

RST 与 SPA 的广义等价

朱红宁 张 斌

(东北大学信息科学与工程学院 沈阳 110004)

摘 要 通过对粗糙集理论(RST)中的基本概念“正区”(Positive Region)、“负区”(Negative Region)和“边区”(Boundary Region)与集对分析(SPA)中的基本概念“同一度”、“差异度”和“对立度”的分析,从人类认识世界的这个广义角度探讨了 RST 与 SPA 的等价性。

关键词 粗糙集理论,集对分析,含糊集,等价

中图法分类号 TP393 文献标识码 A

Generalized Equivalence between RST and SPA

ZHU Hong-ning ZHANG Bin

(College of Information Science and Engineering, Northeastern University, Shenyang 110004, China)

Abstract Based on the analysis of the essential conception about positive region, negative region, boundary region of RST and degree of sameness, difference, opposition of SPA, discussed the equivalence between RST and SPA from the generalized angle of human cognizing the world.

Keywords RST, SPA, Vague sets, Equivalence

近几十年来,模糊集^[1]、粗糙集^[2]、灰色系统^[3]、直觉模糊集^[4]、集对分析^[5]和 Vague 集^[6]等许多数学理论相继涌现,它们虽然从不同角度分别研究了不精确(模糊)、不确定(随机)与不完全(灰色)数据的处理,但这些理论之间存在着密切的关系,有的甚至完全等价。如:1986 年,保加利亚学者 Atanassov 在模糊集的基础上,提出了直觉模糊集(Intuitionistic Fuzzy Sets)的概念,直觉模糊集是模糊集的推广,模糊集是直觉模糊集的特殊情形。1993 年, Gau 和 Buehrer 定义了 Vague 集。1996 年, Bustince 和 Burillo 指出 Vague 集的概念与 Atanassov 的直觉模糊集是相同的^[7]。

本文认为搞清上述理论之间的关系对进一步研究与发展它们至关重要。从人类认识世界的角度,本文运用 Vague 集中的真、假隶属度概念进一步分析了 RST 中的模糊边区;同时,用 SPA 中的同一度、差异度、对立度逐层嵌套地分析了差异度,从而证明了 RST 和 SPA 的广义等价性。

1 基础理论

1.1 RST

粗糙集作为一种处理不精确、不确定与不完全数据的新数学理论,最初是由波兰数学家 Z. Pawlak 于 1982 年提出的。

(1)不可分辨(indistinguishable)

集合 $R \subseteq U \times U$, U 称为论域, R 为满足自反性(reflexive)、对称性(symmetric)及传递性(transitive)的集合,在 U/R 的元素被称为基本集合,并且存在 xRy , 则称 x 和 y 为不可

分辨。

(2)下近似(lower approximations),即正区(positive region, PR)

如果集合 $A \subseteq U$, 则定义下近似为:

$$PR_R(A) = \underline{apr}(A) = \{x \in U \mid [x]_R \subseteq A\} = \bigcup \{[x]_R \in \frac{U}{R} \mid [x]_R \subseteq A\}$$

$$\text{where } [x]_R = \{y \mid xRy\} \quad (1)$$

(3)上近似(upper approximations)

如果集合 $A \subseteq U$, 则定义上近似为:

$$\overline{apr}(A) = \{x \in U \mid [x]_R \cap A \neq \emptyset\} = \bigcup \{[x]_R \in \frac{U}{R} \mid [x]_R \cap A \neq \emptyset\}$$

$$\text{where } [x]_R = \{y \mid xRy\} \quad (2)$$

(4)负区(negative region, NR)

定义

$$NR_R(A) = U - \overline{apr}(A) \quad (3)$$

为 A 的负区。

(5)边区(boundary region, BR)

定义

$$BR_R(A) = \overline{apr}(A) - \underline{apr}(A) \quad (4)$$

为 A 的边区,如果

$$\underline{apr}(A) \neq \overline{apr}(A) \quad (5)$$

则称 A 为粗糙集的边区。

1.2 SPA

中国学者赵克勤于 20 世纪 80 年代末提出集对分析,即

从同、异、反 3 方面表征两集合(集对) $H(A, B)$ 的关系,它主要通过同、异、反联系度来描述。

(1) 联系度

$$\mu(w) = \frac{S}{N} + \frac{F}{N}i + \frac{P}{N}j \quad (6)$$

其中, N 指集对的特性总数, S 指共同具备的特性总数, P 指具有相对立的特性总数, $F = N - S - P$ 指既不相对立, 又不共同具备的特性总数。

S/N 为两集合在问题 w 下的同一度;

F/N 为两集合在问题 w 下的差异度;

P/N 为两集合在问题 w 下的对立度。

式(6)中, $i \in [-1, 1]$, j 一般情况下取 -1 , 可简记为: $\mu = a + bi + cj$ 。其中, a, b, c 满足归一化条件: $a + b + c = 1$ 。

(2) 集对势

当联系度 $\mu = a + bi + cj$ 中 $c \neq 0$ 时, 同一度 a 与对立度 c 的比值 a/c 为所论集对在指定问题背景下的联系势或集对势, 用 “ shi ” 表示,

$$shi(H) = a/c \quad (7)$$

$a/c > 1$, 则称 $shi(H)$ 为集对同势; $a/c = 1$, 则称 $shi(H)$ 为集对均势; $a/c < 1$, 则称 $shi(H)$ 为集对反势。

2 广义的解释

实质上, RST 与 SPA 都是人们认识世界的方法, RST 体现了西方哲学的思辨求是; SPA 体现了东方哲学的实事求是。它们根植于同一个基础——集合论, 体现了东西方哲学在认识世界上的互补与殊途同归。

人类认知是用已知理论去认识未知事物的过程, 是一个逐步“去粗存真”、“去伪存真”的过程。也就是说, 通过不断地认识事物的部分是什么、不是什么, 最终确定事物的全体是什么、不是什么。

本文认为认识事物的关键在于对事物边界的分析。由于人类认识水平的局限, 对事物边界的认识是模糊的。人类对事物的认识程度可以用事物边界区域的大小和模糊状况来描述, 事物边界区域的不断缩小和模糊状况逐步澄清代表着人类对事物认识的加深。

2.1 RST 的广义拓展

RST 是建立在分类机制的基础上的, 它将分类理解为在特定空间上的等价关系, 而等价关系构成了对该空间的划分。RST 将知识理解为对数据的划分, 每一被划分的集合称为概念。RST 的主要思想是利用已有的知识库, 将不精确或不确定的知识用已有的知识库中的知识来(近似)刻画^[8]。

从人类认知的角度, RST 可以表述为: 对于未知事物 A , 用已知事物的知识 $[x]_R$ 确定 A 的 PR, NR 和 BR , 即: $PR \subseteq A; NR \cap A = \emptyset; BR \not\subseteq A \wedge BR \cap A \neq \emptyset$ 。

对于 BR , 引入 Vague 集的真、假隶属度来逐步澄清它。

Vague 集是对模糊集的扩充, 它弥补了模糊集中隶属函数的不足, 提出了从真假(正反)两个方面对研究对象进行描述。

定义 1 U 是论域, 其中的任意一个元素用 x 表示, U 中的一个 Vague 集 V 用一个真隶属函数 t_V 和假隶属函数 f_V 表示。 $t_V(x)$ 是从支持 x 的证据所导出的 x 的隶属度下界, $f_V(x)$ 则是从反对 x 的证据所导出的 x 的否定隶属度下界,

$t_V(x)$ 和 $f_V(x)$ 将区间 $[0, 1]$ 中的一个实数与 U 中的每一个点联系起来, 即:

$$t_V: U \rightarrow [0, 1]; f_V: U \rightarrow [0, 1] \quad (8)$$

其中, $t_V(x) + f_V(x) \leq 1; x \in V$, 可表示为 $[t_V, 1 - f_V]$ 。

由式(1)、式(2)和式(4)可知, BR 由 $[x]_R^{(T)}$ 和 $[x]_R^{(F)}$ 组成, 即:

$$BR = [x]_R^{(T)} \cup [x]_R^{(F)} \quad (9)$$

其中,

$$[x]_R^{(T)} = \{x \in U \mid [x]_R \cap A \neq \emptyset \wedge [x]_R \cap A \neq \Phi\}$$

$$[x]_R^{(F)} = \{x \in U \mid [x]_R - A \neq \emptyset \wedge [x]_R \cap A \neq \Phi\}$$

如图 1 所示, 对于未知事物 A 的边界区域(去掉中央实心圆形的矩形区域), 已知概念 C (右下角圆形)与 A 之间存在部分 $([x]_R^{(T)})$ 属于 A 和部分 $([x]_R^{(F)})$ 不属于 A 。

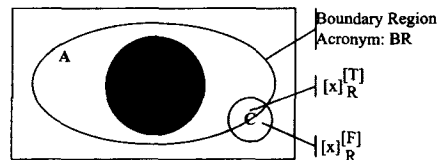


图1 粗糙集边界区域的划分

实际上, 概念 C 与未知事物 A 的属于、不属于关系是很难清晰界定的, 所以 Vague 集用真、假隶属度来确定。 $t_V(x) + f_V(x) \leq 1$ 是一个准确的描述, 因为 $t_V(x) + f_V(x) = 1$ 是一个理想状态, 是人类认识事物的终极状态。 $1 - t_V(x) - f_V(x) > 0$ 是由于认识提高而引发的新一轮模糊。

2.2 SPA 的广义拓展

SPA 从两个集合同(同一)、异(差异不确定)、反(对立)这个角度研究不确定性。在基本概念、研究思路、处理方式等方面都与其它不确定性理论有根本性的区别。采取了对不确定性加以客观承认、系统刻画、具体分析的态度, 从而使研究结果更加贴近实际。

从人类认知的角度, SPA 可以表述为: 对于未知事物 A , 用已知事物的知识 B 分析与 A 的同一(Sameness)、差异(Difference)和对立(Opposition), 如图 2 所示。即: 事物 A 与知识 B 的联系度 $\mu = a + bi + cj$, 其中, a 为同一度; b 为差异度; c 为对立度。

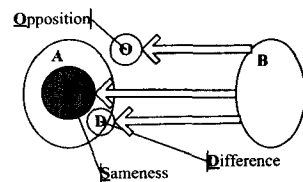


图2 集对分析与人类认知

实际上, 人类认知受人类现有知识的限制, 对差异度的认识更是如此。随着知识的不断涌现, 在差异度中会产生新的同一、差异与对立。也就是说, 对于差异度的逐层嵌套的集对分析势必导致差异度的不断缩小。

2.3 RST 与 SPA 的关系

RST 与 SPA 都是针对不确定性研究的理论。从人类认知的角度看, 它们都是用现有知识确定未知事物的是与否, 更主要的, 它们都是分析未知事物边界区域的方法。

为了说明 RST 与 SPA 的广义等价性, 本文采用 Vague 集的真(t)、假(f)隶属度与逐层嵌套的集对分析作为二者之

间的桥梁,如图3所示。

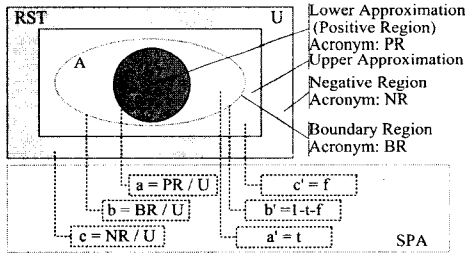


图3 RST与SPA关系

(1)对于论域 U 上的一个未知事物 A ,构造集对 $H(A, [x]_R)$,则有:

$$\text{同一: } S = PR = \bigcup \{[x]_R \in \frac{U}{R} \mid [x]_R \subseteq A\};$$

$$\text{差异: } D = BR = \bigcup \{[x]_R \in \frac{U}{R} \mid [x]_R \cap A \neq \emptyset \wedge [x]_R \not\subseteq A\};$$

$$\text{对立: } O = NR = \bigcup \{[x]_R \in \frac{U}{R} \mid [x]_R \cap A = \emptyset\}.$$

则对于联系度 $\mu = a + bi + cj$,有 $a = PR/U$; $b = BR/U$; $c = NR/U$ 。

(2)对于未知事物 A 的 BR (边界区域或称差异),构建集对 $H'(BR, [x]_{R'})$,则有:

$$\text{同一: } S = \bigcup [x]_{R'}^{(T)};$$

$$\text{差异: } D = BR - S - O;$$

$$\text{对立: } O = \bigcup [x]_{R'}^{(F)}.$$

则,对于联系度 $\mu' = a' + b'i + c'j$,有 $a' = t$; $b' = 1 - t - f$; $c' = f$ 。

一般地, $[x]_R$ 中关于论域 U 的关系集合 R 与 $[x]_{R'}$ 中关于论域 U 的关系集合 R' 存在 $R \subset R'$,因此,随着人类认识的提高,应对 $U \times U$ 作进一步的刻画。

3 等价性探讨

常用的数学证明有两种:构造性证明和存在性证明。本文对于RST与SPA的广义等价的证明采用构造性证明。

3.1 一般性假设

论域 U :由全部可能成员 $M^{(E)}$ 和全部可能关系 $R^{(E)}$ 组成,表示为: $U = \langle M^{(E)}, R^{(E)} \rangle$,其中, $R^{(E)} \subseteq R^{(E)} \times R^{(E)}$, $M^{(E)}$ 代表全部成员及其全部可能组合。另外,用 $M^{(K)}$ 代表已知成员集合;用 $R^{(K)}$ 代表已知关系集合,则: $M^{(K)} \subseteq M^{(E)}$, $R^{(K)} \subseteq R^{(E)}$ 。对于 $\forall m^{(K)} \in M^{(K)}$, $[m^{(K)}]_{R^{(K)}}$ 表示与已知成员 $m^{(K)}$ 是有已知关系 $R^{(K)}$ 的全部可能成员的集合。

事物 A :由成员 $M^{(A)} = \{m_1, m_2, \dots\}$ 和关系 $R^{(A)} = \{r_1, r_2, \dots\}$ 构成,表示为: $A = \langle M^{(A)}, R^{(A)} \rangle$ 。则: $M^{(A)} \subseteq M^{(E)}$, $R^{(A)} \subseteq R^{(E)}$ 。

如果 A 是一个未知事物, $[m^{(K)}]_{R^{(K)}}$ 与 $M^{(A)}$ 之间存在3种关系:

$$(1) [m^{(K)}]_{R^{(K)}} \subseteq M^{(A)};$$

$$(2) [m^{(K)}]_{R^{(K)}} \not\subseteq M^{(A)} \wedge [m^{(K)}]_{R^{(K)}} \cap M^{(A)} \neq \emptyset;$$

$$(3) [m^{(K)}]_{R^{(K)}} \cap M^{(A)} = \emptyset.$$

第一种关系:从RST的角度,说明了未知事物 A 包含论域 U 中的已知成员,也就是说,已知成员的并集构成了未知事物 A 的正区;从SPA的角度,说明了未知事物 A 与论域 U

中的已知成员具有共性,即同一性。

第二种关系:从RST的角度,说明了未知事物 A 不包含论域 U 中的已知成员,但它们之间存在着相同和相异的成分,也就是说,已知成员的并集构成了未知事物 A 的边界区域;从SPA的角度,说明了未知事物 A 与论域 U 中的已知成员具有差异性。

第三种关系:从RST的角度,说明了未知事物 A 不包含论域 U 中的已知成员,且它们之间毫无相似之处,也就是说,已知成员的并集构成了未知事物 A 的负区;从SPA的角度,说明了未知事物 A 与论域 U 中的已知成员具有对立性。

由上可知,在一般性假设前提下,RST与SPA是相通的。

一般地,我们认为:随着已知关系集合 $R^{(K)}$ 的变大,即人类认知水平的提高,对论域 U 的划分变细,基本集合变小,未知事物的正区和负区变大,边界区域变小。也就是说,人类的认识变清。

若存在已知关系集合 $R_{(1)}^{(K)}$,使得 $R^{(K)} \subset R_{(1)}^{(K)}$,

则有:

$$[m^{(K)}]_{R_{(1)}^{(K)}} \subset [m^{(K)}]_{R^{(K)}}$$

但是:

$$\bigcup ([m^{(K)}]_{R^{(K)}}) \subset \bigcup ([m^{(K)}]_{R_{(1)}^{(K)}}) \mid m^{(K)} \in M^{(K)}$$

3.2 等价性构造

用 $\text{card}(S)$ 代表集合 S 的基数,对于论域 U 中的未知事物 A ,作如下构造:

$$\text{Let: } a = \frac{\text{card}(PR)}{\text{card}(U)}; b = \frac{\text{card}(BR)}{\text{card}(U)}; c = \frac{\text{card}(NR)}{\text{card}(U)}$$

对于 A 的边界区域 BR ,进一步构造:

$$\text{Let: } a' = t; b' = 1 - t - f; c' = f$$

则对于 $\mu = a + bi + cj$,有:

$$\mu = \frac{\text{card}(PR)}{\text{card}(U)} + \frac{\text{card}(BR)}{\text{card}(U)}i + \frac{\text{card}(NR)}{\text{card}(U)}j$$

进一步,有:

$$\mu = \left(\frac{\text{card}(PR)}{\text{card}(U)} + \frac{\text{card}(BR)}{\text{card}(U)}t \right) + \left(\frac{\text{card}(BR)}{\text{card}(U)}(1-t-f) \right)i + \left(\frac{\text{card}(NR)}{\text{card}(U)} + \frac{\text{card}(BR)}{\text{card}(U)}f \right)j$$

以上的构造表示,使用Vague集中的真、假隶属度分析RST中的边界区域等价于使用集对理论进一步分析SPA中的差异度。

3.3 逐层嵌套SPA的收敛性证明

特殊地,在SPA中,假定对于差异度 b 的进一步分析,同、异、反的比例不变,即有 $a' = a$; $b' = b$; $c' = c$ 。并且,认为新的差异度 bb' 同样遵循这样的比例。以下是对逐层嵌套SPA收敛性的简单证明。

Proof:

$$\text{Let: } \mu = a + bi + cj$$

$$\text{where } a + b + c = 1$$

more

$$\mu = a(1+b) + b^2i + c(1+b)j$$

...

$$\mu = a \frac{1-b^n}{1-b} + b^n i + c \frac{1-b^n}{1-b} j$$

$$\lim_{n \rightarrow +\infty} \mu = \frac{a}{1-b} + \frac{c}{1-b} j = \frac{a}{a+c} + \frac{c}{a+c} j = a + cj$$

$$\text{where } a + c = 1$$

Q. E. D.

一般地,在 SPA 中,对于差异度的每一次进一步分析都不可能遵循同样的同、异、反比例。即有: $a' \neq a; b' \neq b; c' \neq c$ 。

Proof:

$$\mu = a_0 + b_0 i + c_0 j$$

$$\text{where } a_0 + b_0 + c_0 = 1$$

more

$$\mu = (a_0 + a_1 b_0) + b_0 b_1 i + (c_0 + c_1 b_0) j$$

...

$$\text{Let: } b_{-1} = 1$$

$$\text{for } \mu = a + bi + cj$$

$$\text{where: } a = \sum_{i=0}^n a_i \prod_{j=0}^i b_{j-1}, b = \prod_{i=0}^n b_i, c = \sum_{i=0}^n c_i \prod_{j=0}^i b_{j-1}$$

$$\text{Let: } a_{\max} = \max(a_i), a_{\min} = \min(a_i);$$

$$b_{\max} = \max(b_i), b_{\min} = \min(b_i);$$

$$c_{\max} = \max(c_i), c_{\min} = \min(c_i);$$

$$\text{where: } i = 0, 1, 2, \dots, n$$

$$\lim_{n \rightarrow +\infty} b_{\min}^n \leq \lim_{n \rightarrow +\infty} b \leq \lim_{n \rightarrow +\infty} b_{\max}^n \rightarrow \lim_{n \rightarrow +\infty} b = 0$$

$$\lim_{n \rightarrow +\infty} \sum_{i=0}^n a_{\min} b_{\max}^{i-1} \leq \lim_{n \rightarrow +\infty} a \leq \lim_{n \rightarrow +\infty} \sum_{i=0}^n a_{\max} b_{\max}^{i-1}$$

$$\rightarrow \frac{a_{\min}}{1-b_{\min}} \leq \lim_{n \rightarrow +\infty} a \leq \frac{a_{\max}}{1-b_{\max}}$$

$$\lim_{n \rightarrow +\infty} \sum_{i=0}^n c_{\min} b_{\max}^{i-1} \leq \lim_{n \rightarrow +\infty} c \leq \lim_{n \rightarrow +\infty} \sum_{i=0}^n c_{\max} b_{\max}^{i-1}$$

$$\rightarrow \frac{c_{\min}}{1-b_{\min}} \leq \lim_{n \rightarrow +\infty} c \leq \frac{c_{\max}}{1-b_{\max}}$$

$$\mu = \frac{a_i}{\max(a_i + c_j)} + \frac{c_j}{\max(a_i + c_j)} j = a_i + c_j j$$

$$\text{where } a_i + c_j = 1, i, j = 0, 1, 2, \dots, n$$

Q. E. D.

结束语 本文在分析 RST 与 SPA 基本理论的前提下,从人类认知的角度,引入 Vague 集的真、假隶属度研究 RST 的边界区域,从而达到扩展 RST 的目的;同时,采用逐层嵌套的方式运用集对理论进一步分析 SPA 中的差异度,从而达到扩展 SPA 的目的。进而,为了说明 RST 与 SPA 的理论相通,进行了一般性假设和等价性构造,并且证明了逐层嵌套 SPA 的收敛性。本文认为这对进一步发展 RST 和 SPA 有着重要的意义。

参考文献

(上接第 157 页)

- [9] Kanoun K, Spainhower L. Dependability Benchmarking for Computer Systems[M]. Chapter 1. IEEE, July 2008
- [10] Wang Huiqing, Pang Yonggang, Du Ye, et al. Evaluation of network dependability using event injection[C]//Proceedings of International Workshop on Web-Based Internet Computing for Science and Engineering. Berlin: Springer, 2006: 991-998
- [11] IBM. Georgia Tech implements a cool solution for green HPC with IBM, [EB/OL]. ftp://ftp.software.ibm.com/common/ssi/pm/ab/n/oic03003usen/OIC03003USEN.PDF
- [12] Joseph Ellard D. Trace-based analyses and optimizations for network storage servers[D]. Harvard University, 2004
- [13] http://www.qudong.com/soft/program/Sql%20Server/jiqiao-yunyong/20080328/4795_2.html
- [14] Dimitrijević Z D, Rangaswami R, Cahng E. Design and implementation of semipreemptible IO[C]//Proceeding of the Second Usenix File and Storage Technology. San Francisco, 2003: 145-158
- [15] Aranya A, Wright C P, Zadok E. Tracefs: A file system to trace them all[C]//Proceedings of the 3rd USENIX Conference on File and Storage Technologies. San Francisco: USENIX, 2004: 129-143
- [16] Roselli D, Lorch J R, Anderson T E. A comparison of file system workloads[C]//Proceedings of the Annual USENIX Technical Conference. San Diego: USENIX, 2000: 41-54
- [17] Zhu Qingbo, Chen Zhifeng, Tan Lin, et al. Hibernator: helping disk arrays sleep through the winter[C]//Proceedings of the 20th ACM Symposium on Operating Systems Principles. NY: ACM, 2005: 177-190
- [18] Joukov N, Wong T, Zadok E. Accurate and efficient replaying of file system traces[C]//Proceedings of the 4th Conference on File and Storage Technologies. San Francisco: USENIX, 2005: 336-350
- [19] Feng Wuchang, Oel A, Bezzaz A, et al. TCPivo: a high performance packet replay engine[C]//Proceeding of the ACM SIGCOMM workshop on Models, methods and tools for reproducible network research. NY: ACM, 2003: 57-64
- [20] Cheng Yuchung, Hölzle U, Cardwell N, et al. Monkey see, monkey do: a tool for TCP tracing and replaying[C]//Proceedings of USENIX Annual Technical Conference. Berkeley: USENIX, 2004: 87-98
- [21] Hong Seungsun, WU S F. On interactive internet traffic replay [C]//Proceedings of the 8th International Symposium on Recent Advances in Intrusion Detection (RAID). Berlin: Springer, 2005: 247-264
- [22] Geimef M, Wolf F, Wylie B J N, et al. Scalable parallel trace-based performance analysis[C]//Proceeding in the 13th European PVM/MPI User's Group Meeting (Bonn, Germany). Berlin: Springer, 2006: 303-312
- [23] Kurmas Z, Keeton K. Using the distiller to direct the development of self-configuration software[C]//Proceedings of International Conference on Autonomic Computing 2004. NY: IEEE, 2004: 172-179