因子列来构造灰色关联空间, 用关联序来描述信息之间的相关联程度, 得到信息资源聚焦的一种新的数学模型。它的特征是"序列"聚焦而不是"集合"聚焦, 从系统论的角度来看是对关联信息从无序到有序、从宏观到微观分析的一种突破, 它充分利用了信息结构各个层面在不同背景下的资源映射, 从形式上来看更为细化、可操作性更强。

1 信息、信息维及饱和信息维的结构描述

信息具有本体论和认识论两个层次, 从本体论到认识论, 信息的本质是变化的, 但变化并不等于信息, 只有客体的变化通过某种途径作用到主体, 才构成了信息。所以信息的本质定义应该是相对于主体

* 收稿日期: 2009- 11- 30
基金项目: 国家部委资助项目
作者简介: 吴强(1964—), 男, 副教授。

而存在的客体变化。鉴于此,在信息获取科学的框架内,依据这种“变化”来界定信息的描述方式,以信息差异性的一种作为信息的一维,应用灰色认知模型来刻画信息的结构性差异。

文献[1]指出,智能体为了获得改变世界无序化进程的能力,将外部信息转化为内部信息的过程,称为信息获取;人类在进行信息获取实践活动的过程中积累的经验、技巧和知识,称为信息获取技术;研究信息获取的基本原理和普遍规律的科学,称为信息获取科学。信息与物质的质量、能量一样,也是物质的存在方式,“差异”是信息,凡信息必有差异,没有物质就没有信息,信息的存在以物质的存在为前提。物质的差异性存在方式,称为信息,通常用 ξ, ξ_i 来表示。

灰色系统理论认为,信息是认知的根据,认知只能而且必须以信息作为根据,凡是作为认知根据的,均为信息。

定义1 设 w^0 为白化度,IFM 为一特定信息获取技术下的认知映射:

$$IFM: \xi \xrightarrow{} \Theta,\ \Theta \subseteq (a, 1],\ a \geqslant 0, \quad IFM: \xi \xrightarrow{} \Theta,\ \Theta \in \Theta,\ 0 \leqslant \Theta \leqslant 1,\ \Theta = w^0 \tag{1}$$

则称 ξ 为信息载体(信息元或信息),Θ 为认知程度信息,Θ 为认知程度集,若 $\Theta_j \subseteq \Theta$,则称 Θ_j 为信息载体 ξ 的 j 认知集;若 $\Theta_j \in \Theta$,则称 Θ_j 为 ξ 在 j 指标下的认知度。

定义2 若 θ 为信息 ξ 的内在信息(属性信息),ξ 为一特定信息获取技术下 ξ 的白化表现,则 ξ 在形式上可表述为

$$\xi \leftarrow (\xi \| \theta) \quad 或 \quad \xi \xrightarrow{} (\xi \| \theta) \tag{2}$$

此式也称为信息 ξ 在一特定信息获取技术下的信息表现模式。

命题1[7] 令 $\Theta_j \in \Theta_j \subseteq \Theta$, $\Theta = \{\Theta_j \mid j \in J\}$, $J = \{1, 2, \cdots, n;\ \text{or}\ 1, 2, \cdots, n, \cdots\}$

记 U_j 为 Θ_j 的邻域族:

$$U_j = \{\Theta_j \mid \Theta_j \in \Theta_j, \quad j \in J\}$$

且记 Θ_j 的拓扑为 Γ,若 U_j 满足下述条件(常称为邻域族条件):(1) $\forall \Theta_j \in U_j \xrightarrow{} \Theta_j \in \Theta_j$;(2) 若 $\Theta_j \supseteq \Theta_i$,则 $\Theta_j \in U_i$, $\forall \Theta_i$;(3) 若 Θ_i, $\Theta_j \in U_i$,则 $\Theta_j \cap \Theta_i \in U_i$;(4) 若 $\Theta_i \in U_i$,则有 $\Theta_j \in U_i$,进而,若 $\Theta_j \in \Theta_j$,则 $\Theta_i \in U_j$。则 (Θ, Γ) 称为 ξ 的认知程度拓扑空间,亦称为 ξ 的认知拓扑空间。

命题2 信息载体 ξ 的认知度 Θ 可以进行结构上(纵向或横向)的剖分。

证明 由定义1,2及命题1可直接得到。

信息载体 ξ 的这种拓扑性和相容性,为刻画信息的结构提供了背景和基础,在 $W(\Theta)$ 中可按定义1,2分解出子集和元素(属性),且认知模式的数字域可以扩大到负区,比如将虚的、假的认知程度作为负数,而且负值越大,虚假程度越高,为简化定义,以 $\Theta = 0$ 表示认知为假、为虚、为零,无认知与假认知未加区别。相对于 ξ,纵向的剖分得到信息的不同层次,横向的剖分得到不同的构成元素(属性),它们存在着差异性。

定义3 信息 ξ 的一种横向的分解得到不同的构成元素(属性)集合(也称为差异性集合)称为信息的维,简称维(Dimension)。维用符号 d_i 表示,信息 ξ_i 的第 i 维信息表示为 $d_i(\xi_i)$。信息 ξ 第 i 维 $d_i(\xi_i)$ 的属性差异状态描述,称为该信息中维的取值。

值得注意的是信息中维的取值可以是一个白化数(确定的实数值),也可以是具有某种差异性描述背景的灰数(只知大概范围,而不知其准确数值)。

在一般的信息获取系统中,为了更加有效地获取、处理和传输信息,可以将信息的维进行排序(可以为有限也可以为无限),用序号代替具体的差异性描述。

定义4[3] 用 Ω 表示全体信息组成的集合

$$\Omega = \{\xi \mid \xi \in S_{Inf}\} \tag{3}$$

若在 Ω 上赋予信息外延公理、空间外延公理和空间分离公理,则 Ω 称为信息空间。

信息空间 Ω 中某些维的取值被确定后,得到维数降低的信息空间 Ω_1,称为信息子空间,记为 $\Omega_1 \subset \Omega$。由具体的信息获取系统中有限维构成的信息空间,称为信息获取子空间,记为 Ω';信息获取子空间中的信息,称为具体信息。

在不同的信息获取子空间中, 具体的维及表现形式会有所不同, 这取决于信息系统的信息获取能力。但一般来说其维数总是有限的, 经过初步信息获取过程以后的信息为

$$\xi = f(d_1, d_2, \cdots, d_i)\ (\text{在一 } IFM \text{ 下}) \quad i \in I = (1, 2, \cdots, n) \tag{4}$$

在信息获取技术的形式化结构下, 信息获取子空间是有限维的笛卡儿坐标空间。文献[3]给出了信息获取过程中的数学描述通式为(无限到有限的过程):

$$\xi_1 \to \xi_2,\ \text{where} \begin{cases} \xi_1 = f_1(d_1, d_2, \cdots, d_\infty),\ \xi_2 = f_2(d_1, d_2, \cdots, d_n) \\ d_i(\xi_1) \subseteq d_i(\xi_2), \quad 1 \leqslant i \leqslant n \end{cases} \tag{5}$$

信息中维的"取值"并不完全是数学意义上的等于, 而是相当于集合意义上的属于。由于信息获取子空间中的具体信息"取值"是各个维上的区间(或一点), 因此, 信息空间中信息的实际存在状态总是这些取值区间中的一部分或者一点。

定义 5　设 Ω' 为信息获取子空间, $\xi_0 \in \Omega'$, 若 $D(\xi_0) = n$, 记

$$\xi_0 = f(d_1, d_2, \cdots, d_s, d_{s+1}, \cdots, d_n), \quad \xi' = f(d_1, d_2, \cdots, d_s)$$

$$\xi'' = f(d_1, d_2, \cdots, d_s, d_{s+1}, \cdots, d_n, d_{n+1}, \cdots, d_t) \quad (s \leqslant n \leqslant t)$$

则称 ξ_0 为饱和维信息, 称 ξ' 为关于 ξ_0 的不饱和维信息, 称 ξ'' 为关于 ξ_0 的超饱和维信息。

不饱和维信息可以通过"加维"的方式扩张为饱和维信息, 即在缺维的部分补充定义, 使其"取值"均为 $\overline{\cong}$ 即可, 理解为"取值覆盖了该维的全部"。

命题 3　ξ_0 饱和维的增减技术能够张成一个可以从 Ω' 中分离出的子空间, 它们具有与 ξ_0 有关的共同性质。

证明　由空间分离公理可直接得到。

定义 6　具有与 ξ_0 有关的共同性质张成的一个从 Ω' 中分离出的子空间, 称为 ξ_0 的具有某种性质的关联子空间, 记为 Ω'_{ξ_0}。

数值与信息的对立, 是信息获取科学中的难点之一。认知的根据是信息, 根据非唯一, 认知也非唯一, 但真认知是唯一的, 既然存在着唯一的真认知的信息表现模式, 那如何才能找到这个唯一的真认知呢? 灰色系统理论认为, 任何一种复杂现象, 理论上都可以通过量化的一组信息完全或不完全的数值(其中里面的任一数值实际上是影响其它各种内外因素耦合作用的结果在某一时空上的明白显示, 是一种态势, 我们称其为数值态)来直观描述。

2　基于灰色关联序的信息聚焦的数学模型

2.1　关联子空间数值态的形式化描述

设 Ω' 为包含智能体关注的信息点 ξ_0 的信息获取子空间, 根据 ξ_0 的整体信息特征, 经过信息获取技术的处理, 得到 ξ_0 为饱和维信息的表示模式:

$$\xi_0 = f(d_{0w}(1), d_{0w}(2), \cdots, d_{0w}(t), d_{0g}(1), d_{0g}(2), \cdots, d_{0g}(h)) \tag{6}$$

其中, $d_{0w}(1), d_{0w}(2), \cdots, d_{0w}(t)$ 表示 ξ_0 的 t 个白化(确定)的信息特征, 而后面的 $d_{0g}(1), d_{0g}(2), \cdots, d_{0g}(h)$ 表示 ξ_0 的 h 个灰色(不完全确定)的信息特征, 且有 $t + h = n$。

同时得到 Ω' 中与 ξ_0 相关联的其它信息(一般来说总是有限的, 不妨设为 m 个)的表示模式:

$$\xi_i = f(d_{iw}(1), d_{iw}(2), \cdots, d_{iw}(t), d_{ig}(1), d_{ig}(2), \cdots, d_{ig}(h)) \quad (i = 1, 2, \cdots, m) \tag{7}$$

信息获取的过程, 是对信息各维取值范围不断明确的过程, 此处有如下约定: (1) 不饱和维信息通过"加维"扩张为饱和维信息, $\overline{\cong}$ 的对应部分取值为零; (2) 超饱和维信息把多余维去掉变成饱和维信息(即"超"的部分不认知); (3) 相关信息各维的取值性态亦相对应。

综上即为 ξ_0 关联子空间的数值态形式化的具体描述。

2.2　基于关联序的信息资源聚焦的数学模型

信息资源聚焦[1], 是围绕智能体关注的信息点, 将相关关联的信息资源聚集起来形成一个以智能体

关注的信息点为中心的关联信息资源集合的过程,是将相关的信息资源汇聚而忽视无关信息资源的过程,其特征是"集合(邻域)聚焦"。我们运用饱和信息维的增减技术对信息获取子空间进行结构上的剖分,试图通过收敛的信息列,得到一种新的聚焦模式。

对式(6)、(7)的白化部分,令

$$\gamma(\xi_0(k),\xi(k))=\frac{\min\limits_i\min\limits_k|d_{0w}(k)-d_{iw}(k)|+\rho\max\limits_i\max\limits_k|d_{0w}(k)-d_{iw}(k)|}{|d_{0w}(k)-d_{iw}(k)|+\rho\max\limits_i\max\limits_k|d_{0w}(k)-d_{iw}(k)|} \tag{8}$$

$$\gamma(\xi_0,\xi)=\frac{1}{t}\sum_{k=1}^{t}\gamma(\xi_0(k),\xi_i(k))\quad(i=1,2,\cdots,m) \tag{9}$$

容易验证 $\gamma(\xi_0,\xi_i)$ 满足灰关联四公理,$0\leqslant\rho\leqslant1$ 称为分辨系数。

定义7 记 $\xi=\{\xi_i\mid i=0,1,2,\cdots,m\}$,$\gamma\in\Gamma$,则称$(\xi,\Gamma)$为基于饱和信息维的灰关联空间,式(8)、(9)即为信息关联度的一般计算公式。

灰关联分析是距离空间与点集拓扑空间的升华与结合,在式(8)中,$|d_{0w}(k)-d_{iw}(k)|$是距离测度,而$(\min\limits_{i,k}\Delta_i(k),\max\limits_{i,k}\Delta_i(k))$是$\xi(k)$与$\xi_0(k)$的比较环境,也是$\xi_i(k)$的邻域,它含有点集拓扑的信息,$\rho$的作用在于调节比较环境的强弱。当$\rho=0$时,环境消失;当$\rho=1$时,环境被原封不动地保持着。灰色关联分析依据关联度来确定因子之间的关联序,是研究系统中多个因素之间相互作用、相互关联的一种统计方法,是各因素发展态势的量化比较。当系统内存在复杂的相互影响,在其效果、结构、整体性能、优劣度、权重等方面所吸收与采用的信息不明确、不完整时,它是一种有力的定量分析工具。

进行关联分析时,首先进行白化(确定)的信息特征的关联度计算,其一般步骤为:

(1)构造关联空间(Ω'_{ξ_0},Γ),在获取信息基础上进行数据可比性处理;

(2)求差序列,记$\Delta_i(k)=|d_{0w}(k)-d_{iw}(k)|$,$\Delta_i=(\Delta_i(1),\Delta_i(2),\cdots,\Delta_i(t))$;

(3)求两级最大差与最小差,记$M=\max\limits_{i,k}\Delta_i(k)$,$m=\min\limits_{i,k}\Delta_i(k)$;

(4)求关联系数,$\gamma_{0i}(k)=m+\rho M/\Delta_i(k)+\rho M$,$\rho\in(0,1)$ $(k=1,2,\cdots,t;\ i=1,2,\cdots,m)$。

然后进行灰色(不完全确定)的信息特征的关联度计算,采用灰中心最大原则(即中心白化值的认同程度最大(文献[6-7]))。

考虑到信息不同维的不同权重,赋予权重数值列:

$$\delta=(\delta_1,\delta_2,\cdots,\delta_k,\cdots,\delta_n)\quad\text{且}\sum_{k=1}^{n}\delta_k=1$$

于是得到$\xi_i(i=1,2,\cdots,m)$与ξ_0相对相关联程度的描述:

$$\gamma(\xi_0,\xi_i)=\frac{1}{n}\sum_{k=1}^{n}\delta_k\gamma_{0i}(k) \tag{10}$$

用式(10)确定的序关系:

$$\gamma(\xi_0,\xi_{i_1})\geqslant\gamma(\xi_0,\xi_{i_2})\geqslant\gamma(\xi_0,\xi_{i_3})\geqslant\cdots\geqslant\gamma(\xi_0,\xi_{i_n}) \tag{11}$$

反映了信息获取子空间中的信息列$\xi_{i_1},\xi_{i_2},\xi_{i_3}\cdots,\xi_{i_n}$与$\xi_0$相关联程度的关联序,也可以解释为接近程度的序关系。

设Ω'_{ξ_0}为包含智能体关注的信息点ξ_0的信息关联子空间,由式(11)确定的接近ξ_0的序列$\xi_{i_1},\xi_{i_2},\xi_{i_3}\cdots,\xi_{i_n}$的形成过程称为基于灰色关联序的信息资源聚焦。

2.3 数值实例分析

设$G_{\alpha,\beta}$表示一生存域为$[\alpha,\beta]$的灰数,在一信息获取技术(数值映射)下,由智能体关注的信息点$\xi_0=f(1.26,3.17,G_{2.2,4.2},2.10)$所张成的关联子空间为:$\Omega_{\xi_0}=\{\xi_1,\xi_2,\xi_3,\xi_4,\xi_5\}$,其中$\xi_i$的维数不相同:

$$\xi_1=f(1.98,2.60,G_{1.2,3.2}),\ \xi_2=f(1.84,2.53,G_{1.8,3.6}),\quad\xi_3=f(1.11,2.75,G_{1.8,2.8},2.65)$$

$$\xi_4=f(1.36,3.08,G_{2.0,3.4},2.80,4.10),\quad\xi_5=f(2.01,2.90,G_{1.9,2.1},3.20,3.96,2.35)$$

显然 ξ_1, ξ_2 是不饱和维, ξ_3 是饱和维, ξ_4, ξ_5 是超饱和维。用本文提到的信息维的增减技术将关联子空间各元素统一到四维, 取权重列 $\delta=(0.2, 0.3, 0.3, 0.2)$, 这里略去计算过程而直接给出结果:

$$\gamma(\xi_0, \xi_1)=0.15,\ \gamma(\xi_0, \xi_2)=0.16,\ \gamma(\xi_0, \xi_3)=0.18,\ \gamma(\xi_0, \xi_4)=0.21,\ \gamma(\xi_0, \xi_5)=0.17$$

由此得到聚焦序列: $\xi_0 \leftarrow \xi_4 \leftarrow \xi_3 \leftarrow \xi_5 \leftarrow \xi_2 \leftarrow \xi_1$。

3　讨 论

(1) 由于信息获取技术的复杂性及关联度计算的柔性环境, 由式(11)确定的接近 ξ_0 的序列 ξ_{i_1}, ξ_{i_2}, ξ_{i_3} …, ξ_{i_n} 的序关系非唯一, 它符合灰色系统"解是非唯一"的思想, 也是信息获取技术逐步深入的体现;

(2) 将本体思想引入到元数据模型中来[1], 建立基于本体的元数据模型, 提出一种面向语义内容的更为全面的信息资源描述方式, 为关联度分析的可操作和可实现提供了一种系统、完整的思路;

(3) 仙农信息熵是信息领域中广泛使用的一种度量信息的方法, 用拓扑结构来描述信息空间, 在此基础上建立基于信息熵的分层、分维的信息度量空间, 折射出这种聚焦方法的可行性和应用前景。

参 考 文 献:

[1] 黄宏斌. 基于语义关系的× × 信息资源聚焦服务方法及关键技术研究[D]. 长沙: 国防科技大学, 2007.

[2] Feng L S, Yi L. On Measure of Information Content of Grey Numbers[J]. The International Journal of Systems & Cybernetics, 2006, 35(6): 1256 – 1264.

[3] 汪小龙. 信息获取科学的若干问题研究[D]. 合肥: 中国科技大学, 2003.

[4] Papazoglou M P, Proper H A, Yang J. Landscaping the Information Space of Larg Multi-database Networks[J]. Data and Knowledge Engineering, 2001, 36(3): 251– 281.

[5] 腾书华, 周石琳, 孙即祥, 等. 基于条件熵的不完备信息系统属性约简算法[J]. 国防科技大学学报, 2010, 32(1): 90– 94.

[6] 罗党, 刘思峰. 不完备信息系统的灰色关联决策方法[J]. 应用科学学报, 2005, 23(4): 57– 60.

[7] 刘思峰, 谢乃明. 灰色系统理论及应用[M]. 北京: 科学出版社, 2008.

(上接第 140 页)

3　结 论

本文通过对两类特殊类型 Feistel 密码等价结构的详细刻画和分析, 给出了基于等价结构的改进 Square 攻击。并分别以SNAKE(2)和 CLEFIA 为例, 给出了基于等价结构 Square 攻击的具体过程。攻击结果表明, 等价结构可大大降低特殊类型Feistel 密码 Square 攻击的时间复杂度, 从而能较好地改进这两类Feistel 密码的 Square 攻击。

参 考 文 献:

[1] Daemen J, Knudsen L, Rijmen V. The Block Cipher Square[C]// FSE 1997, LNCS 1267: 149– 165.

[2] Lucks S. The Saturation Attack-a Bait for Twofish[C]// FSE 2001, LNCS 2355: 1– 15.

[3] Biryukov A, Shamir A. Structural Cryptanalysis of SASAS[C]// EUROCRYPT 2001, LNCS 2229: 394– 405.

[4] Knudsen L, Wagner D. Integral Cryptanalysis[C]// FSE 2002, LNCS 2365: 112– 127.

[5] Duo L, Li C, Feng K. New Observation on Camellia[C]// SAC 2005, LNCS 3897: 51– 64.

[6] Shirai T, Shibutani K, Akishita T, et al. The 128-Bit Blockcipher CLEFIA[C]// FSE 2007, LNCS 4593: 181– 195.

[7] Lee C, Cha Y. The Block Cipher: SNAKE with Provable Resistance Against DC and LC Attacks[C]// JW– ISC 97, 1997: 3– 17.

[8] The 128-Bit Blockcipher CLEFIA: Security and Performance Evaluation[R]. Sony Corporation, Revision 1.0, June 1, 2007.