第17卷 第9期
2007年9月

Vol.17 No.9
Sep. 2007

不同粒度中的保序性

王向阳[1],张燕平[2]
(1.安徽电气工程职业技术学院 动力系,安徽 合肥 230022;
2.安徽大学 人工智能研究所,安徽 合肥 230039)

摘 要:信息的不完备、不确定是复杂决策环境中不可避免的问题。从不完备、不确定的海量信息中发现某种特定目标的潜在有用的知识,若没有一种降低不确定性、变复杂为简的有效处理海量数据集的方法,而直接利用相应的海量数据进行挖掘,不仅数据处理效率低,更由于传统的方法基本是以不可分辨关系为基础的,因此,其变复杂为简的方法要求形成互不相交的等价数据子集,这在实际处理中不易满足。文中给出商空间的保序性的有关概念,给出求商(拟)半序直接方法及其数学证明,然后以序关系代替传统的不可分辨关系,形成逼近某种特定目标的子集序列,从而获取潜在有用的知识。
关键词:商空间;保序性;等价关系;半序
中图分类号:TP18 **文献标识码**:A **文章编号**:1673-629X(2007)09-0071-04

The Holding Order Relation in Different Granula

WANG Xiang-yang[1], ZHANG Yan-ping[2]
(1. Anhui Electrical Engineering Professional Technique College, Hefei 230022, China;
2. Institute of Artificial Intelligence, Anhui University, Hefei 230039, China)

Abstract: There are some questions that cannot be avoided in complex decision making environment, such as imperfect and uncertainty information. If there is no method that can reduce the nondeterminacy and make complicated thing to simple in dealing with mass data set usefully, but using the matching mass data directly to find out the potential information with some special target from the imperfect and uncertainty mass data. If you do in this way, you will find it is inefficient. And the traditional method is based on the undistinguishability, so the method that makes the complicated thing to simple requires to form a mutually disjointed equivalent data subset, but it is hard to meet in reality. Firstly, this paper offers the relevant concept about quotient space, then provides the direct method to obtain the quotient semi-order and its mathematical proof, at last in order to get the potential useful information, it uses the order relation to replace the traditional one that is undistinguished and forms a sequence of subsets which is close to some special target.
Key words: quotient space; holding order property; equivalence relation; half order

0 引 言

由于客观世界的复杂、多变性和人类自身认识的局限、主观性,致使人们所获得、所处理的信息知识中,往往含有不肯定、不确定、不完全甚至不一致的成分。不确定性是人们思维过程中经常出现的一种心理状态,人们在日常生活中要处理大量的不确定性问题,并在此基础上做出决策,没有被信息的不确定性所困惑。

文献[1]提出这样一个粒度与信息的关系假设:对某个不确定性信息 I,存在一个层次 X_1,I 在 X_1 中可表示为确定的信息。反之,一个在某层次 X_1 上为确定的信息 I,在比 X_1 粒度更小的层次 X_2 上,I 将表现出某种程度的不确定性。并给出描述这种从不同粒度观察、分析世界的数学模型——商空间理论。

信息的不完备、不确定是复杂决策环境中不可避免的问题。从不完备、不确定的海量信息中发现某种特定目标的潜在有用的知识,若没有一种降低不确定性、变复杂为简的有效处理海量数据集的方法,而直接利用相应的海量数据进行挖掘,不仅数据处理效率低,更由于传统的方法基本是以不可分辨关系为基础的,因此,其变复杂为简的方法要求形成互不相交的等价数据子集,这在实际处理中不易满足。

1 粒度划分中保序性

1.1 商空间理论简介

人类在问题求解过程中能在不同层次的分析中,

收稿日期:2007-04-03
基金项目:国家自然科学基金资助项目(60675031);安徽省自然科学基金资助项目(05042000208);安徽省教育厅重点自然科学研究项目(2006KJ015A)
作者简介:王向阳(1961-),男,浙江宁波人,副教授,研究方向为人工智能、机器学习及应用;张燕平,博士,教授,研究方向为人工智能、机器学习、人工神经网络应用。

正确地使用与它们相适应的信息，也就是能正确地把不同的信息应用到和它相应的层次。也许这就是人类能自如地处理各种不确定性的原因所在，用商空间模型描述信息的不确定性，恰好反映出这种问题决策与信息粒度之间的关系。

在商空间模型中，用一个三元组(X,f,T) 描述一个问题。X 表示问题的论域；$f(.)$ 表示论域的属性，可用函数 $f:X\rightarrow Y$ 表示；T 是论域的结构，指论域 X 中各元素的相互关系。分析或求解问题(X,f,T)，是指对论域 X 及其有关的结构、属性进行分析、研究。商空间法就是将不同的粒度世界与数学上的商集概念统一起来，表示对象模型的方法，即以商集作为不同粒度世界的数学模型的方法[2]。

"粒度"，顾名思义，就是取不同大小的对象。也就是说，将原来"粗粒度"的大对象分割为若干"细粒度"的小对象，或者把若干小对象合并成一个大的粗粒度对象，进行研究。将原来的空间分成粒度粗细不同的空间，这与数学中的"划分"(Partition)概念完全吻合，而数学上"划分"与"等价关系"是等价的，故"等价关系"就成为建立商空间理论的一个基本概念。

在此处定义的粒度世界，与 Pawlak 所提出的信息粒度(IG)不完全相同。所谓的信息粒度，是指人类在解决处理和存储信息的有限能力上的一种反映，即人类在解决和处理大量复杂信息问题时，由于人类的能力有限，需把大量复杂信息按其各自的特征和性能将其划分成数个较简单的信息块，以方便处理，每个如此划分的信息块就被认为是一个粒度[3]。由于信息粒度按等价类进行划分，因此这种划分法仅仅是问题论域和所属关系的粒度变化[4,5]，但其空间结构并没引入粒度的变化。与 Zadeh 所讨论的粒度计算相比[6]，利用模糊等价关系可以将原来的商空间理论推广成模糊商空间理论，从而将两者联系起来[7]。

从不同粒度来考察一个对象，希望能从较粗的粒度空间上，尽可能多地观察到感兴趣的性质。但是，并不是随意地划分都可以达到此目的。因此，如何划分对象，使人们能用尽可能少的计算获得尽可能多的有用信息，是一个需要研究的问题。文中从粒度划分形成的序关系进行讨论。

1.2 半序空间商结构的求法

下面仅讨论半序空间情况。

定义 1：设(X,T) 是半序空间，即在 X 的部分元素上有如下的关系 $<$，满足：

(1) $x<y$ 且 $y<x\Rightarrow x=y$

(2) $x<y,y<z\Rightarrow x<z$

(若条件(1) 不成立，则称(X,T) 是拟半序空间)。

1.2.1 商拟半序的求法

设[X] 是半序空间(X,T) 的商集，在[X] 中引入商结构[T]，问什么情况下[T] 也是半序，以及如何求[T]。

在文献[1] 中给出求商结构[T] 的方法如下：

(1) 将(X,T) 中的 T，化作某种相应的拓扑(右序拓扑 T_R)，得(X,T_R)；

(2) 求(X,T_R) 商拓扑空间([X]，[T_R])；

(3) 将[T_R] 化成为对应的(拟) 半序，则称$[T_R]_S$ 是 T 对应的商拟半序，简记为[T]。

定义 2：(T_R 的定义) 设(X,T) 是半序集，对 $\forall x\in X$，定义集合

$u(x)\hat{=}\{y\mid <x<y,y\in X\}$

令 $B=\{u(x)\mid x\in X\}$，称以 B 为拓扑基的拓扑 T_R 为 X 的右序拓扑，记为(X,T_R)。这样，就将一个半序集变为对应的拓扑空间。

定义 3：(T_S 的定义) 设(X,T) 是一拓扑空间，用 T 定义如下的"$<$" 关系：

对 $x,y\in X$，$x<y\Leftrightarrow\forall u(x)$，有 $y\in u(x)$，其中 $u(x)$ 表示 x 的开邻域，记为 T_S。

由上述定义的"$<$" 满足传递性，但一般未必能满足自反性"$x<y,y<x\Rightarrow x=y$"，故称为拟半序。

定义 4：设(X,T) 是拓扑空间，由定义 3 定义的 T_S，满足"$x<y,y<x\Rightarrow x=y$"，则称拓扑 T 与 T_S 是相容的。

若关系 T_S 与(X,T) 的拓扑 T 是相容的，则 T_S 是半序，且 X 在关系 $<$ 下成一半序集，记为(X,T_S)，它是(X,T) 拓扑空间对应的半序集，在不发生混淆的地方也简记为(X,T)。

设商拓扑空间([X]，[T_R]) 的[T_R] 与$[T_R]_S$ 是相容的，简称 R 是相容的，则([X]，$[T_R]_S$) 在 $<$ 关系下是半序集，记为([X]，[T])。称这个半序集是由半序集(X,T) 导出的商半序。

命题 1：(保序性) 设 R 是半序空间(X,T) 上的一个等价关系，([X]，[T]) 是对应于 R 的商空间，若 $x,y\in(X,T)$ 且 $x<y$，则

$[x]<[y]$，其中$[x],[y]\in([X],[T])$。

1.2.2 商半序的判别

命题 2：设 R 是相容的，若 $x,y\in[x]$，且 $x<y$，则区间$[x,y]=\{x\mid x<z<y,z\in X=\subset[x]\}$

推论 1：设 R 是相容的，则每个[x] 必是由若干个全连通的两两互不能比较的集合组成。

这里 A 是全连通是指 $\forall x,y\in A\Rightarrow$ 区间$[x,y]\subset A$。所谓集合 A,B 不能比较是指 $\forall x\in A,y\in B$，x 与 y 均不能比较。

命题2及其推论给出商拟半序是商半序的充要条件。为选取划分,提供一个有用的准则。

以上命题详见文献[1]。

1.3　求半序商结构的直接法

上面给出求半序空间的商拟半序的方法,虽可行但很复杂,在文献[1]中还给出另一个直接求法,但未给出证明,下面给出该求法的严格的数学证明。

定义5:设(X,T)是一半序空间,R是一等价关系,做有向图$E(T)$:$e=(x,y)\in E\Leftrightarrow \exists x_0=x,x_1,\cdots,x_k=y,x_i<x_{i+1}(T),i=0,1\cdots$称$E(T)$为$T$的有向图表示。

求商结构的直接法:

仅讨论X是有限的情况。

定义6:设$[X]$是半序空间(X,T)的商集,定义$[X]$上的(拟)半序$[T]$为:$a,b\in[X],(a,b)\in E([T])\Leftrightarrow \exists x\in a,y\in b,(x,y)\in E(T)$。

引理1:$(X,[T_R])$是按定义2定义的右序商拓扑,则$\forall a\in[X]$,有最小开集$u(a)$。

证明:令$u_1(a)=[\bigcup_{x\in a}u(x)]$,其中$u(x)$是$T$中含$x$的最小开集。

一般令$u_i(a)=[\bigcup_{x\in u_{i-1}(a)}u(x)]$,因$X$有限,故上述迭代进行到某一步有$u_n(a)=u_{n-1}(a)$。令$u_n(a)=u(a)$。

下面证明$u(a)$是$[T_R]$拓扑中含a的最小开集。

$u(a)$是开集是显然的,设$v(a)$是$(X,[T_R])$中含a的任一开集,故有$u_1(a)\subset v(a)$,得$[u_1(a)]\subset v(a)$,于是得$u_2(a)\subset v(a)$。利用归纳法,易得$u(a)\subset v(a)$,最后得$u(a)$是含a的最小开集。证毕。

命题3:按定义2、3定义的(拟)半序$[T_R]_S$与定义6求到的$[T]$是等价的。

证明:设按$[T_R]_S$有$a<b$。按$[T_R]_S$的定义,对任给开集$v(a)$,有$b\in v(a)$。由引理1,取$u(a)$是a在$([X],[T_R])$中的最小开集。

于是有$b\in u(a)$,故$\exists y\in b$,及$x_k,x_{k-1},\cdots,x_1,x_0,x_i\in u_i(a),x_0\in a,x_i\in u(x_{i-1}),y\in u(x_k)$。得

$x_0<x_1<\cdots<x_i<y$,即$\exists x_0\in a,y\in b,x_0<y$,得$(x_0,y)\in E(T)$

反之,设$x\in a,y\in b$,且$x<y$,按右序拓扑定义,得$y\in u(x)$,$u(x)$是含x的最小开集,按商拓扑定义有,$y\in u(x)\subset\bigcup_{x\in a}u(x)$,得$b=[y]\in[\bigcup_{x\in a}u(x)]=u_1(a)$。由引理1得$u_1(a)\subset u(a)$是$a$的最小开集,于是对任意$v(a)$(含$a$的开集)有$b\in u_1(a)\subset u(a)\subset v(a)$。按定义3,得$a<b$,即$(a,b)\in[T]$。证毕。

1.4　保序性分析

从命题1,虽然对$[T]$是否是半序,保序性都成立,但是当$[T]$是半序时与$[T]$只是拟半序其性质仍有很大区别。下面先看一个例子。

例:设半序集(X,T)如图1所示,取等价关系R_1:$\{a=\{y_1,y_2,y_3,y_4\},b=\{y_5,y_6,y_7,y_8\},c=\{y_9,y_{10},y_{11}\}\}$,得在$\{X\}$上的拟半序$[T]_1$为:$a<b,b<c,c<a$。显然$[T]_1$不是半序。虽由$y_2<y_{10}$,有$c=[y_{10}]<a=[y_2]$,但又有$a<b<c$。这样$a,b,c$实际上是无区别的(在半序意义下)。

若取R_2:$\{a_1=a,b_1=\{y_5,y_8,y_7\},c_1=\{y_6,y_9,y_{10},y_{11}\}\}$,则得$c_1<a_1<b_1$,得对应的$[T]_2$是半序。其中三个元素$c_1,a_1,b_1$是可区别的。

这个简单例子说明,若划分不得当,可能会落个"事倍功半"的结果(如上例商空间都是三个元素,一个能将三个元素区别开来,一个却不能)。所以若商空间的规模一样,则商结构是半序时,其分辨率最高(指不等价的元素个数最多)。

另一方面,保序性说明不管划分如何取法,在商空间中不会出现相反的情况,即不会有$x<y,x\neq y$,而在商空间中反而有$[x]>[y],[x]\neq[y]$的情况。

这说明元素之间的序关系是元素之间本质的关系之一。

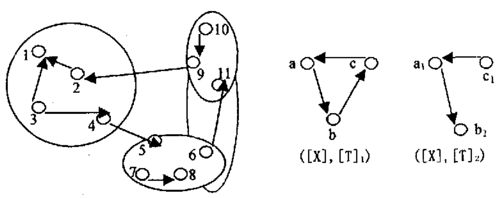

图1　半序集(X,T)

什么样的划分,其对应的商结构是半序,在文献[1]中给出判别的充分必要条件。这里以求商拟半序的直接法给出另外的判别条件。

命题4:设R是半序空间(X,T)上的一个等价关系,$([X],[T])$是对应于R的商空间,则$[T]$是商半序$\Leftrightarrow[T]$对应的有向图$E([T])$无有向圈。

2　应　用

设半序空间(X,T),若用某个方法得到关于X的分层递阶结构$\{(X(\lambda),0\leqslant\lambda\leqslant1\}$,由定义3得具有粒度结构的知识:$\{(X(\lambda),d(\lambda)),0\leqslant\lambda\leqslant1\}$。若记$X$上的半序为$T(1)$,于是由商空间理论对每个商空间

$X(\lambda)$ 将诱导出一个商结构 $T(\lambda)$(一般是拟半序结构),于是有 $\{(X(\lambda), d(\lambda), T(\lambda)), 0 \leqslant \lambda \leqslant 1\}$。这样就得到一个更丰富的具有粒度结构的知识。比如,可利用拟半序商空间的保序性,在不同的商空间上利用“保假原理”进行推理。

总之,得到具有粒度结构的知识后,就可以按照求解的需要,取不同的粒度,从 X' 中抽取出所需不同粗细的信息,再利用商空间理论中的各种保持性原理进行演绎,这样就把这两种思维的模型结合起来。下面利用信息由粗到细的商空间,给出逼近、最后到达目标节点的路由选择。

随着网络的增大,路由器路由选择表也会成比例地增大。增大的表格不仅占用路由器的内存,而且需要更多的CPU时间扫描表格,以及更大的带宽来发送关于表格的状态报告。在某一时刻,网络可能会增大到不可能让每一个路由器都给出至其它每一个路由器的路径表项。因此,对路由的选择不得不引入分级(不同粒度)的概念,这正是满足保序性的不同粒度的数据子集。

首先需确定的是对一个非常大的网络应分几种粒度。这可引用 Kamoun 和 Kleinrock 的结论[7],他们发现有 N 个路由器的子网的最优级数为 $\ln N$,其中每个路由器需要的表项总数为 $e\ln N$。由此给出对应的描述(以簇、区、组等粒度为例):

算法1:相邻簇点的区路由描述:

输入:相邻网络簇点 C_1 和 C_2,约束条件 f,相邻簇的商空间 T_C(即簇的网络拓扑结构图)。

输出:满足约束条件的由 C_1 到 C_2 的区路由 $R(C_1, C_2)$。

(1) 确定 C_2 的每个区点 n($n \in R$ 在 T_C 所属的等价类),并对任一 n,设 m 是与 n 相邻且满足约束条件的区节点,生成一个扩展段,由 n 至 m,记这样的结构为 D_2。

(2) 设 S_1 由 C_1 的区节点及按(1)形成的扩展段构成的一种结构。

(3) 将 D_2、S_1 按(1)的方式再扩展,直至 D_2、S_1 重叠,则 D_2、S_1 的组合结构就是 C_1 到 C_2 的区路由 $R(C_1, C_2)$。

算法2:不相邻簇点的商空间描述:

输入:约束条件 f,相邻簇的商空间 T_C 及相邻簇点的区路由 $R(C_1, C_2)$。

输出:不相邻簇点的商空间 T_{NC}。

(1) 通过扫描 T_C 的边界,确定关于 C_1, C_2 的相邻的簇点 C_C。

(2) 对不相邻的簇点,遍历每一点,按上述生成扩展段的方法,生成各自的扩展结构,直至和 T_C 中的某簇的区路由 R 相重叠。

(3) 设不相邻的簇点为 A,重叠的区路由为 $R(C_1, C_2)$,将 A 作为 C_1, C_2 的相邻簇添加到商空间 T_C,其对应的区路由分别为 $R(A, C_1)$, $R(A, C_2)$。

算法3:网络中簇点的商空间描述:

输入:约束条件 f,网络环境 W 及 K 个簇路由器。

输出:簇商空间 T_C。

(1) $C_1 = \varnothing, C_2 = K, k = 0, T_C$ 就为 K。

(2) 如果 $C_2 = \varnothing$,则停止,否则转(3)。

(3) 利用算法1,计算 $R(C_1, C_2)$。

(4) 利用算法2,计算 $R(A, C_1)$, $R(A, C_2)$,并扩展商空间至不相邻的簇点。

(5) 让 $C_1 = A, C_2 = K - A, k = k + 1$,转(2)。

3 结　语

文中表示了复杂网络中簇点的商空间结构,同理,可将区、组等不同粒度的商空间进行表示,从而在保序性的条件下达到了目标节点。

可见对复杂的对象,可以用序关系代替传统的不可分辨关系,形成逼近某种特定目标的子集序列,从而获取潜在有用的知识。

参考文献:

[1] 张　钹,张　铃. 问题求解的理论及应用[M]. 北京:清华大学出版社,1990.

[2] 张燕平,张　铃,吴　涛. 不同粒度世界的描述法——商空间法[J]. 计算机学报,2004,27(3):328-333.

[3] Pawlak Z. Rough Sets: Theoretical aspects of reasoning about data[M]. Dordrecht: Kluwer Academic Publishers, 1991.

[4] Yao Y Y. Granular computing: basic issues and possible solutions[C]// In: Proc. of Fifth Joint Conference on Information Sciences, Vol. I. Atlantic City, New Jersey, USA: [s. n.], 2000:186-189.

[5] Yao Y Y, Li X. Comparison of rough-set and interval-set models for uncertain reasoning[J]. Fundamental Informatics, 1996,27(2/3):289-298.

[6] Zadeh L A. Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic[J]. Fuzzy Sets and Systems, 1997,90(2):111-127.

[7] 张　铃,张　钹. 基于商空间模型的粒度计算[J]. 软件学报,2003,14(4),770-776.